Truth or Dare: Management Theory at a Crossroads

A dissertation presented

by

Ray Fung

to

The Technology and Operations Management Unit at Harvard Business School

in partial fulfillment of the requirements

for the degree of

Doctor of Business Administration

in the subject of

Technology and Operations Management

Harvard University
Cambridge, Massachusetts

May, 2014

Dissertation Advisor: Clayton Christensen Ray Fung

Truth or Dare: Management Theory at a Crossroads

Abstract

This dissertation examines the state of management theory, whether as espoused by the (largely self-proclaimed) gurus, or by management academics. Given that philosophers of science have determined that theory is supposed to provide reliable, non-obvious predictions I test whether management theory meets those requirements. I examine certain famous guru works for reliability through a case-study method. I examine the published results of management academia through a statistical analysis. I examine the non-obviousness of published management academia's hypotheses through a series of interviews and by posing a survey of those hypotheses to volunteers untrained in management knowledge to determine whether they find those predictions to be obvious. Management theory is currently found to be wanting. However, I then propose a set of prescriptions that might allow management to become a truly progressive discipline as well as what management academics and gurus could fruitfully deliver to audiences today.

1. Management Theory Today – Gurus and Academics

A mere three decades ago, the quest in search of excellence in managerial decision-making and organizational strategy theory was declared over. No longer would managers need to agonize over how to optimize firm performance, nor expend countless hours perusing the business press in hopes of gleaning some new-fangled insight, nor invest years of their time and pricey tuition on MBA & Executive Education coursework to improve their knowledge, and certainly would not require the advice of pricey management consultants. By spending a mere $9.99 at their local bookstore, they could unlock the secrets of managerial excellence…and so can you.

Or so Tom Peters and Robert Waterman - authors of the seminal 1982 business book "In Search of Excellence" (hereafter known as 'Search') - would have the world believe. And armies of aspiring and actual managers surely did believe them. Not only did 'Search' become one of the best-selling books of all time, it also became a veritable literary phenomenon within the business space. While 'Search' sold over 3 million copies within its first four post-launch months alone and over 6 million copies to date, mere sales figures do not do justice to the superlatives of its popularity. The WorldCat library catalog service listed 'Search' as the top-held book held by libraries in the United States from 1989-2006. NPR named 'Search' as "one of the top three business books of the 20th century."[1] A survey by Bloomsbury Publishing anointed 'Search' "The Greatest Business Book of All Time"[2]. To this very day - a full three decades after publication - Search continues to be cited by columnists at top business periodicals such as Forbes as "having certainly shaped my view of the world and of business" [3]

But perhaps the most important contribution that 'Search' provided was not its contents, nor even the myriad accolades that it won, but rather the cottage industry of follow-on popular-press management books that it inspired. If imitation is the sincerest form of flattery, then Peters and Waterman have been eulogized to the heavens, for while the quest in search of excellence had ostensibly been declared completed upon publication in 1982, the quest in search of improved firm performance – or at least the quest in search of book royalties – had apparently only just begun. Need to distinguish between firms that are 'Big Winners and Big Losers', presumably so that your firm can be one of the former? Care to have your firm embark upon a 'Blue Ocean Strategy'? Would you like to develop a firm that is 'Built to Last'? Need to understand how to transform a firm from 'Good to Great'? Want to understand 'How the Mighty Fall' presumably so that you won't do likewise? Want to learn how your firm can be 'Great by Choice' (apparently with the presumption that other firms are simply 'mediocre by choice')? No matter how exotic and obscure your management problem may be, the gurus are all-too-happy to provide convenient theories to solve that problem, if you would only purchase their books. Indeed, the only management question that never seems to be answered is: if all management problems can truly be solved simply by reading a few books - then why

[1] http://www.google.com/finance?cid=13257409
[2] http://www.tompeters.com/printer_friendly.php?note=bio/bio.php
[3] http://www.forbes.com/sites/erikaandersen/2012/01/11/tom-peters-still-rocks/

hasn't every manager in the world done exactly that? The entire raison d'etre of the guru industry therefore rests upon persistent managerial myopia, or perhaps desperation for some "scientific" guidance.

Yet if the popular-press gurus represent one key source of management theories today, then surely the other source is comprised by no less than the faculty of the world's eminent business schools. In fact, development of management theory over the last generation has become the dominant raison d'etre for faculty at the top business schools. The unquestionably dominant in much of modern-day B-school faculty hiring and promotion process is publication within a select group of prestigious, peer-reviewed academic journals. Such journals place a primacy upon theory development. Indeed, one such top academic journal – the *Academy of Management Review* - publishes only purely theoretical papers as a matter of formal policy, with papers containing actual empirical data being expressly forbidden. Those other journals that do accept empirical evidence invariably <u>also</u> demand theoretical contributions to be paired with the empirical evidence that (almost always) support the new theory. As former President of the Academy of Management Donald Hambrick observed: "The gatekeepers for the top journals in management first screen manuscripts for basic readability and technical adequacy, and then they apply one pivotal test, above all others: Where's the theory?"[4]. Another business academic stated wryly: "new theory development has emerged as the ultimate end."[5]

---

[4] Hambrick, D. 2007.
[5] McKinley 2010.

Given the emphasis on management theory by both popular-press gurus and business school faculty alike - and the resulting avalanche of theories that it has precipitated - it is high time that the quality of those theories be assessed, along with the ancillary philosophical issue of what the purpose of theory is in the first place. How should theories be assessed? How should they progress? What *should* management theories contribute, perhaps as opposed to what the extant management theories *actually* contribute? In short, what is the true purpose of theory?

One highly promising framework for answering these questions was provided by Clayton Christensen and David Sundahl, who developed a schema to categorize and classify managerial theory development. I now turn to this framework to examine what it entails.

1.1 The Christensen & Sundahl 'Theory of Theories'

Christensen and Sundahl propose a 'theory of theories': a meta-theory of the evolution of management theories.[6] They proffer a pyramidal structure that denotes the state of knowledge regarding a particular management phenomenon. Theory induction and theory deduction – the intellectual divide that pervades the management community – are therefore modeled as antiparallel processes that respectively ascend and descend the pyramid of knowledge. The base of the pyramid represents the lack of knowledge and the concomitant collection of data – the steps of observation/description/measurement - to acquire initial knowledge regarding the phenomenon. The data are then classified and

[6] Clayton M. Christensen, David L. Sundahl, "The Process of Theory Building," HBS Working Paper, 02.016.

packaged into *proposed* theory through a process of *inductive corroboration*. The proposed theory is "a statement of what causes what, and why, and under what circumstances". The theory is then used to deductively generate predictions which are then tested upon additional data. Theory that survives repeated deductive testing may eventually ascend to paradigmatic status.

Arguably, the most innovative aspect of the Christensen-Sundahl is its explicit exploitation of *anomalies* within the theory-generation process. Anomalies – defined to be data points that fail to conform to a particular theory – are neither data points to be avoided nor disappointments if encountered. Rather, anomalies are *actively pursued.* Scholars are explicitly tasked with finding 'odd' data-points that fail to conform to extant theory. For example, given the highly popular theory that lean-manufacturing boosts operational performance, might there be firms that would be *hurt* by implementing lean? Given the theory that team diversity fosters innovation, might there be certain types of teams or innovation where diversity *decreases* innovation? The goal is therefore not only to find and test data points that support a particular theory, but also to find those data points that fail to support the theory. Such failures are not disasters but rather opportunities, for they allow the theory-development process to begin anew. Researchers can return to the bottom of the pyramid to collect data regarding the anomalous behavior with the goal of ultimately building another theory that integrates both the old theory and the anomalies. Theory development therefore is modeled as a never-ending cycle of theory development and redevelopment, perpetually renewed by the discovery and integration of anomalies. Given the continuous cycle of theory renewal coupled with the ever-changing nature and complexity of management, the pyramidal apex status of 'paradigm' might likely never be reached, for there will always be new anomalies remaining to explain. Paradigmatic status could therefore be treated as an ideal that researchers would strive to achieve rather than a goal to be accomplished.

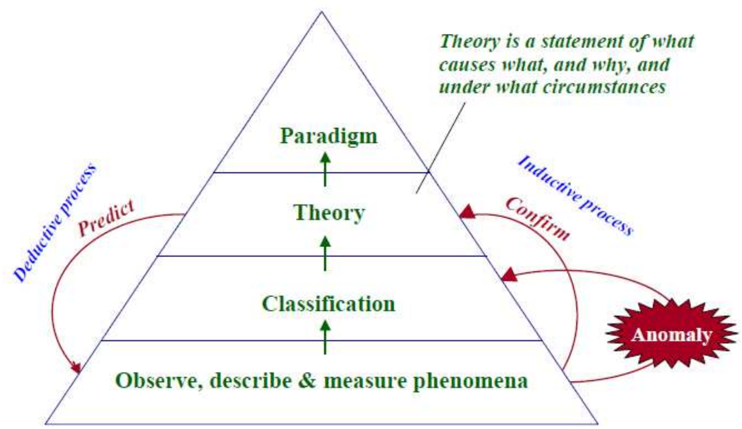


Figure 1

Yet however useful the Christensen-Sundahl model may be in understanding the theory development process in management, several key questions remain.   The process by which data observations/measurements are classified and converted to theory is unstated. What constitutes a confirmed (prospective) theory? How should that theory then be used to deductively predict and test new data? Perhaps most importantly of all, what happens – or more specifically – what *ought* to happen to previous theory once anomalies have been discovered for which new theory can explain?   Such questions have vexed philosophers of science throughout the sweep of history.   I now turn to the following poignant historical vignette that encapsulates epistemological thought regarding how theory is assessed.

1.2 A Vignette on Epistemology and the Purpose of Theory

One crisp autumn day in Cambridge Massachusetts in 1939, two fast friends of the august Harvard Society of Fellows were discussing the finer epistemological details of their disciplines. The first man was Stanislaw Ulam, a Polish-Jew who had narrowly escaped the Nazi onslaught of his homeland. Already a world-famous topologist by 1939 at the tender age of 30, in a few short decades he would originate the Monte Carlo method of mathematical simulation, the concept of nuclear pulse spaceship propulsion, and – perhaps most notoriously of all – co-develop the Teller-Ulam thermonuclear weapons design and thereby be forever dubbed the "Father of the Hydrogen Bomb"[7]. The second man was Paul Samuelson, who would later become the first American to win the Nobel Memorial Prize in Economics for formulating the Neoclassical Synthesis of classical microeconomics with Keynesian macroeconomics, and be dubbed by the New York Times upon his passing as "the foremost academic economist of the 20th century"[8]. On that day in 1939, Ulam challenged Samuelson the following pithy yet deliberately

[7] While the popular press generally dubs Edward Teller as the father of the hydrogen bomb, Nobel Physics Laureate Hans Bethe is quoted in Schweber p. 166 as saying: "After the H-bomb was made, reporters started to call Teller the father of the H-bomb. For the sake of history, I think it is more precise to say that Ulam is the father, because he provided the seed, and Teller is the mother, because he remained with the child. As for me, I guess I am the midwife"
[8] http://www.nytimes.com/2009/12/14/business/economy/14samuelson.html

provocative question that, while conceptually elementary, would bedevil not only Samuelson at the time, but also all social scientists from that day forward:

> **"Name me one proposition in all of the social sciences which is both true and non-trivial**."[9]

Samuelson notably had no response to Ulam's request until 1969 – a full three decades after the question was asked and the year in which Samuelson won the Nobel Prize in Economics – when he replied that Ricardo's Theory of Comparative Advantage, the ideological and theoretical foundation of free trade, fulfills both of Ulam's stipulations. [10] Indeed, free trade has been noted as arguably the only non-obvious policy upon which all economists can agree.[11]

What Ulam captured in his question was the distilled essence of the 'Demarcation Problem' commonly identified by epistemologists to distinguish valuable empirical theories. Similar sentiments are periodically expressed by those hoping to define theory. The commentator Jim Manzi similarly pronounced that: "*The purpose…is to create useful, <u>reliable</u> and <u>non-obvious</u> rules that allow us to <u>predict</u> the effects of potential interventions".* [12]Biologist Eric Lander proclaimed: "You only believe theories when they make non-obvious predictions that are confirmed." The noted philosopher Imre

---

[9] Samuelson, Paul (1969), "The Way of an Economist", in Samuelson, P. A., *International Economic Relations: Proceedings of the Third Congress of the International Economic Association*, London: Macmillan, pp. 1–11

[10] See Chang, Ha-Joon. Kicking Away the Ladder: Development Strategy in Historical Perspective. London: Anthem Press, 2002. It should also be noted that some economists maintain that Samuelson's answer fails to address Ulam's question because comparative advantage is a mathematical identity about what ought to happen rather than an empirically validated proposition.

[11] http://www.nytimes.com/1993/09/17/us/a-primer-why-economists-favor-free-trade-agreement.html?pagewanted=all&src=pm

[12] http://theamericanscene.com/2011/03/19/jim-and-noah-s-excellent-adventure-part-1

Lakatos dichotomized 'research programs' as being either progressive or degenerate. Progressive programs consistently and systematically propose "stunning…hitherto unknown novel facts" that are confirmed to be true. [13]Degenerate programs in stark contrast make predictions that are not novel, not true, or never make any predictions at all but produce only post-hoc explanations of known facts,[14] which one prominent scholar dismissed as amounting to little more than "journalism with regressions"[15]. Karl Popper similarly proposed the notion of the 'risky but valid prediction' in determining a theory's value.

> *"Confirmations should count only if they are the result of risky predictions; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory — an event which would have refuted the theory…"* [16]

Paul Meehl likewise argued at length in his classic 1978 paper:

> *"A theory is corroborated to the extent that we have subjected it to such risky tests; the more dangerous tests it has survived, the better corroborated it is. If I tell you that Meehl's theory of climate predicts that it will rain sometime next April, and this turns out to be the case, you will not be much impressed with my "predictive success." Nor will you be impressed if I predict more rain in April than in May, even showing three asterisks (for p < .001) in my t-test table! If I predict from my theory that it will rain on 7 of the 30 days of April, and it rains on exactly 7, you might perk up your ears a bit, but still you would be*

[13] http://www.lse.ac.uk/philosophy/about/lakatos/scienceandpseudosciencetranscript.aspx
[14] http://www2.lse.ac.uk/philosophy/About/lakatos/scienceAndPseudoscienceTranscript.aspx
[15] Davis, G & C. Marquis. 2005. Prospects for Organizational Theory. Organization Science. 16:4 332-343.
[16] Karl Popper, *Conjectures and Refutations*, London: Routledge and Keagan Paul, 1963, pp. 33-39; from Theodore Schick, ed., *Readings in the Philosophy of Science*, Mountain View, CA: Mayfield Publishing Company, 2000, pp. 9-13

> *inclined to think of this as a "lucky coincidence." But suppose that I specify which 7 days in April it will rain and ring the bell; then you will start getting seriously interested in Meehl's meteorological conjectures. Finally, if I tell you that on April 4th it will rain 1.7 inches (.66 cm), and on April 9th, 2.3 inches (.90 cm) and so forth, and get seven of these correct within reasonable tolerance, you will begin to think that Meehl's theory must have a lot going for it."* [17]

The upshot is that, as a consensus opinion of philosophers of science, a theory is an intellectual construction that produces one or more empirically reliable, non-obvious predictions. To assess a theory is therefore to assess whether it in fact produces such predictions. I therefore propose a framework with which such theories can be validation.

1.3 The Enhanced Christensen-Sundahl Model:

Given the sentiments expressed in the previous section by the gamut of philosophers of science, we now have criteria to assess both the inductive theory building and deductive theory testing procedures of the Christensen-Sundahl model. The process of data observation, description, and measurement – including the collection and evaluation of anomalies – ultimately serves the purpose of inducing theoretical statements that necessarily possess the following two qualities:

(1) They must be *empirically testable*

[17] Meehl, P. 1978. Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology, Journal of Consulting and Clinical Psychology.

(2) They must produce *non-trivial predictions* : the more non-trivial the predictions are, the more worthy the theory becomes (assuming those risky predictions are validated)

The deductive theory prediction step encompasses the following third quality:

(3) The predictions of the theory must be empirically validated.

However, as stated previously, any lack of empirical validation is not a failure by any means, but rather presents the opportunity to develop new theory by virtue of an anomaly. While certain empirical validation failures may be dismissed as methodological issues such as measurement errors or statistical identification difficulties, data points that reliably and consistently defy validation by the existing theory illustrate the shortcomings of the extant theory and the consequent new phenomenon that the new theory must encompass.

A progressive academic discipline must therefore periodically generate *theoretical ejecta*: theories that were formerly believed, but which the scholarly community now considers to be superceded by newer theories as demanded by anomalies. A short synopsis of such ejecta from astronomy would include the geocentric model of the solar system (where the Earth served as the immovable central rotational axis of the universe) which was ejected in favor of the heliocentric model (where the Sun served as immovable central rotational axis of the universe) which in turn was ejected in favor of our modern view of the universe that has no central rotational axis at all and where the Sun is simply one of an innumerable quantity of stars. Similarly, physics has ejected Aristotelian Physics in favor of Newtonian mechanics, which in turn has been ejected, at least in principle, and replaced by quantum mechanics and relativity (although it should be noted that Newtonian physics is still widely applicable as a first approximation). Quantum mechanics and relativity are both broadly expected by the physics community to eventually be ejected in favor of a new theory that integrates both theories. [18]

I therefore humbly proffer a modest proposal: a modification to the Christensen-Sundahl model. As before, the bottom two levels of the pyramid comprise the data observation and classification stages, whose purpose is to generate a proposed theory which are then interrogated to ascertain whether they produce empirically testable, nontrivial predictions. Proposed theories that are untestable, or whose testable predictions are trivial should be reformulated until the theories can product testable, non-trivial predictions. Proposed theories that do produce testable, non-trivial predictions are promoted to candidate theory status. Candidate theories then undergo at first an initial validation process, followed by a continual revalidation process to test and retest said non-trivial predictions. Failures to validate non-trivial predictions represent an opportunity to identify an anomaly and subsequent generation of better theory. In principle, a theory that survives numerous repetitions of revalidation may eventually ascend to paradigmatic status in the eyes of the academic community, but as previously

[18] Indeed, certain highly successful specialized theories such as quantum electrodynamics have already integrated prior aspects of quantum mechanics and relativity, implying that those prior aspects have been ejected.

stated, given the diversity and complexity of management phenomenon, paradigmatic status would likely serve only as an ideal. Ejecta of failed theories are therefore continually generated by the theory validation stage. The accumulation of theory ejecta serves as an indicator of a healthy, progressive academic discipline.

To be clear, none of this discussion is meant to imply that ejecta are not a *causal* determinant of academic progress, but the natural result of making risky predictions. Obviously researchers could easily accumulate a never-ending stream of ejecta simply by deliberately proposing a litany of unreliable theories that they knew would surely never survive empirical scrutiny. I assume that researchers are legitimately attempting to produce true, non-trivial theories; the presence of ejecta then serves as an indicator that researchers are discarding failed theories in favor of better ones.

My modified Christensen-Sundahl pyramid is presented below. Each of the modifications comprises a chapter topic.

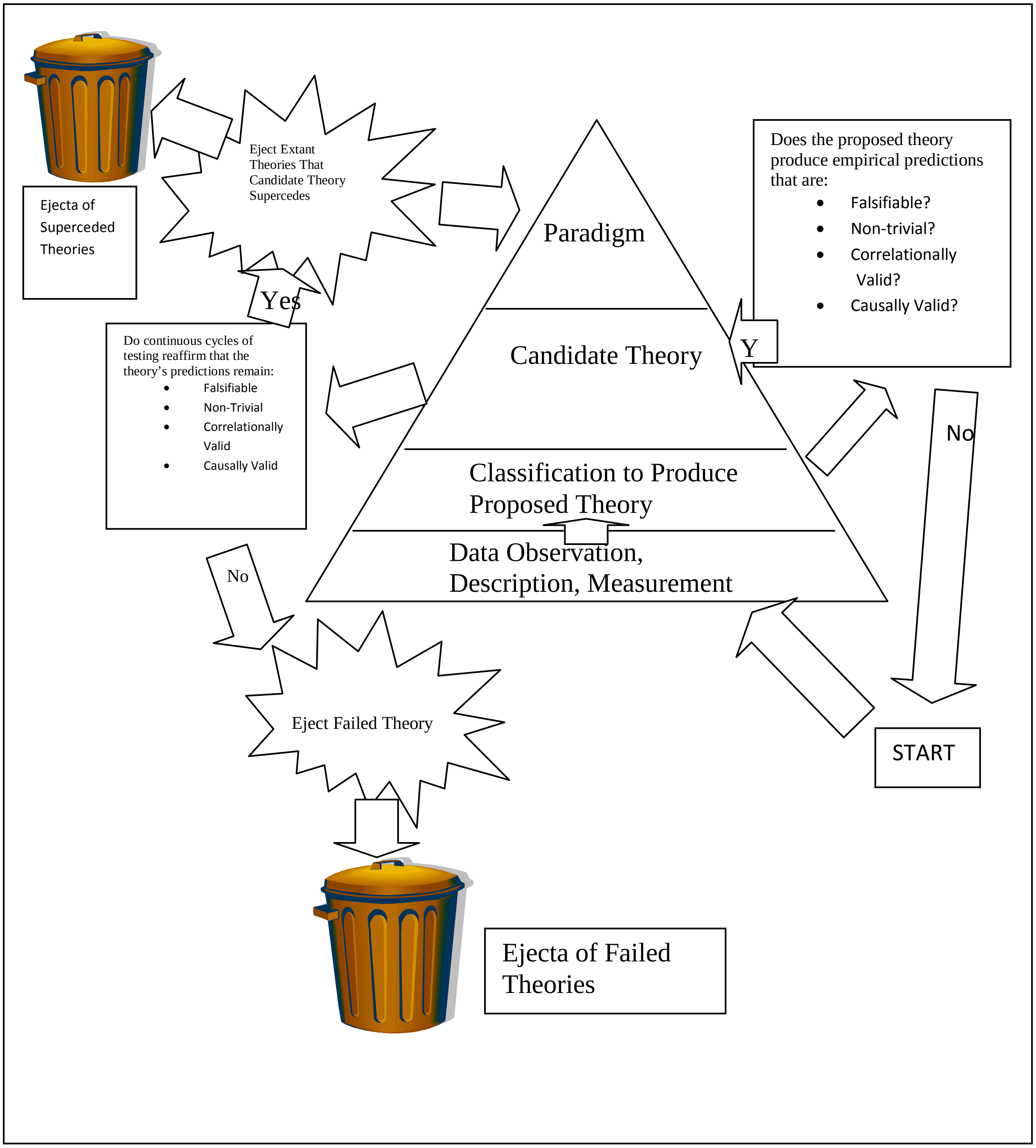


Figure 2

1.4 Structure of the Dissertation

The dissertation therefore proceeds thusly:

- **Chapter 2: Empirical Testability and Falsifiability:** Valid theories must provide propositions that run the bona fide risk of falsification. But are management theories legitimately falsifiable? I investigate the falsifiability of some of the most influential management theories in history.

- **Chapter 3- Non-triviality**: The entire premise behind any theory is that it produces surprising (yet empirically reliable) propositions. Management propositions are examined for their non-triviality. I examine a random sample of recent predictions within both the practitioners' and academic literature along with their accompanying theory, examined for their non-triviality.

- **Chapter 4 - Empirical Testing & Prediction Validation**: Predictions of theories must be empirically validated before they can be considered to be reliable. Propositions must be developed *before* the data are collected and analyzed; I discuss why this step is necessary and how it is often times subtly violated.

- **Chapter 5 – Causality:** Mere empirical correlations are insufficient for management academia to become the truly useful *professional* discipline that it claims to be. What practicing managers want to know is whether certain correlations have *causal* interpretations, and those causal interpretations are what management researchers should aim to supply. I discuss why this is so and how causality is established.

- **Chapter 6 Continuous Testing –** Given that empirical results should be viewed as tentative, they therefore should be subjected to replicated revalidations, with retractions occurring as necessary. Revalidations should occur not only with the discovery of new datasets, but also with new theoretical and methodological advances. A key sign of scientific progress is that theories are ejected when found retesting finds them wanting. I discuss the lack of such ejection within management academia and its implications.

- **Chapter 7 - Have Any Management Theories Become Paradigms? -** Perhaps certain management theories do indeed survive a repeated sequence of retesting. One might imagine that that would imply that such a theory would be promoted to paradigmatic status. Unfortunately, given the pervasively mutually conflicting tenets of the management theory landscape, the promotion of one theory to paradigmatic status necessarily implies ejecting a conflicting theory. Yet again, that ejection apparently never seems to occur. I discuss one particularly enlightening case study of theoretical conflict – that between Transaction Cost Economics and Resource Dependence Theory – and show that despite mutually exclusive predictions, neither theory has instigated the ejection of the other. I also show that this is typical of a field such as management where scholars simply cannot or will not agree upon a criteria of ejection, and that epistemological dissensus is the true cause of lack of field progression.

- **Chapter 8 – What is to be Done? Conclusion and Final Thoughts**

  The aforementioned myriad epistemological challenges that stymie progress in management academia have understandably spurred a plethora of potential reform proposals. I review a few popular ones – the revival of purely qualitative research, the reinstitution of practitioners as faculty, and the launch of 'dedicated and insulated research cadres'. I conclude that while such reforms may indeed generate more and better theories, none of them are likely to succeed in achieving the true marker of progress: convincing the community to eject obsolete theories. Indeed, such reforms are likely to exacerbate the theoretical clutter by adding yet more theories to the landscape that will never be ejected.

  Rather, what may ultimately generate ejection and resultant theoretical progress is the enforcement of the principles of the scientific method as laid out in the previous chapters. I propose a suite of reforms that the academic community – likely through journal editors as the instrument – can enforce a set of rules and challenges to extant theories that may ultimately cause theories to be ejected.

  However, I recognize all too well that such reforms are unlikely to be enacted anytime soon, for the academic community at this time, frankly, has little reason to reform itself. That then raises the natural question of, given the seemingly perpetual dissensus within the theoretical landscape, how the academic community can fulfill its mandate of delivering reliable, non-obvious, consensus advice to practitioners. One answer to that question is that the <u>dissensus itself</u> may be the practical deliverable that academia can provide, such that the community can and should publicly challenge the litany of gurus and consultants who claim to offer easy, sweeping answers. A second answer is that <u>methodology</u> may well be a practical <u>pedagogical</u> deliverable that academia can provide to practitioners. While academia may not be able to provide students with rigorous answers, academia can at least students how to rigorously discover their own answers, or at least to critically assess the easy answers supplied by others.

2 Empirical Testability and Falsifiability

For a proposed theory to be promoted to candidacy for subsequent empirical testing, the theory obviously must provide empirically testable predictions in the first place. To invoke the syntax of Karl Popper, such theories must be empirically falsifiable (a point to which I shall return later). To clarify, the term ‘falsifiable’ must be distinguished from the term ‘false’; the former *prospectively* determines that a theory *could be found* to be empirically invalid, whereas the latter term *retrospectively* has determined that a theory *has been found* empirically invalid.

We shall discuss the underlined line in the Christensen-Sundahl pyramid.

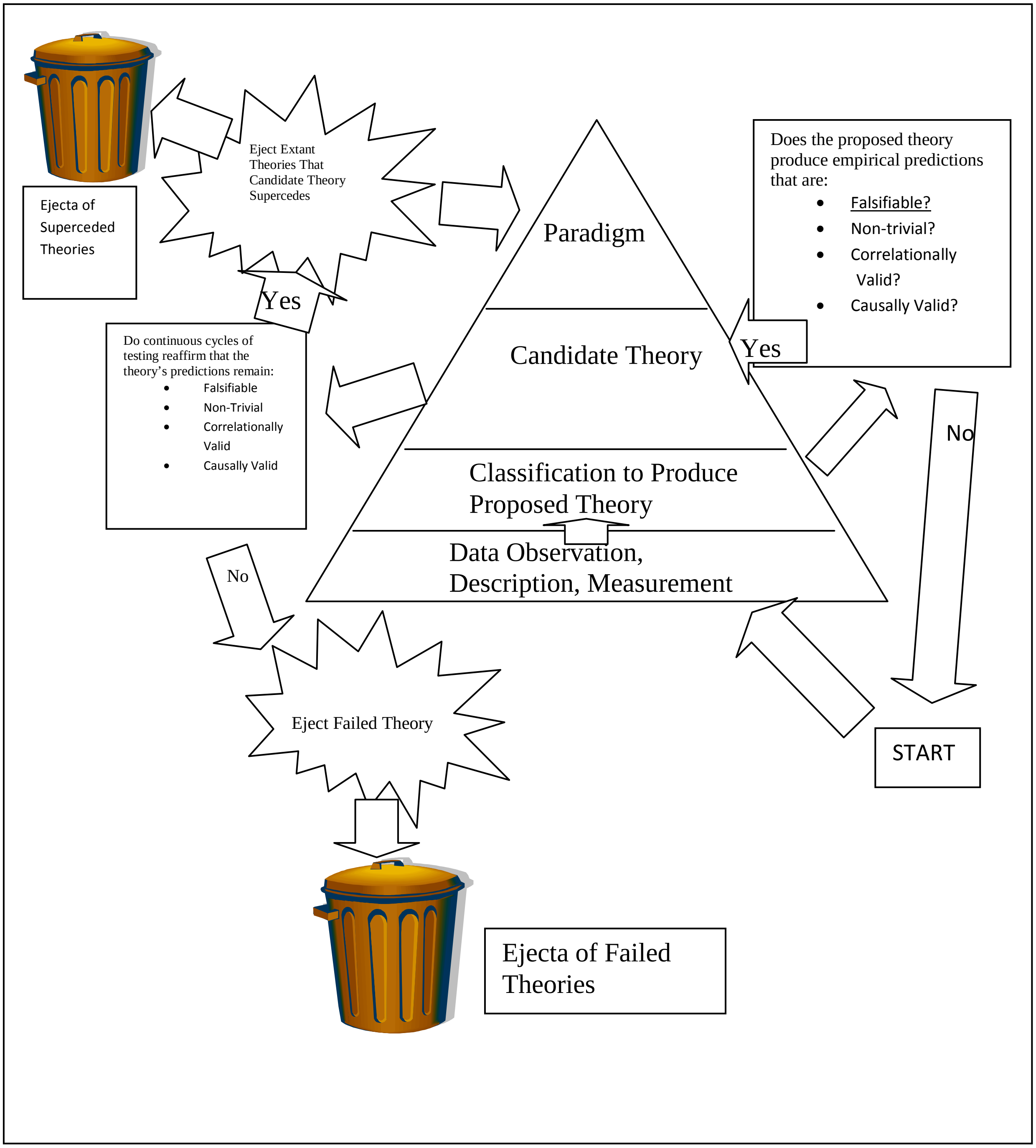

Ejecta of
Superceded
Theories
Eject Extant
Theories That
Candidate Theory
Supercedes
Yes
Paradigm
Candidate Theory
Classification to Produce
Proposed Theory
Data Observation,
Description, Measurement
Does the proposed theory
produce empirical predictions
that are:
Falsifiable?
Non-trivial?
Correlationally
Valid?
Causally Valid?
Yes
No
START
Do continuous cycles of
testing reaffirm that the
theory's predictions remain:
Falsifiable
Non-Trivial
Correlationally
Valid
Causally Valid
No
Eject Failed Theory
Ejecta of Failed
Theories

Figure 3

As a pedagogical device, I propose a simple 2x2 grid framework that assesses the empirical testability of a proposed management theory. I then demonstrate how some of the most prominent management theories in history fail to conform to the grid framework.

2.1 The 'Grid Framework of Empirical Validation'

Consider a proposed correlation between handsomeness of your salesmen and their individual sales revenue. Or the correlation between happiness of your employees and profits of the firm. Or the size of the CEO's office and improved stock price performance. Each of these relationships can be abstracted as an elementary bivariate relationship between X and Y. Indeed, almost any proposed relationship amongst social science phenomena can be modeled via a set of similar elementary bivariate relationships.[19] [20] However, whether such propositions are <u>valid</u> theories is another

[19] For example, moderation and mediation can itself be modeled in this fashion by assigning X to be the moderator/mediator construct and Y to be the strength of the moderation/mediation in question. Nonlinear relationships can be modeled as a (potentially large) sequence of relationships between two variables. Classifiers could be modeled as a simple Success/Failure classification relationship.

[20] Note that the X/Y bivariate relationship does not rule out relationships with other variables. In other words, Y may well be correlated with other variables; changes in X are not necessary for changes in Y. However, ceteris paribus, changes in X will be associated with changes in Y.

matter. Each proposition can be assessed by the following simple, 2x2, 4-cell grid framework.

Figure 4 - The Grid Framework of Mgmt Theory Validity

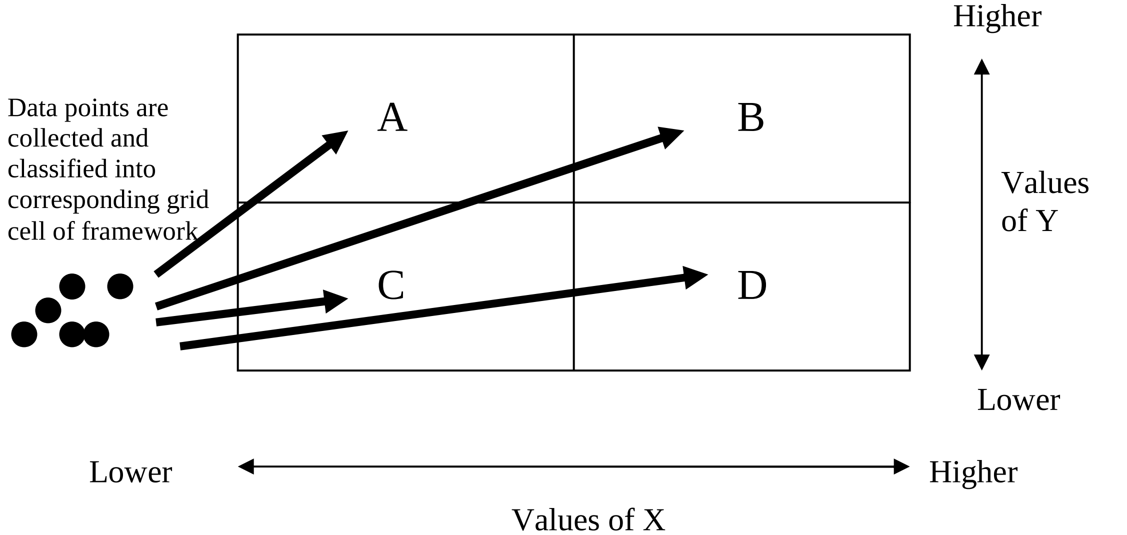


Figure 4

If X and Y are proposed to be positively correlated[21] (the default assumption that I make throughout this text), then we are proposing that data should predominantly lie within cells B and C. However, note that cells A and D serve a far more important purpose beyond simply serving as a catch-basin for stochastic outliers. They also denote disconfirmations of the theory that could in principle be found. Therein lies the crux of the difference between theories that are false vs. theories that are unfalsifiable. An

[21] An inverse relationship can be modeled by simply switching the signs of one of the variables

unfalsifiable proposed theory of the positive correlation between X and Y must be subject to the risk that data could legitimately crop up in cells A and D. Theories that are immune, whether definitionally or empirically, from such risk are *unfalsifiable*. As trivial as this may seem, some of the most heavily cited management theories are in fact unfalsifiable and are therefore invalid theories – a point we shall revisit shortly.

The Grid Framework and the Falsification Test
Both Cells A and D Must be Feasible

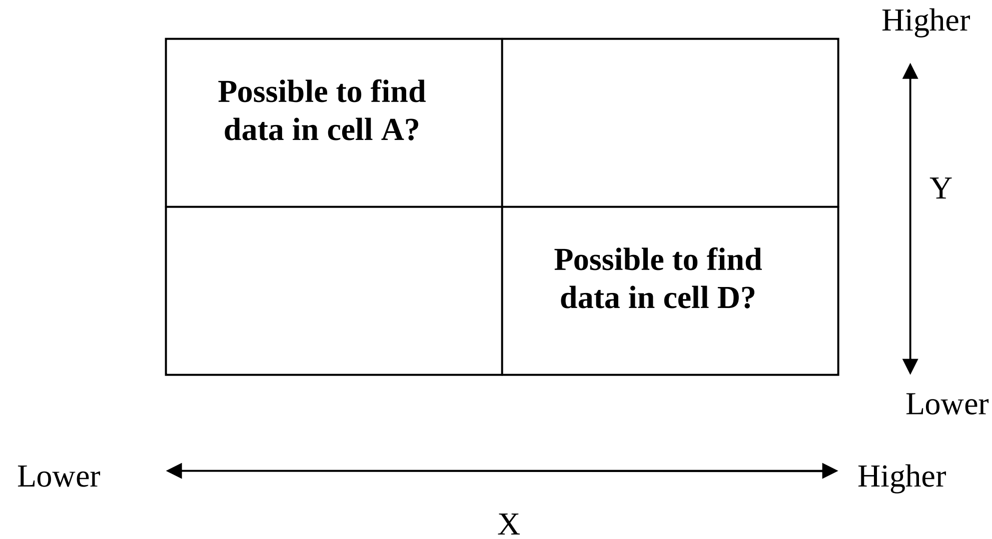


Figure 5

If the relationship between X and Y is fully deterministic, then the data should lie *only* within cells B and C (with cells A/D serving the purpose of checking for falsifiability). However, the stochastic nature of the business world seldom allows us to propose statements with such certainty. Nevertheless, the bulk of the datapoints ought to be found to lie within cells B and C, with only statistical outliers found within A and D. These scenarios are shown in the following figures, where filled circles indicate data that are tested upon pre-established hypotheses, with the size of the circle indicating the proportion of data found within the cell.

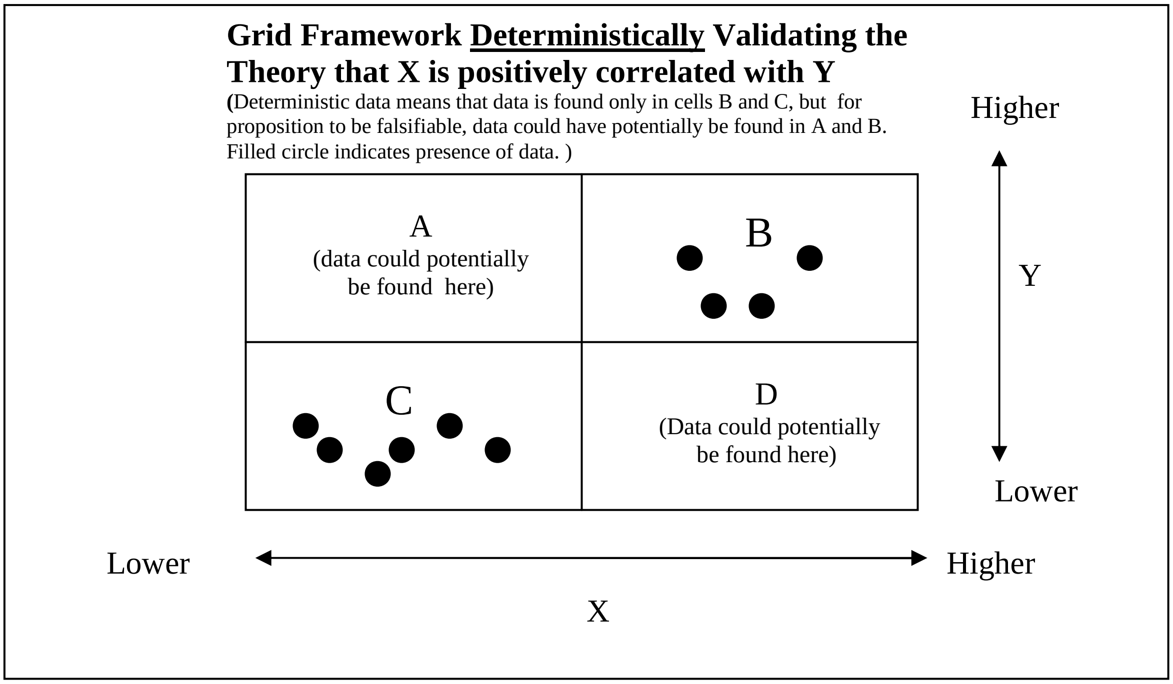


Figure 6

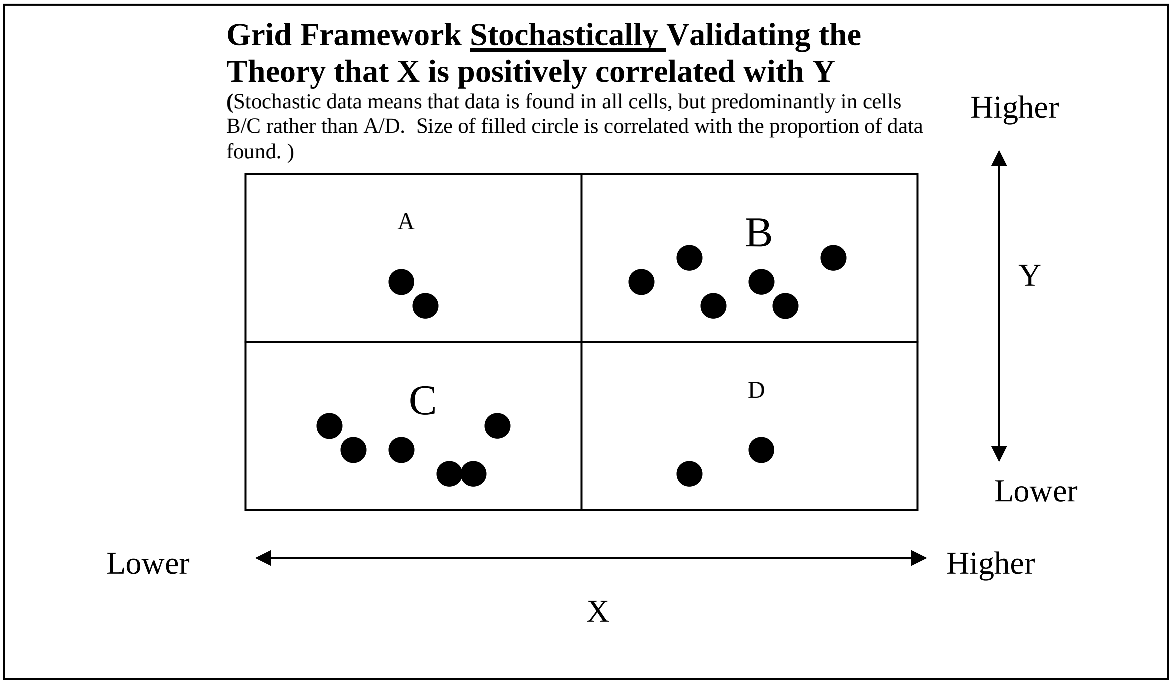


Figure 7

With the 2x2 assessment tool in hand, let us turn our attention to one of the most influential academic management theories in history, the Resource Based View Theory,

2.2 The Curious Case of the Resource Based View (RBV) Theory

As "the leading theory of competitive advantage"[22] and arguably the most heavily cited theory in the entire subfield of strategic management, the resource-based view (RBV) theory offers the promise of explaining why certain firms seem to wield long-term competitive advantage over others despite not possessing monopoly power or obvious external barriers to entry.[23]

[22] Powell 2001.
[23] Such external barriers to entry tend to be the focus of the Michael Porter school of strategy.

RBV encourages firms to search within themselves to discover valuable ‘resources’ that others do not possess. Such resources could be employee skillsets, managerial insights, organizational learning capabilities, advantageous social networks, or other such attributes.

The central theoretical prediction made by RBV is that resources generate competitive advantage. To understand this prediction, the definitions of the terms ‘resources’ and ‘competitive advantage’ must therefore be unpacked and defined[24]..

2.2.1 How does RBV Define ‘Resources’?

- Resources are “all assets, capabilities, organizational processes, firm attributes, information, knowledge, etc. controlled by a firm that enable the firm to conceive of and implement strategies that improve its efficiency and effectiveness”[25], or "firm attributes that may enable firms to conceive of and implement value-creating strategies”.

- Resources are “valuable” to the extent that they “enable a firm to conceive of or implement strategies that improve its efficiency and effectiveness" & “when they exploit opportunities or neutralize threats in a firm's environment"[26]

2.2.2 How does RBV Define ‘Competitive Advantage’?

[24] Note, rather than imputing my own interpretation regarding the definitions of the RBV and thereby imposing potential skew upon those definitions, I invoke the original quotes from the academic papers that represent the foundation of RBV.
[25] This definition is from Daft 1983, and is cited by Jay Barney, a leading proponent of the RBV school, in his 1991 paper: the most heavily cited paper of the RBV school.
[26]. Priem & Butler 2001.

- Competitive Advantage (CA) is defined as "implementing a value creating strategy not simultaneously being implemented by any current or potential competitors"[27]

RBV also theorizes that certain types of resources provide the strongest, longest-lasting, and therefore greatest value, relative to other resources. I therefore note the following additional sub-prediction.

2.2.3 Which Resources Provide the Strongest Competitive Advantage?

- The most valuable resources are characterized by "intangibility, invisibility, complexity, causal ambiguity" and otherwise "difficult-to-specify interactions among complex, technological and behavioral variables" [28] The foundational idea is that competitors will quickly imitate valuable resources by developing or purchasing their own, so it is precisely those resources that are difficult to quantify that convey the most long-term competitive advantage.

2.3 Tautology of RBV

The central prediction of RBV theory therefore is, via amalgamating the points above, that firms with unique valuable resources that other firms do not possess will enjoy

[27] Taken from Priem & Butler 2001, from the direction citation of Barney 1991.
[28] See Powell 2001 for a philosophical consideration of the role of intangibility in relation to RBV. Such considerations also form the core of Diericks & Cool 1989.

competitive advantage. But that is precisely where tautological trouble rears its ugly head. If valuable resources are defined to be assets that provide value as per point (1) above, and competitive advantage is a value-creating strategy as per point (2), then that means that the notion of resources is linked by definition to competitive advantage through the shared construct of 'value'. Put another way, CA and resources are overlapping terms. Resources and competitive advantage simply entail each other in the formal sense of entailment. In other words, by definition it is impossible for an asset to be a resource if it does not confer on its firm competitive advantage. If not, then the asset in question must therefore not be a resource, or not unique, or both. **So my challenge is: name me a resource that is valuable but that does not provide competitive advantage. Is such a concept even possible? Is RBV therefore a falsifiable theory?**

The unfalsifiability of RBV is demonstrated vividly through use of the 2x2 grid as shown below. Consider the thought experiment of how an empirical researcher would attempt to investigate RBV's central prediction regarding the connection between resources and CA. Such a researcher would presumably obtain a dataset where some firms demonstrated CA over others. He would then, either through qualitative case studies, or econometric methods, attempt to link the construct of CA to resources. He would then likely find that firms with CA possess such resources, and conversely that firms without CA lack such resources. He might conceivably even find a few firms that have resources but nevertheless lack CA, perhaps because they suffer from other organizational weaknesses that squander the advantage conferred by their resources. Cells B&C are therefore populated, and D is plausible and indeed, perhaps possibly populated with a few data.

But the key question is: could we ever fill Cell A? Honestly, what would happen if he found firms who do have CA but who have no valuable, unique resources? Remember, we are talking about long-term competitive advantage rather than any temporary or murky advantage that might be dismissed as ephemeral statistical noise. How would the RBV researcher react? Likely, his reaction would be that those resources must exist somewhere in the firm[29], and that the researcher simply couldn't find them, but by definition, they must still exist. The possibility of a firm with CA not possessing valuable, unique resources is entirely precluded within the unfalsifiable tautology of the RBV theory.

[29] Powell made a similar point in his 2001 paper.

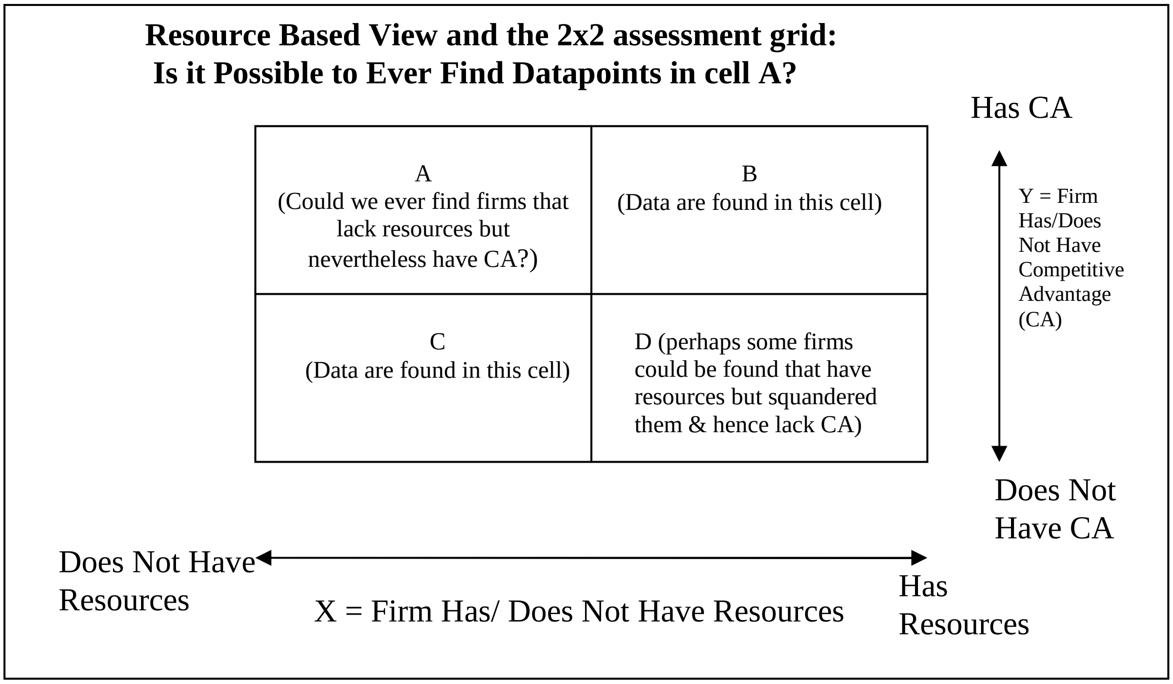


Figure 8

But it gets worse. If after intense combing through the data, the researcher still cannot locate any unique, valuable resources possessed by a firm with CA, what is that researcher likely to do? A savvy RBV researcher would likely resort to invoking point (3) above: <u>that the resource must be characterized by "intangibility, invisibility, complexity, causal ambiguity" or otherwise "difficult-to-specify interactions</u>". In other words, not only would the very inability to measure a resource be interpreted as evidence of the existence of a resource, but as (3) states, it is precisely those resources that are difficult to measure that are the most valuable![30] Point (3) therefore serves as the researchers' proverbial 'Get Out of Jail Free Card". They can always dismiss the inability to find a resource as evidence that the resource must be 'intangible' or 'difficult to specify'. The

[30] Powell remarks that in this way does RBV render itself "refutation-proof".

RBV school perfectly insulates itself from falsification. As shown in the below figure, any datapoints that seem to belong to cell A by definition must actually belong to cell B.

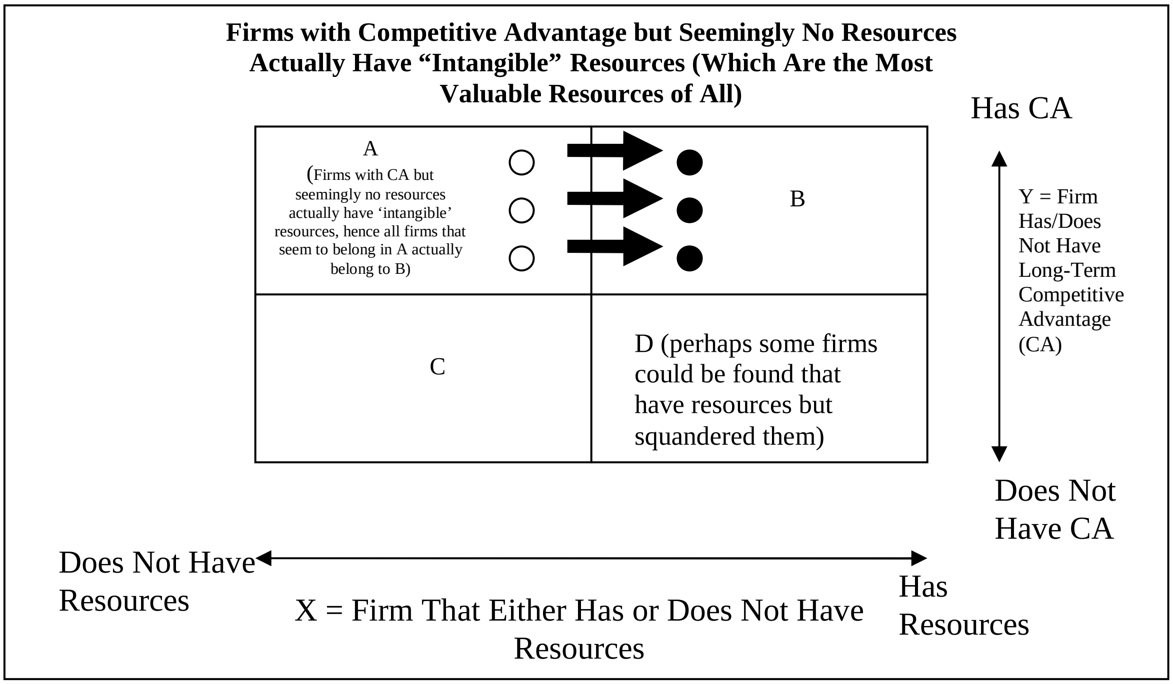


Figure 9

By definition, the tenets of RBV are unfalsifiable, and the proponents of RBV therefore can never be wrong. More fairly, RBV should not be classified as a theory at all, for Cell A of the Validation Grid is not logically possible. Recall that both off-diagonal cells (both A and D) of the validation grid must be possible for a theory to be valid. Yet under the definitions of RBV a firm that has no resources yet nevertheless enjoys CA is a contradiction in terms.

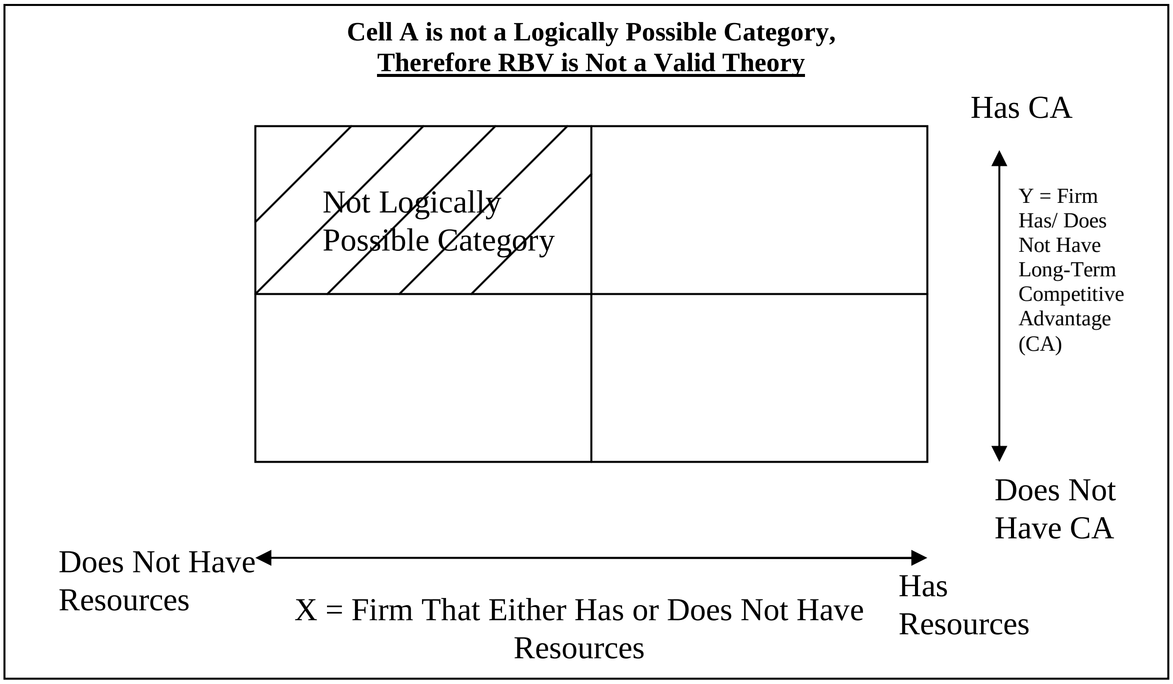


Figure 10

The upshot is that the Resource Based View, the heart of the subfield of academic management strategy, is a fundamentally unfalsifiable and therefore untestable theory, true only by virtue of construction. No data that could ever possibly be discovered would prove it false; which then raises the question of why so many empirical RBV papers – of which there are likely thousands at the time of this writing - have been published to validate a theory that could never possibly be wrong.

To be fair, RBV is not the only popular unfalsifiable management theory. The theory of 'core competency' – a favorite amongst practitioners – is likewise unfalsifiable. That should not be surprising given that core competency seems to be little more than rebranded RBV for the practitioners' set. Let us now turn our attention to the theory of core competency through the lens of the 2x2 grid falsifiability framework where I demonstrate that the theory of core competency suffers from the same logical problems as does RBV.

2.4 The Theory of Core Competency: RBV, Practitioner-Style

Consider the recent example of the dotcom boom of the late 90s, now perhaps now little more than a bitter faded memory for the grizzled tech cognoscenti: a cautionary tale of business hubris and inanity. But during its heyday, the boom was notable not only for the myriad brand-spanking new startups founded to seize seemingly endless opportunities, but also for established firms creatively repositioning their businesses around their 'core competency' that extended to the dotcom wave. The superhero Plastic-Man never exhibited such impressive self-contortion! IBM asserted that its true core competency was actually providing the overall computing reliability that Internet sites required, not their once-touted, outdated core competency of selling venerable mainframe systems whose architectures stretched back to the 1960s. Oracle Corporation boasted that rather than merely selling database systems, its true business expertise stretched to general data access and management that the dotcom firms needed. Perhaps most brazenly of all, Intel asserted that Intel Online Services, its short-lived website-hosting business, was a natural outgrowth of Intel's decades-long core competency in managing large-scale semiconductor fabrication plants.[31]

[31] This was precisely the pitch sold by the sales representatives of Intel Online Services to this author during the dotcom boom era.

But the ‘core competency’ concept was not merely some marketing spiel to bamboozle unsuspecting customers. It was also meant to convey academic respectability, as defined by progenitors C.K. Prahalad and Gary Hamel:

> "Core competencies are the collective learning in the organization, especially how to coordinate diverse production skills and integrate multiple streams of technologies….   First, a core competence provides potential access to a wide variety of markets. …Second, a core competence should make a significant contribution to the perceived customer benefits of the end product. Clearly, Honda’s engine expertise fills this bill. Finally, a core competence should be difficult for competitors to imitate….” [32]

Unfortunately, Prahalad’s and Hamel’s explanation of core competency seems to confuse more than it illuminates.   How do you know *what* a firm’s “collective learning in the organization” might be, other than by imputing its existence through success of the firm? How do you know whether a particular characteristic of the firm “make[s] a significant contribution to the perceived customer benefits” or whether it is “difficult for competitors to imitate”? [33]

In practice, the ‘core competence’ motif shares the same retrospective plasticity and resultant unfalsifiability as does RBV.   Core competency, in distilled form, seems to translate into the hypothesis: “A firm will succeed in providing a product/service to the extent that it has valuable unique internal expertise delivering that product/service.” Replace the term ‘valuable unique internal expertise’ with the word ‘resource’, replace the term ‘succeed in providing a product/service’ with the term ‘competitive advantage’ and you have RBV.

[32] http://www.economist.com/node/12231124
[33] Ibid

We invoke the validation grid to demonstrate core competency's inherent unfalsifiability. We check the feasibility of both cells A&D. Akin to RBV, it might well be true that a firm with valuable core competencies might nevertheless fail to exploit those competencies to produce a successful product, perhaps because the firm suffers from other weaknesses that hinder the exploitation of its competencies. Such a firm might reasonably be categorized into cell D. But here's the kicker: Might a firm be categorized in cell A in developing a successful product for which it had <u>no</u> core competency? What would happen if we found an example of such a firm? Surely, we would argue that the firm actually possesses heretofore unrecognized 'latent' core competency corresponding to that product after all, for how else would one explain the success of that product?

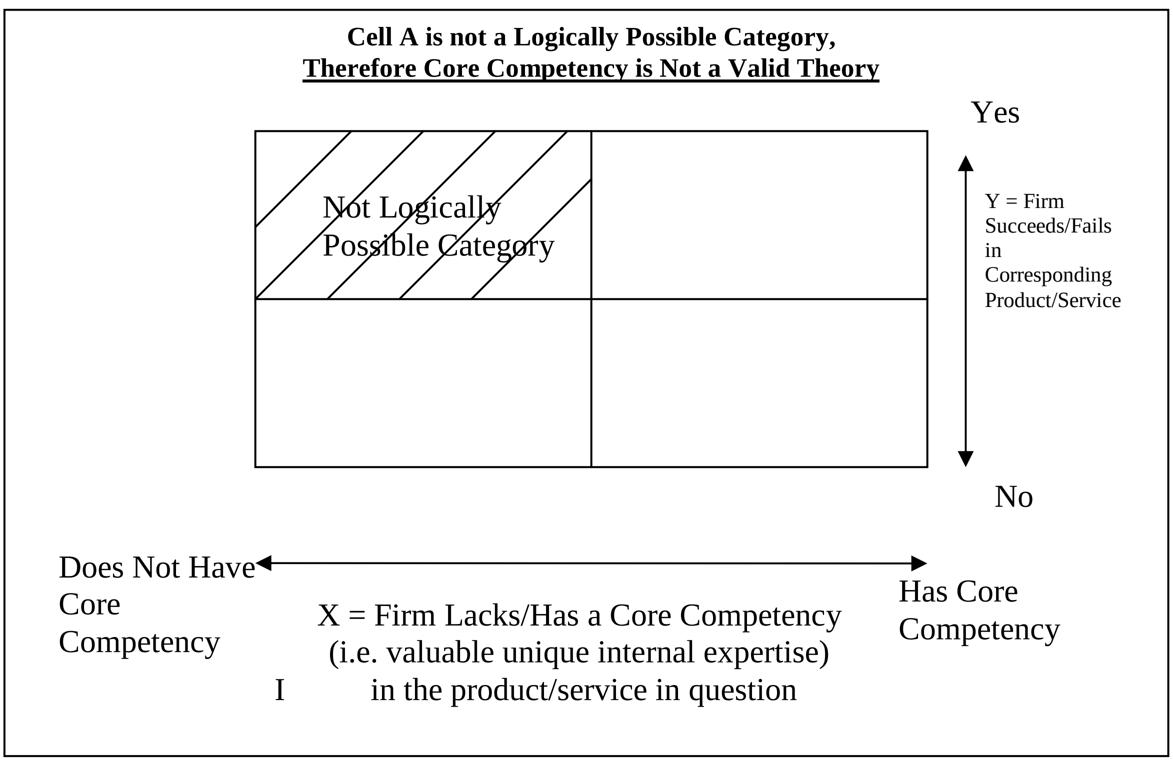


Figure 11

2.5 Principal-Agent Theory and Unfalsifiability

The savvy reader might note that the aforementioned unfalsifiability of RBV and core-competency theory stems from an invocation of an *inferred* characteristic. That is, the presence of either a core competence or a resource is *inferred* from the presence of success or competitive advantage, which irrevocably fuses the definitions of the terms and hence renders them unfalsifiable. A key corollary to unfalsifiability is that certain constructs are difficult to measure and hence require such inferral. The concepts of resources and core competencies are defined as constructs that provide performance advantages, but may otherwise comprise any characteristic of the firm. Ingenious R&D scientists, strong sales/marketing skills, efficient distribution networks – any or all of these traits could be considered resources or core competencies.

This measurability "wiggle-room" built into many other influential theories renders their falsifiability suspect as well. Consider the influential academic management theory known as the 'Principal/Agent Theory' (PA). PA starts with the basic economics-based assumption that everybody's behaviors are governed by a utility function that they are perennially maximizing. PA then holds that the utility functions of a firms' owners - known as *principals* - will differ from or be *misaligned* from the utility functions of the firms' managers, known as *agents*. Principals also cannot perfectly monitor the actions of the agents, so agents enjoy great freedom to behave as they wish. Because the utility functions of agents and principals may differ, and everybody is assumed to always maximize their own utility, agents will inevitably engage in the shirking of duty or other private profiteering that maximizes their utility at the expense of the utility of the principals. Stock-options, performance bonuses, and other such incentive programs are the recommended prescriptions by proponents of PA theory by which the agents' utility functions can be modified to align with those of the principals. For example, because principals are usually stockholders whose utilities increase with stock price appreciation, then compensating the agents through stock will modify the agents' utility functions to conform to that of the principals'. Indeed, much of the recent boom in executive stock-based compensation in the last few decades is attributed to the influence of the Principal-Agent theory[34]

Yet PA suffers from a fundamentally unfalsifiable core because of its reliance upon the concept of utility maximization. Truth be told, utility maximization is a cloaked tautology, as any action that the individual ever chooses to take, by definition, must be directed towards maximizing his utility. As morbidly stated by economist Robert Frank:

> "…Suppose, for example, that we see someone drink a gallon of used crankcase oil and keel over dead. [Utility-maximization] can "explain" this behavior by saying that the person must have really *liked* crankcase oil…" [35]

The underlying problem is that utility – just like 'resources' or 'core competency' - is a fundamentally unmeasurable construct. No tool presently exists with which one can

[34] To quote Frank Vermeulen: "This practice — of offering CEOs stock-based pay — is a recommendation straight out of something called "agency theory." It is one of the few academic theories in management academia that has actually influenced the world of management practice. It is basically a theory, stemming from economics, that says that you have to align the interests of the people managing the firm with the interests of its shareholders. Otherwise, they will only do things that are in their own interest, and will be inactive, lazy, or plain deceitful. Yep, these economists have an uplifting worldview." - http://blogs.hbr.org/vermeulen/2009/04/why-stock-options-are-a-bad-op.html

[35] Frank, Robert H. Microeconomics (3rd Edition).

independently measure somebody's utility. Rather, utility is an *imputed construct,* whose very existence is inferred by the behaviors of which individuals choose to partake. But such inference definitionally fuses together the construct of utility with behavior. Somebody drinking used crankcase oil and dying in agony is necessarily (and unfalsifiably) maximizing his utility.

The unfalsifiability of PA therefore stems from the fundamental fact that no evidence could fall into the A quadrant—nothing would interpreted as violating someone's utility function. Because neither the principals' nor the agents' utility functions can ever be directly measured, any misalignment of those functions can likewise never be directly measured but rather can be inferred only through the <u>differing behavior</u> of the agents relative to that of principals. Yet that differing behavior is precisely what the difference of utilities between principals and agents is attempting to explain in the first place! Put another way, what if certain agents, lacking any incentives that reduce utility function misalignment, nevertheless behaved exactly according to how the principals would wish? For example, let's assume that we found certain CEO's who tirelessly work to boost the company stock price, but without any personal financial incentives to do so. The researcher would then likely conclude that the principal/agents' utility functions must not have been misaligned after all. Using the 2x2 grid framework, could it ever be logically possible for the principals' and agents' utility functions to be misaligned, yet the agents nevertheless dutifully execute the wishes of the principals?

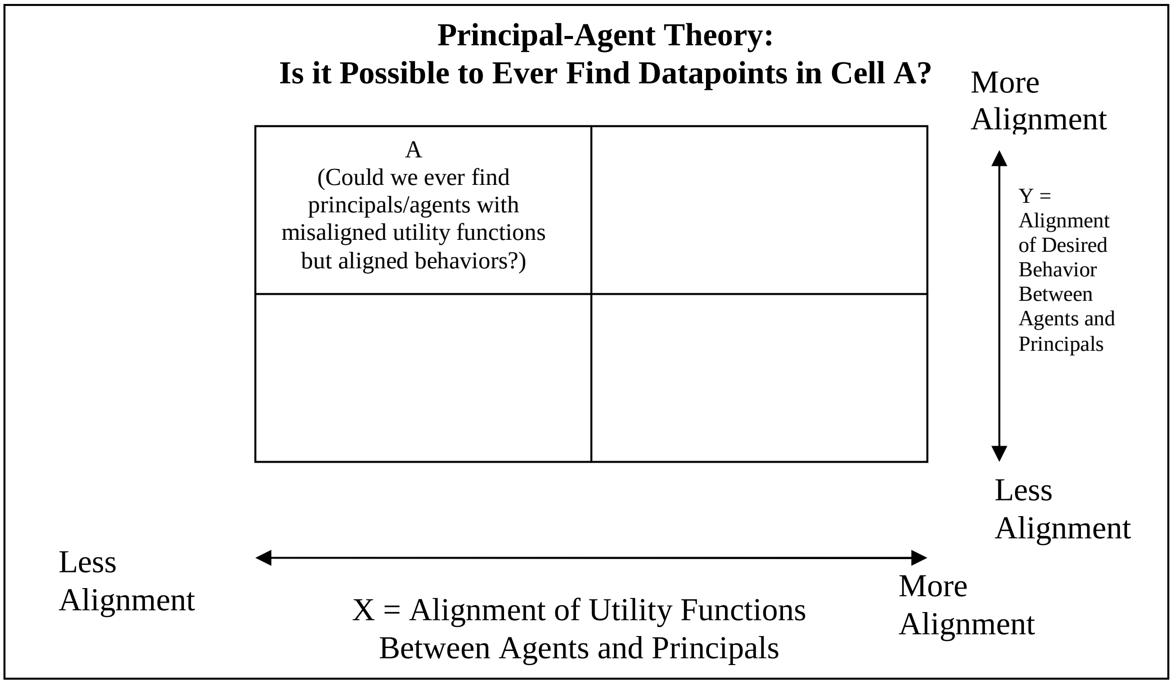


Figure 12

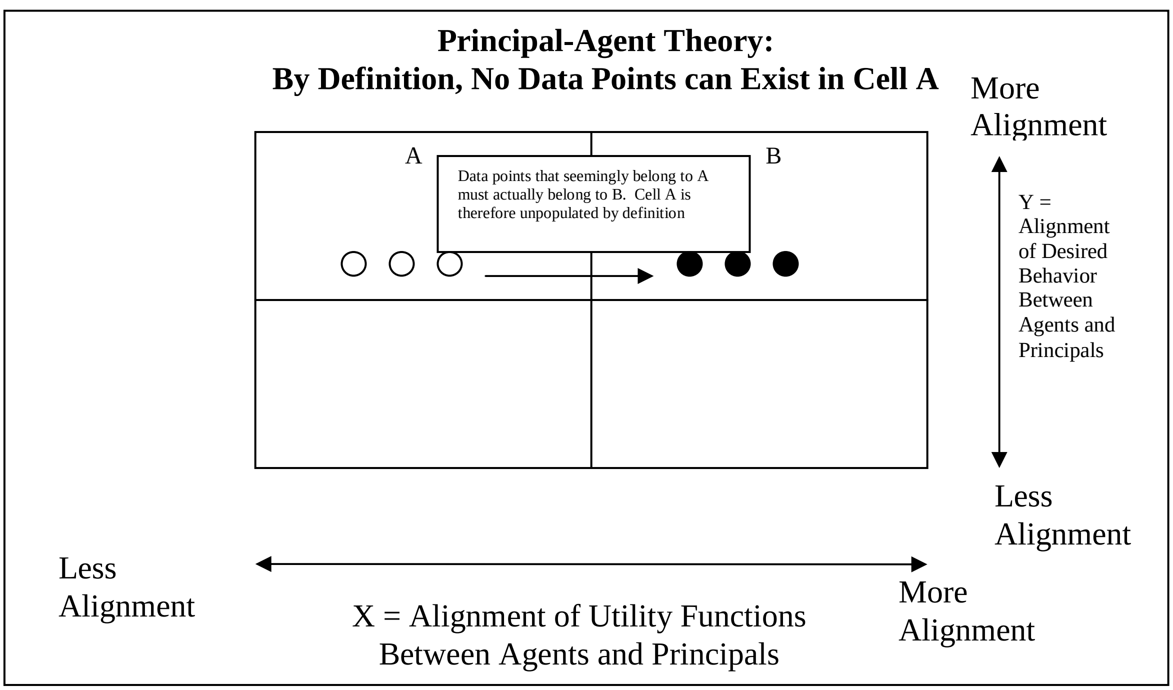


Figure 13

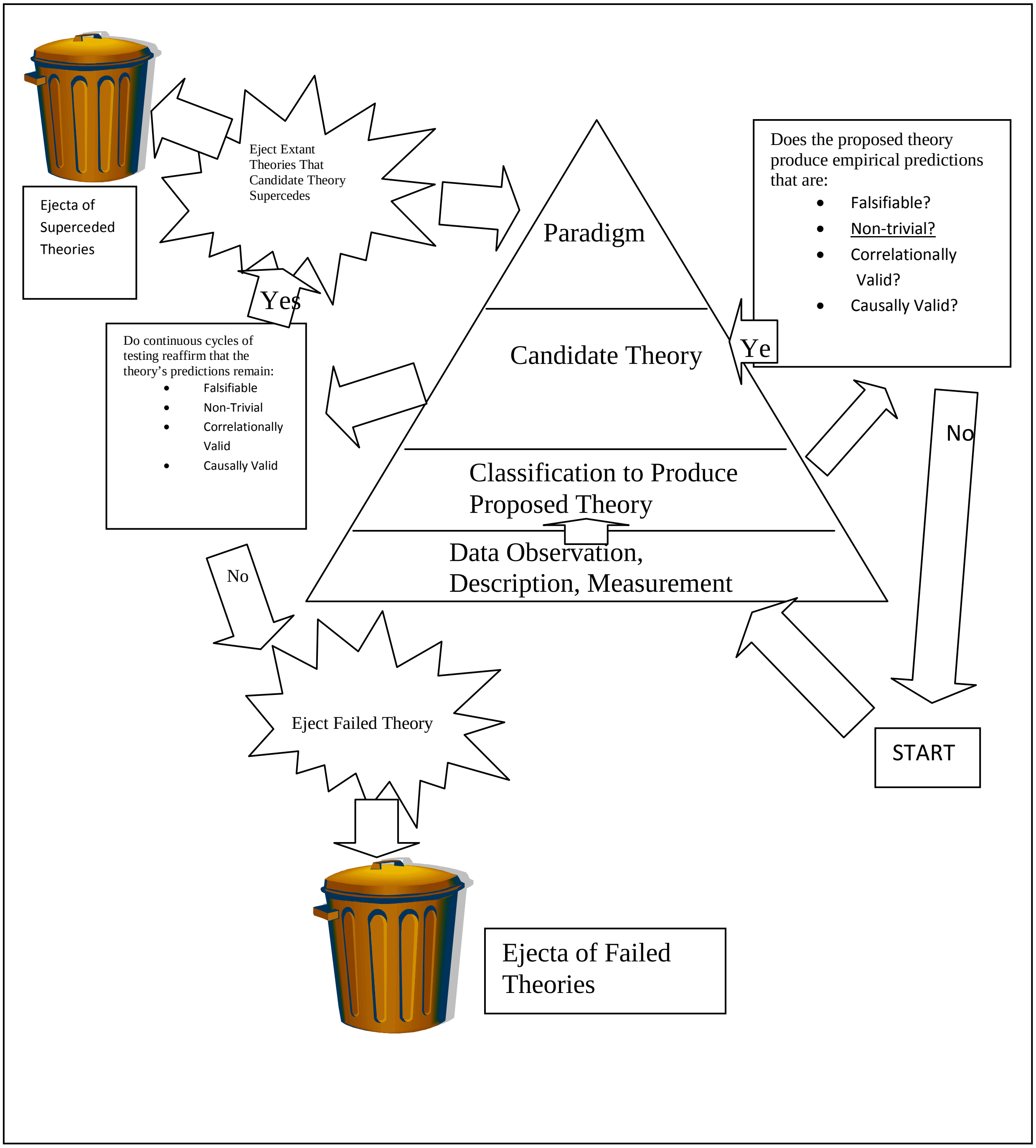
Because the X and Y axes of PA theory are definitionally linked, PA theory cannot offer falsifiable empirical predictions. Just as crank-case oil connoisseurs must have really enjoyed crank-case oil, similarly, agents who behave exactly according to their principals' desires must really have utility functions that are perfectly aligned to their principals' desires. Any purported utility misalignment must not actually exist. PA, despite its popularity, is therefore not empirically falsifiable and hence cannot be accepted as a true candidate theory.

2.6 Conclusions

Proposed theories must produce empirically falsifiable predictions to be promoted to the status of candidacy. I believe that my 2x2 grid framework is an intuitive yet remarkably effective tool for assessing falsifiability. A candidate theory must potentially populate each and every cell of the framework without definitionally excluding any cells. Proposed theories that fail to do so – apparently including some of the most popular and influential management theories in history – must be reformulated until they do produce empirically falsifiable predictions.

3 Non-Triviality

To be sure, falsifiability alone is insufficient to ensure the value of a theory. A theory that predicts cold weather in Siberia next winter is indeed falsifiable – since perhaps the weather might be warm. But it is undoubtedly trivial; nobody would be surprised if the prediction was found to be true. As stated by Popper, theories are valuable to the extent that they produce surprising, bold predictions (that are true). This chapter examines the ability of management theory to proffer such bold predictions. We are at the underlined step of the modified pyramid.



Ejecta of
Superceded
Theories
Eject Extant
Theories That
Candidate Theory
Supercedes
Yes
Paradigm
Candidate Theory
Classification to Produce
Proposed Theory
Data Observation,
Description, Measurement
Does the proposed theory
produce empirical predictions
that are:
Falsifiable?
Non-trivial?
Correlationally
Valid?
Causally Valid?
Ye
No
START
Do continuous cycles of
testing reaffirm that the
theory's predictions remain:
Falsifiable
Non-Trivial
Correlationally
Valid
Causally Valid
No
Eject Failed Theory
Ejecta of Failed
Theories

Figure 14

3.1 Introduction

One currently popular comedic tropes is the mocking of news, particularly the mocking of inane newspaper headlines. The “Headlines” comedic bit by Jay Leno run on Monday nights of The Tonight Show collates some of the sillier headlines that his viewers send to him, and Leno has even sold several books collecting the most infamously banal headlines. A number of comedic websites are likewise dedicated to the collection of such mass media inanity. This comedic trope has also proved popular with non-English audiences who also enjoy mocking the fatuousness of their own news organizations.

So let’s play a game. Consider the following list of statements, and ask yourself to which category they belong:

(A) An obvious news headline pulled from a comedic website

(B) An actual hypothesis published in an actual A-level management paper

The statements are

1. Females likelier to test for women’s disease

2. Stores offering the best bargains are most popular
3. Women are higher in femininity than men
4. Human capital is positively related to performance
5. The passage of prohibition regulation by counties and states will increase brewery failures.
6. Career competencies are positively related to employability
7. A lower average investment analyst recommendation for the company's stock will result in a greater probability that the CEO will be dismissed
8. Close look at dating finds men choose attractive women
9. [Employee] turnover damages performance more when leavers are proficient rather than novice.

The answer is that (1), (2), and (8) are drawn from a popular comedy website.[36] The rest are bona-fide hypotheses proposed within management papers published in top-level journals since 2006. I would venture to say that those hypotheses are indistinguishable in terms of inanity from those of the sillier news headlines that warrant Internet mockery. The academic management literature would therefore seem to offer a comparable vein of rich comedic material. Should we expect a future episode of The Tonight Show with Jay Leno featuring an edition of "Headlines – Management Academia Edition"?

[36] Specifically from 'The 25 Most Obvious Headlines Ever' at http://www.huffingtonpost.com/2012/05/24/most-obvious-headlines-ever-photos_n_1542847.html#slide=974315

3.2 Why is Non-Obviousness Important?

Now, to be clear, the above discussion was not referring to the empirical *results* of management papers. Obvious results may arguably still carry epistemological value. The issue is regarding obvious *hypotheses*¸ which have yet to be validated by data. Recall that one critical element of good theories is to generate non-obvious (yet empirically reliable) hypotheses that *predict* how the world should behave. That then raises the question – what's the point of theory, if doing so only allows you to generate trivial hypotheses? As per Lakatos and Popper, theories demonstrate their value when they allow you to generate non-obvious – indeed, seemingly ridiculous – hypotheses, that are then shown to be empirically correct.

Stephen Cole forcefully discussed the value (or lack thereof) of obvious theorizing in his 1994 paper in Sociological Forum. While his comments were directed specifically at the field of sociology, the same criticism applies to management studies, which is an outgrowth of sociology.

> **"Whether or not this type of criticism [of obviousness] is valid depends on the aim of the research. If the aim of the research is primarily descriptive, then to say that the results are obvious is an illegitimate critique. If we want to know what the facts are -and they turn out to be what we thought they would -then this is useful information. Because as Lazarsfeld (1949) has shown, many "facts" become obvious only after empirical data have shown them to be true. However, if the aim of the research is theoretical-that is, to bring an understanding to the facts-then the critique of being obvious carries more weight. It carries more weight for two reasons. First, [it] is supposed to be a discipline that yields knowledge that is not simply common sense; otherwise, why the need for the expensive discipline? Second, a contribution to science is supposed to tell us something that we did not already know. To find what is already known…is not usually judged to be a**

significant contribution to new knowledge (Popper, 1963/1972). Thus, in all sciences the theories that develop counterintuitive or unexpected results are more likely to be judged to be additions to knowledge."[37]

3.3 Obviousness of Academic Hypotheses in Top Journals:

In light of the epistemological metric of non-triviality as proposed by Popper and other philosophers of science, I therefore test whether management academic hypotheses in top journals are indeed obvious. I randomly selected 50 empirical hypotheses from A-level management journals, to be tested for obviousness.

To translate the extensive jargon employed in management research articles that may mask the obviousness of a particular hypotheses, I empanelled a group of 7 volunteers (call this group 1), none of whom have ever worked as managers or ever studied management as a formal scholarly discipline. These volunteers were then randomly selected to peruse a subset of the 50 hypotheses. At least 2 volunteers were assigned to each hypothesis, and no volunteer was assigned more than 15 hypotheses. Volunteers were assigned to translate any jargon they found in their assigned hypotheses to colloquial language that could be commonly understood by untrained subjects. The volunteers performed this translation by perusing the front-ends[38] of the papers and relying upon the terminology and definitions within those front-ends. Given that every

[37] Cole 1994.

[38] The 'front-ends' of management papers consist of the Introduction and Theoretical Development sections of the paper that precede any empirical results Volunteers were specifically told not to examine the empirical results of papers and, whenever possible, were provided with paper front-ends that excluded the results.

assigned hypothesis is directional in nature - that is, the hypotheses proposed that certain variables were either positively or negatively correlated with each other[39] - volunteers were tasked with restating their assigned hypotheses in 'neutral' directional format in order to remove any indication of the direction of the hypothesis as proposed by the article's author... For example, if a hypothesis states that variable A is positively related to variable B, then I had my volunteers restate the hypothesis as a *choice* question of *whether* A was positively or negatively related to B. An example is shown below.

All group 1 volunteers assigned to a particular hypothesis would agree on a particular translation. I maintained veto power over a translation if I felt that it had severely misconstrued the meaning of the hypothesis, but I never once had to invoke that power. *Importantly, none of these volunteers was told what the ultimate purpose of the translations*, in an attempt to avoid translation bias. All that these volunteers knew is that they were being instructed to translate jargon into common language and to eliminate any indication of the hypotheses' predicted direction.

[39] As stated in Edwards (2010), almost every management hypothesis published in the top journals is directional in nature.

Example of an original management hypothesis and its translation by the Group 1 volunteers

Note, the translation was expressed in neutral directional format to conceal the positive direction of the original hypothesis. Also, the definition of "high self-monitors" included in the translation is a direct quote from the text of the original paper.

Original Text of Hypothesis:

"The higher the self-monitoring score, the larger the number of new friends an individual will attract over time."

Hypothesis after translation from jargon and converted to neutral directional format:

Given the following definition of "high self-monitors":

"High self-monitors have been described as "consummate social pragmatists," able and motivated to project images designed to evoke positive affect and conferrals of status in their relations with others."

What would you guess is the correlation between the following two variables:

Variable 1: High self-monitor score

Variable 2: Number of new friends an individual will attract over time

Figure 15

With translations in hand, I then developed a survey tool using the 50 translated hypotheses. I then submitted that survey tool to a group of 50 volunteers, who I call group 2, none of whom overlapped with group 1. Group 2 volunteers were asked to guess what they felt the most likely directions of each of the hypotheses would be. Similar to group 1, no members of group 2 had ever worked as managers, nor had they ever studied management as a formal academic discipline. As an incentive to answer honestly, a prize was paid to whoever could 'guess' the largest number of correct hypothesized directions.

I plot the results in Figure 1. Each data point corresponds to one of the 50 post-translation hypotheses, and the percentage value indicates the percentage of the 50 Group

2 volunteers who correctly guessed the direction of the hypothesis. A 100% meant that every single Group 2 volunteers correctly guessed the authors' proposed direction of the hypothesis and hence the hypothesis was entirely obvious to everybody, whereas a 0% meant that everybody guessed wrong and therefore the hypothesis was not obvious to anybody.

Results of the 50 volunteers guessing

the correct directions of the post-translation hypotheses

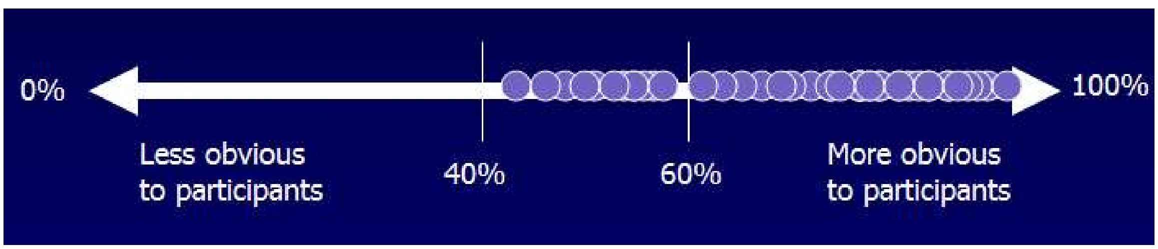


Figure 16

The interesting aspect of Figure 1 is not only is that the bulk of the observations skewed strongly toward the right side of the graph, which means that the majority of observations were relatively obvious. More importantly, *not a single hypothesis's direction* was correctly guessed by less than 44% (or 22/50) of volunteers. Hence, the direction of even the least obvious hypothesis was nevertheless correctly guessed by nearly half of the volunteers. None of the hypotheses could be said to be truly non-obvious (which I define to be less than 40%).

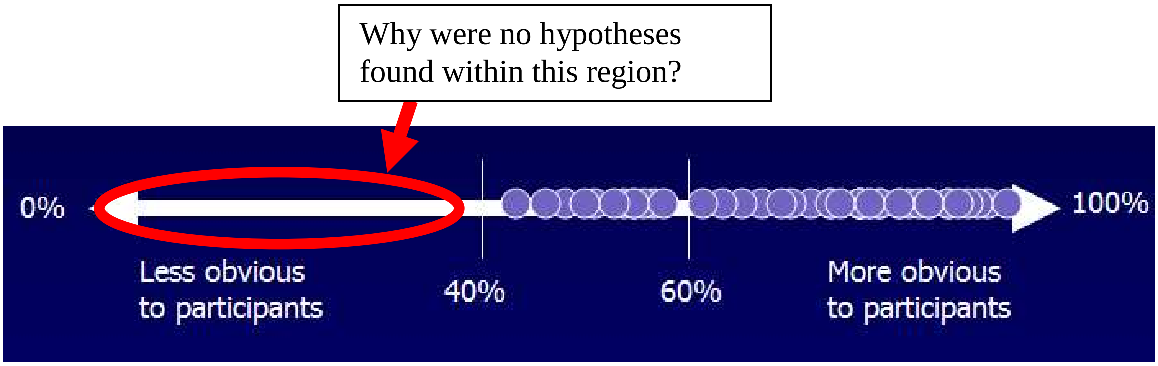


Figure 17

As a thought experiment to serve as a point of reference to compare the above results, consider what would happen if you knew absolutely nothing about a directional hypotheses in question. You could always flip a coin, as a directional hypothesis implies that the 2 variables comprising the hypotheses are either positively or negatively correlated. Even the proverbial alien from outer space who has no familiarity with humanity could guess the correct direction of a directional hypothesis 50% of the time. We could therefore assign that 50% probability as a highly conservative 'null hypothesis' equivalent to the entire set of Group 2 volunteers being aliens who know nothing about the phenomena that management scholars research. The graphical results illustrate that the 'aliens' null hypothesis can be clearly rejected (P<0.01) and that the volunteers clearly found management research to be obvious.

As a further check, I then returned to the group-1 volunteers for and asked them to submit their personal thoughts about the translation procedure they had been asked to complete. To reiterate, none of them were ever told the purpose of the translation exercise, and none of them reported having discussed or having even met any of the group-2 volunteers. Hence, the reason for why they were tasked to perform those translations was ostensibly still a mystery to them. All they knew is that the hypotheses they were asked to translate came from management academic journals, but they did not know the quality of those journals (which unbeknownst to them were all A's) Nevertheless, the responses that I obtained ranged from disbelief, to sarcastic, to scathing. Some of the quotes I received are illustrated in the following tables. More importantly, not a single group-1 volunteer had a positive word to say about the hypotheses they translated.

Disbelief quotes	Sarcastic quotes	Scathing quotes
"I don't know business schools. Is this what business school faculty do all day long? Really?"	"Here's my summary: People want respect. The End."	"I can't believe people spend time actually writing this ****. I can't believe that people spend time reading this ****."
"These are all low-level journals, am I right?"	"I really learned something. I learned that if you have more social status, you will get more flattery. See, there was my problem all along. I thought that more social status means more insults. Now I know better."	"Straight-up tosh"
"Where did you get these? Is this a joke?"	"They'd be educational…for a 10-year-old."	Why why why why why why why why why why do people spend their time [writing] this?"

Figure 18

So evidently, the hypotheses that top-level management journals publish are obvious. As Kenneth Thomas and Walter Tymon observed in their paper in the Academy of Management Review in 1982 (which was itself citing a 1976 paper by G.C. Lundberg), "most of the fields' findings are trivial and add little to common sense." [40] And if it adds little to common sense even to non-practitioners, surely real-world managers are even more likely to find much management academic work to be trivial. That then raises the natural question of what then is the value of management theory as it seems to generate hypotheses of such obviousness -or perhaps more accurately, only ever seems to generate *publishable* hypotheses of such obviousness.

Of course the underlying assumption is that the true purpose of management academic theory is to advance truth. If that is so, then a Popperian/Lakatosian approach to theory is entirely appropriate: theories that generate bold, non-obvious hypotheses that are then empirically validated should be accorded the greatest status, in the same manner that quantum mechanics and relativity produce a litany of downright ludicrous hypotheses., but which have nevertheless been empirically validated time and time again.

3.4 Conclusion

The academic management theories published in top management journals as well as my experimental design seem to produce relatively few non-trivial predictions, and certainly no counterintuitive predictions. The most nontrivial predictions they produce are ones

[40] Thomas & Tymon 1982.

whose results were predicted by roughly half of a population possessing no prior training in management – such predictions being essentially akin to a coin flip. That is, actual obviousness fares worse than randomness. Indeed, the theories produced by gurus may be less trivial than those produced by academia (although this notion was never empirically investigated and is therefore a target for further research). Nevertheless, the paucity of nontrivial management theories is striking. According to the modified Christensen-Sundahl model, theories that fail to produce nontrivial predictions are supposed to be reformulated/reclassified – yet that apparently would entail reformulating a vast litany of extant management theory.

Once a proposed theory is verified as having indeed generated falsifiable, non-trivial predictions, that theory is promoted to candidacy status. That theory is then subject to empirical validation, a topic to which we now turn in the next chapter.

4 Empirical Testing & Prediction Validation

After verifying that a proposed theory indeed generates empirically testable (that is, falsifiable), non-trivial predictions, we can promote the theory to 'candidate theory' status. The next step of the framework is naturally to then test the predictions of the candidate theory. We are therefore at the underlined stage of the framework.

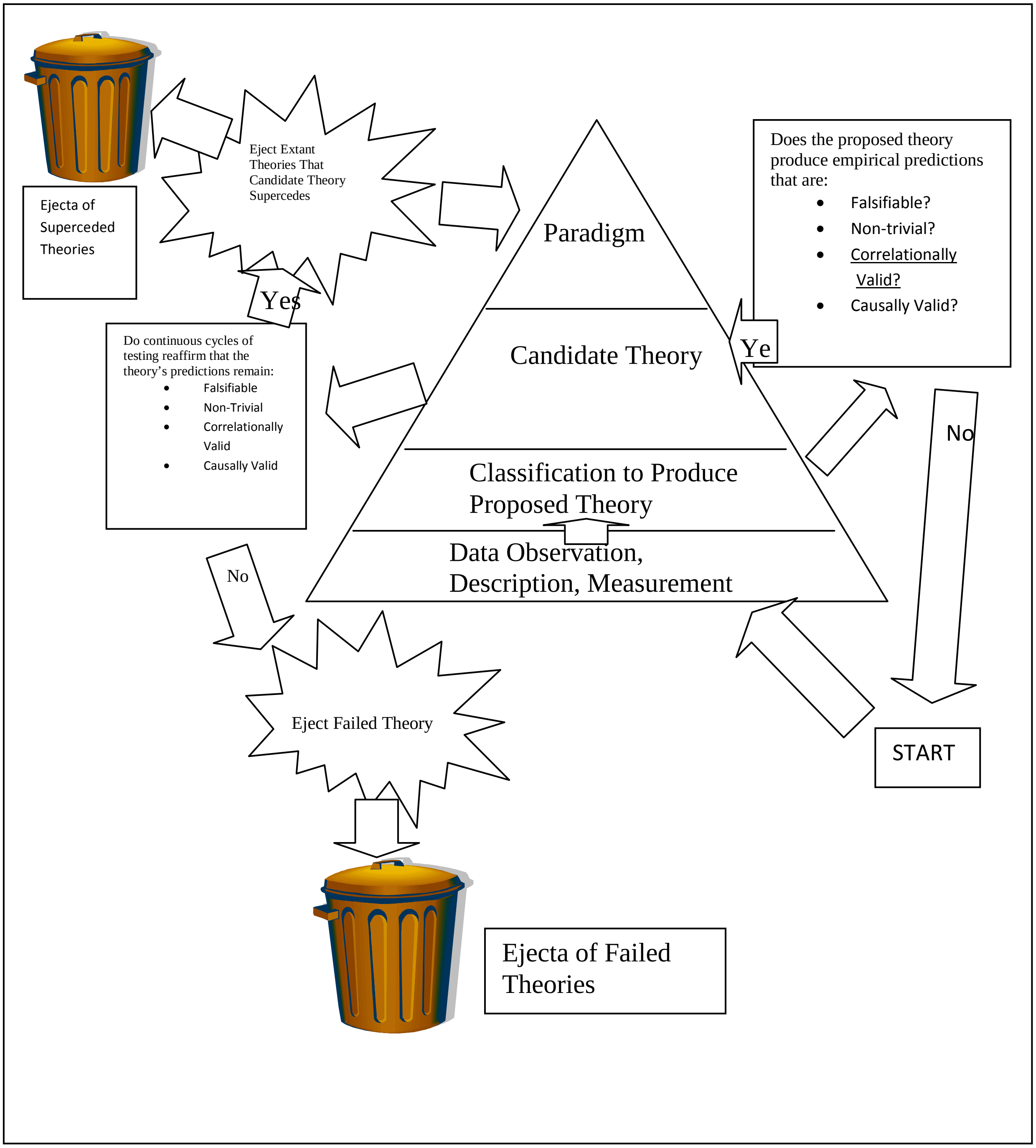

Ejecta of Superceded Theories
Eject Extant Theories That Candidate Theory Supercedes
Yes
Paradigm
Candidate Theory
Classification to Produce Proposed Theory
Data Observation, Description, Measurement
Does the proposed theory produce empirical predictions that are:
Falsifiable?
Non-trivial?
Correlationally Valid?
Causally Valid?
Ye
No
START
Do continuous cycles of testing reaffirm that the theory's predictions remain:
Falsifiable
Non-Trivial
Correlationally Valid
Causally Valid
No
Eject Failed Theory
Ejecta of Failed Theories

Figure 19

In this chapter, I reinvoke the 2x2 grid framework used in Chapter 2 as a simple pedagogical tool to check whether a candidate theory has been properly validated. I then invoke the theories of two famed management guru books - 'In Search of Excellence' by Tom Peters & Robert Waterman and 'Good to Great' by Jim Collins - as case studies regarding how to conduct proper empirical validation. Note that both of these books are widely regarded as two of the most influential managerial theory books ever published. As previously mentioned, In Search of Excellence has sold over 6 million copies to date and was the most widely held library book in the United States from 1989-2006 according to the WorldCat database.[41] By comparison Good to Great is no certainly no slouch, having sold over 4.5 million copies to date and has been cited by several members of the Wall Street Journals' CEO Council as "the best management book they have ever read."[42] Yet despite their sales accolades and laudatory blurbs, I demonstrate through use of the 2x2 grid framework that the proposed theories espoused by each of those books have not been subjected to proper validation. I then examine management academic papers published in the following leading journals: Administrative Science Quarterly, the Academy of Management Journal, Organization Science, Management Science, Strategic Management Journal and the Academy of Management Review[43] and

[41] WorldCat Database 2006.
[42] Murray, A. 2010 The Wall Street Journal Essential Guide to Management.
[43] Every one of these journals earned scores of no less than A/A+ rating from the 2013 Journal Quality List(JQL) of Anne-Wil Harzing. JQL is available here: http://www.harzing.com/jql.htm

demonstrate that, despite their quantitative sophistication, they are little different from the books of the gurus in that they lack proper empirical validation. Indeed, I show that those theories, rather than merely being dispassionate reports of empirical validation tests, instead exhibit surprising parallels to Hollywood screenplay tropes. I conclude with a discussion of the state of validation in management theory.

4.1 The Grid Framework

As discussed in Chapter 2, falsifiability requires that countervailing evidence not be *definitionally* impossible to discover. Such definitional impossibility can be assessed by the 2x2 cell framework by validating that each and every framework cell could conceivably be filled; nothing inherent in the theory would force datapoints that ostensibly belong to a certain cell to be reclassified into a different cell.

Candidate theories must then be subject to a validation process. The 2x2 matrix can be invoked once again to determine whether such a validation process has occurred. Each of the 2x2 framework cells must not only be definitionally viable so that the theory is falsifiable, but also *methodologically* viable so that it can be properly validated. That is, the research methodology that validates the theory must actually run the risk of collecting datapoints that fail to conform to the theory.

As an example, consider a generic theory that X is positively associated with Y. One would therefore expect to collect datapoints that would predominantly belong to cells B and C, with relatively few belonging to cells A & D. Indeed, if the theory in question was deterministic in nature, then the only populated cells would be B & C. However, because most management theories are stochastic in nature, it is sufficient to demonstrate that cells B & C are predominantly filled compared to cells A & D.[44] that scenario is illustrated in the following figure.

[44] Technically speaking, what matters is that the contingency cross-product quantities within B and C exceed the cross-product quantities within cells A & D by a statistically significant amount. See Forgues (2012).

For a Generic Theory That X and Y are Positively Associated.
The Theory is Validated if cells B/C are
More Populated than Cells A/D

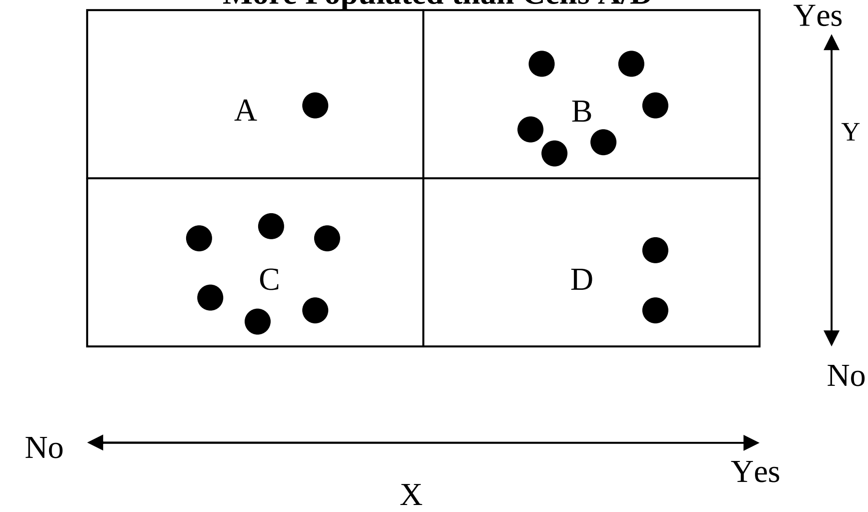


Figure 20

The risk of encountering disconfirming data is a necessary component of the validation process. For the generic theory that X is positively associated with Y, the researcher must run the risk that he might find numerous datapoints within cells A/D - perhaps numerous enough to call the purported relationship between X and Y into question. For example, if from a statistical standpoint, roughly the same number of datapoints was found in cells A & D relative to cells B & C, then one might conclude that the theory is wrong for no positive relationship between X and Y would seem to exist. If *more* datapoints were found in cells A & D relative to cells B & C, then one might conclude that the theory is not only wrong but actually diametrically wrong, for the correct relationship between X and Y is not in the purported positive direction, but rather is negative. Those two scenarios are illustrated in the following two figures.

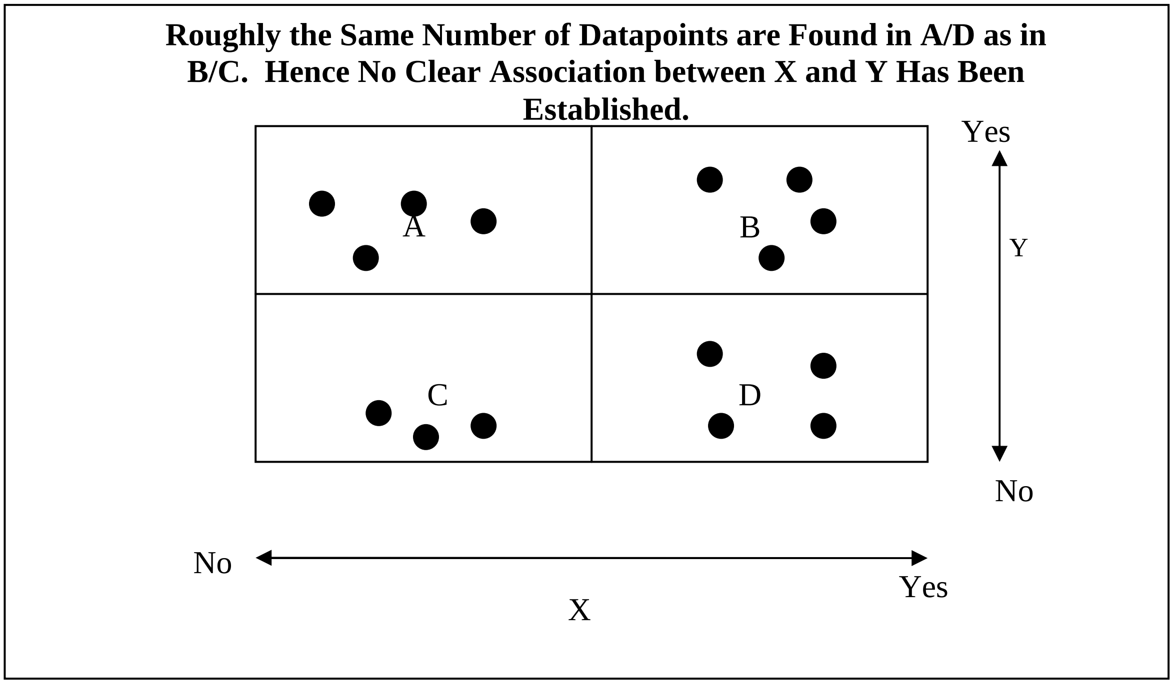


Figure 21

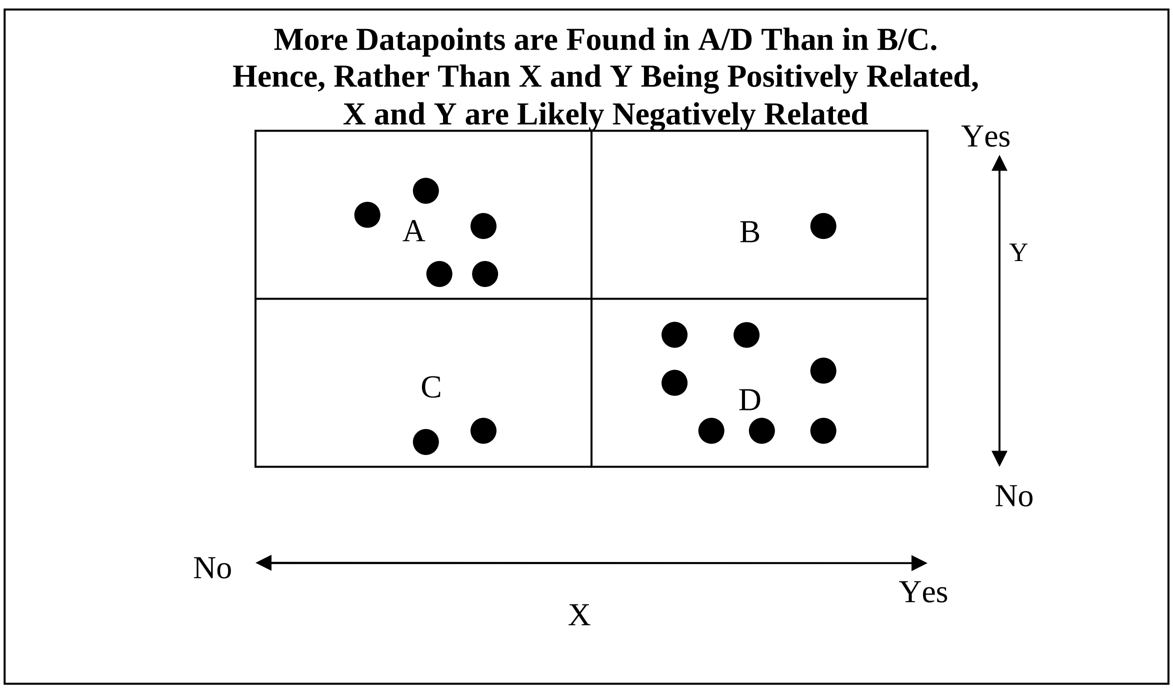


Figure 22

4.1 Case Study #1 'In Search of Excellence'

Perhaps the most prominent example of a best-selling management theory that fails to conform to *experimental* falsifiability is the theory of managerial success proposed by Tom Peters and Robert Waterman in the famed book 'In Search of Excellence'. Peters & Waterman investigated a group of firms[45] that they deemed as 'excellent' as measured by a number of metrics ranging from financial performance, respect amongst managers at peer firms, and the general opinions of Peters & Waterman themselves[46]. Peters & Waterman then extensively investigated by thorough, but exclusively inductive work, the histories of those firms that they deemed to be excellent to uncover the cluster of strategies and managerial practices shared amongst all of them. That strategy-set – which turned out to be a set of 8 management components[47] – were then declared by Peters & Waterman to be the strategy-set associated with firm excellence.

Peters & Waterman's theory is valid from a definitional standpoint, and application of the 2x2 grid framework demonstrates that to be so as shown in the following figure. Use/non-use of their strategy-set is not definitionally linked to excellence/non-excellence. Excellent firms might conceivably not have adopted the strategy-set; conversely, firms

[45] As pointed out by Stewart (2009), while Peters & Waterman claimed that their research study was based on a group of 43 excellent firms,"15 of the excellent firms are instantly forgotten upon making the list; a further 20 or so receive only cursory references; and only 7 are discussed in depth."

[46] See Peters & Waterman's Appendix and Chapters 1-2 for their precise definition of excellence.

[47] The 8 components of the strategy-set were found to be: (1) A bias for Action; (2) Being close to the customer; (3) Autonomy and Entrepreneurship; (4) Productivity through People; (5) Hands-On, Value-Driven Management; (6) Stick to the Knitting; (7)Simple Form, Lean Staff; (8) Simultaneous loose-tight properties. An explanation of each of these components is beyond the scope of this text; interested readers are directed to Peters & Waterman's original text.

that did adopt the strategy set might not achieve excellence. Every single cell of the framework is viable and so the theory is indeed falsifiable. Assuming that the Peters' & Waterman's theory also generates nontrivial predictions – which it arguably does[48] - the theory can be promoted to candidacy status.

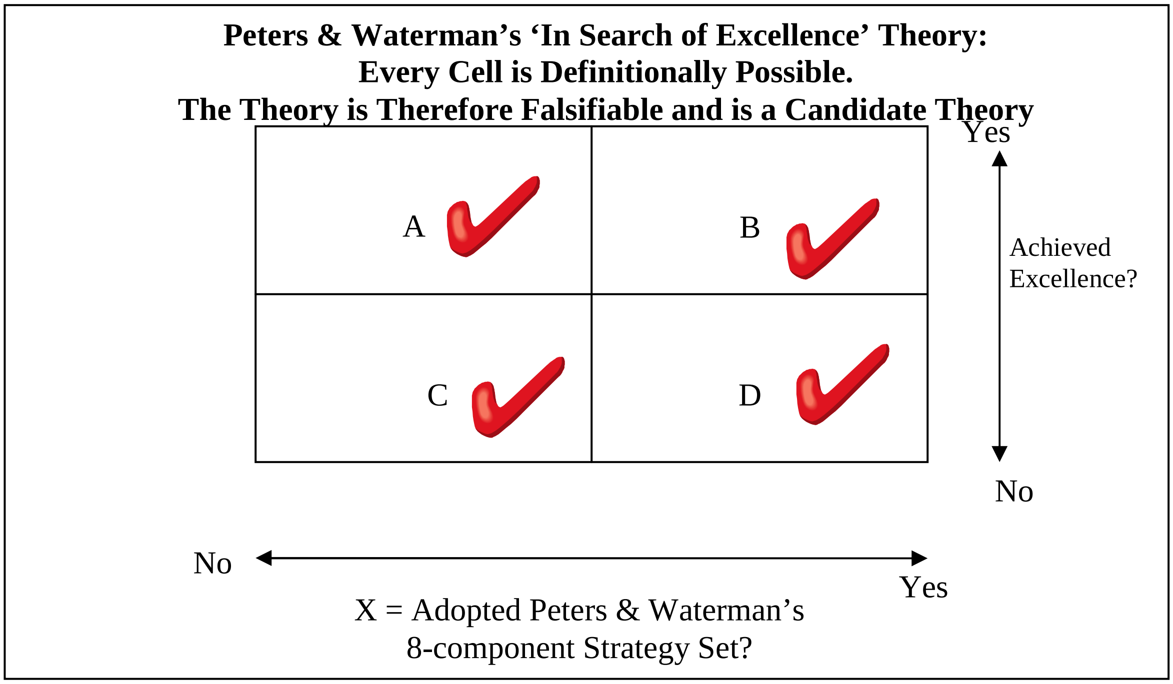


Figure 23

Yet however viable all cells might be from a *definitional* standpoint, only cell B is viable from a *methodological* standpoint. Recall that Peters and Waterman specifically selected only firms that had attained their definition of performance excellence. Data on Non-excellent firms were never collected and hence cells C and D are rendered methodologically impossible. Peters & Waterman furthermore then proceed to search for a specific strategy set *shared* amongst those excellent firms, thereby precluding the inclusion of any datapoints about the firms they studied that did not share the same strategy. In the course of searching, Peters & Waterman must have surely encountered

[48] Follow-on research should confirm whether the predictions of Peters & Waterman are indeed nontrivial.

numerous strategy characteristics that were not shared amongst the entire set of excellent firms, only to be discarded by Peters & Waterman who were searching only for a strategy set mutually shared by their sample of excellent firms. Peters & Waterman therefore *methodologically* exclude all unshared strategy characteristics from the declared final strategy set, hence systematically excluding all datapoints except those in cell B.

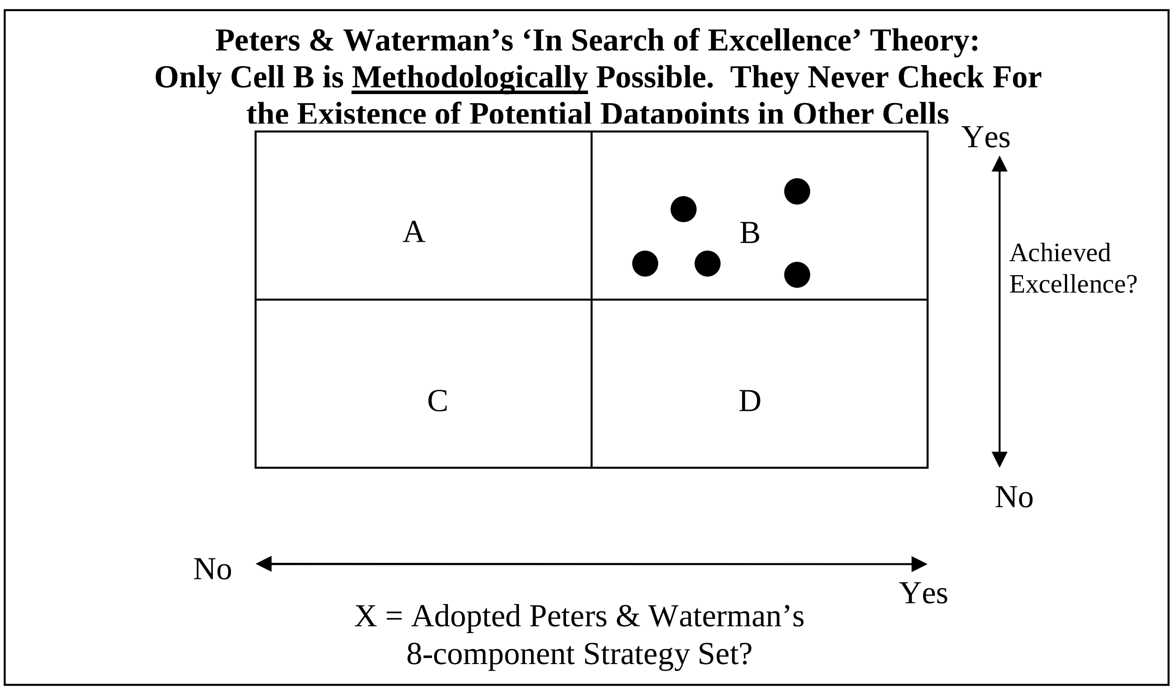


Figure 24

'In Search of Excellence' is therefore not a validated theory, however much Peters & Waterman might insist it to be so, because their methodology neglects any potential *control group* outside their studied firms or even across the firms they studied. Recall that proper validation of a theory that purports that X and Y are positively associated requires placing that theory at risk by examining whether cells A and D might be sufficiently populated as to cast doubt upon the theory. Might there in fact be numerous firms that adopted the Peters & Waterman strategy set yet were not excellent and hence belonged to cell D? Might there be firms aplenty that achieved excellence without using the Peters & Waterman strategy set and hence belonged to cell A? They'll never know; their methodology filters away any such potential datapoints. Generating validated theoretical statements regarding the strategy-set of a set of excellent firms requires a control group with which to compare. Such a control group would serve to classify datapoints with low values of X and/or Y: the more control-group datapoints that are found to be populate cell C rather than A/D, the more valid is the theory.

As a thought experiment, compare the following two figures, where the first figure illustrates what Peters & Waterman demonstrated, whereas the second figure is a hypothetical study regarding the datapoints that they could have collected with a methodology that did include a control group? What if they found that cells A and D are actually *more* heavily populated than are cells B and C? That would imply that adoption of the Peters & Waterman strategy set actually correlates with a *lower* probability of excellence than would non-adoption. That would indicate that Peters & Waterman's strategy set is a strategy set that firms might be well-advised to avoid. But again, Peters & Waterman's methodology prevents them from discovering this. That they examined only the contents of cell B tells us nothing about the populations of the other three cells, and therefore prevents us from drawing any conclusions regarding the association between strategy set and performance—especially as it concerns excellence.

Peters & Waterman Show Only This Data and Argue that Their Strategy Set is Positively Associated with Excellence

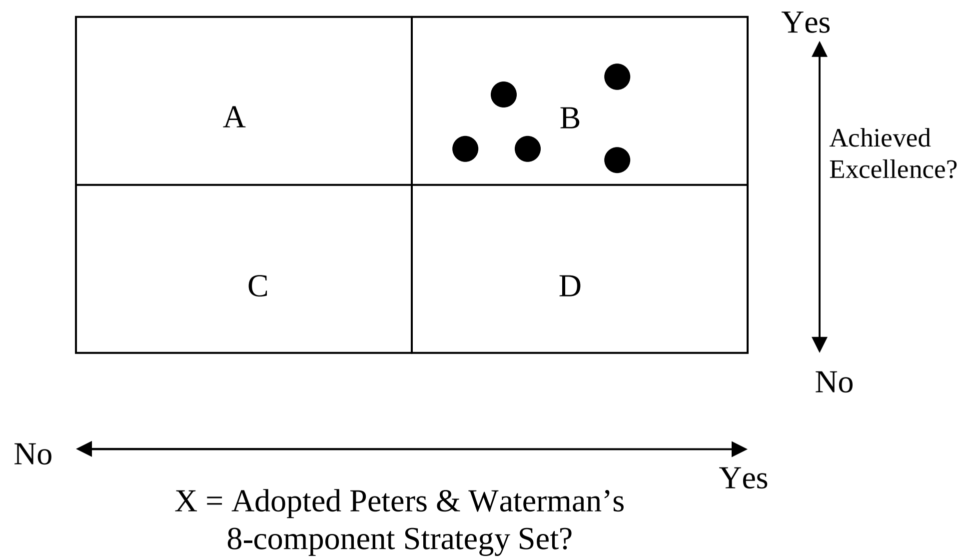


Figure 25

What if the following represented the true population of the 2x2 matrix according to the Peters & Waterman strategy set? Cells A/D are actually <u>more heavily</u> populated than are cells B/C. Peters & Waterman's strategy set is therefore <u>negatively</u> associated with excellence.

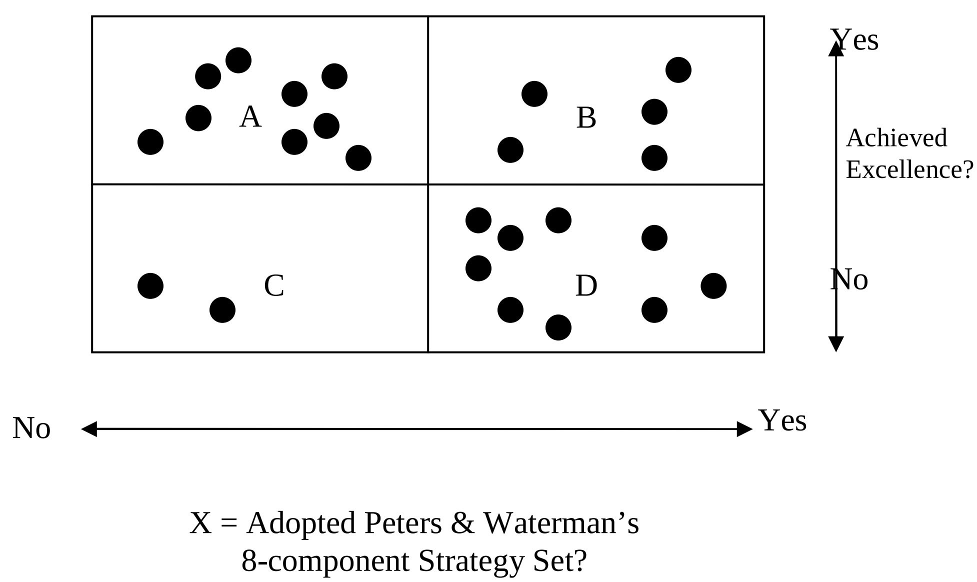


Figure 26

To be sure, while the lack of a control group is a crucial methodological, flaw in Peters & Waterman's methodology, having a control group alone is insufficient to validate a theory – a point I illustrate in my next case study of Good to Great by Jim Collins.   I now segue to this case study to reveal another methodological flaw that demonstrates the lack of validation, and how the 2x2 grid framework helps to reveal that flaw.

4.2 Case Study #2: 'Good to Great'

For the purposes of brevity, I refrain from discussing the extensive methodology of Good to Great in its entirety; interested readers are directed to Chapter 2 as well as Appendix 2-3 of the book. However, in a nutshell, Collins and his research team first builds a list of 11 examples of firms that became 'good-to-great': firms that were once 'good' but that later became 'great' as defined by an strong upwards shift in share price performance both relative to their industries and relative to a prior period of time for those same firms when they were performing at a merely 'good' level.   They then build a matching set of candidate 'good' firms by using the same metrics as that of the 'great' firms, but where those good firms share prices did not exhibit the upwards share price performance shift that the 'great' firms exhibited. Then they devise an algorithm (based on the average value of six criteria to be explained later) to find the closest match that pairs a good firm to each great firm to generate 11 great/good matched pairs.

With the matched pairs in hand, Collins and team then delves through media reports, press releases, cases studies, and any other literature they can find regarding their

Great/Good matched pairs. They also supplement that literature with interviews with company executives to generate a strategy-set of differences between the good/great matched pairs. Collins openly declared that he had no pre-set theory, stating in a highlighted paragraph:

> "It is important to understand that we developed all of the concepts in this book by making empirical deductions directly from the data. We did not begin this project with a theory to test or prove. We sought to build a theory from the ground up, derived directly from the evidence."[49]

Collins' research leads him to enumerate the following five features as the expositive difference between Good and Great firms.[50] I refer to this set of five features as the Good to Great Strategy Set**' (G2GSS)**: all quoted descriptions below are pulled directly from the corresponding chapters of Good to Great.[51]

1. Level 5 Leadership: ""Self-effacing, quiet, reserved, even shy-these leaders are a paradoxical blend of personal humility and professional will"
2. First Who Then What: "People are not your most important asset. The right people are."
3. Confront the Brutal Facts: "Every good-to-great company embraced what we came to call the Stockdale Paradox: You must maintain unwavering faith that you can and will prevail in the end, regardless of the difficulties, AND at the same time have the discipline to confront the most brutal facts of your current reality, whatever they might be."
4. Hedgehog Concept: "If you cannot be the best in the world at your core business, then your core business absolutely cannot form the basis of a great company"

[49] Collins, p. 10.
[50] Note, Collins Good-to-Great strategy set seems to include seven elements rather than five. However, a close examination of his work reveals that the final two features – the Technology Accelerator and the Flywheel/Doom-Loop - seem not to be truly distinct strategy features. However, my resulting analysis changes little if Collins strategy set indeed consists of all seven elements.
[51] Collins.

5. Culture of Discipline: “Fanatical adherence to the Hedgehog Concept and the willingness to shun opportunities that fall outside the three circles”

To verify that G2GSS is a legitimate candidate theory, we must first examine whether it is falsifiable. That is, could a researcher conceivably find disconfirming evidence that isn’t definitionally ruled out by the tenets of the theory? To assess this question, I again invoke the 2x2 verification framework. The X axis indicates whether a firm utilized the G2GSS; the Y axis indicates whether a firm exhibited good or great performance. As shown in the previous section, the X axis is not definitionally linked to the Y axis, as none of the five tenets of the G2GSS are definitionally linked to good or great performance. For example, a firm with ‘Level 5 Leadership’ is not definitionally linked to any particular performance standard, nor is a firm that ‘Confronts the Brutal Facts’ or a firm that employs the ‘Hedgehog Concept’—and therefore such statements are not simply tautological. One could conceivably imagine measuring each of the tenets of the G2GSS and find that some firms possessing each of those tenets nevertheless do not exhibit great performance, and vice versa. Hence, the G2GSS is a definitionally falsifiable theory that can be promoted to candidacy.

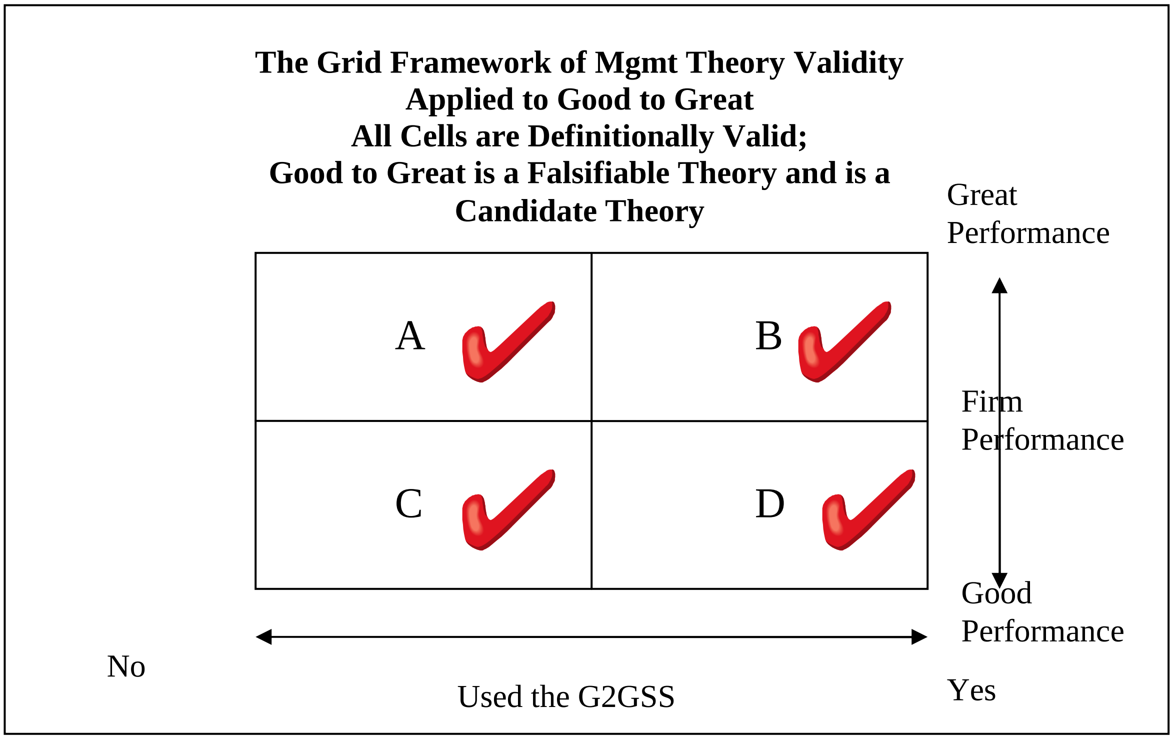


Figure 27

To his credit, Collins also sidesteps the pitfall that ensnared Peters & Waterman of failing to include a control group. Recall that Peters' & Waterman's methodology collected only datapoints regarding firms that both adopted the Peters & Waterman strategy set and achieved excellence performance. Hence, every one of their datapoints had high values on both the X and Y axis and therefore occupied cell B. Peters & Waterman <u>strongly implied</u> that if additional datapoints had been collected that belonged to the remaining cells, those datapoints were likely to be found predominantly in cell C rather than A or D. However, Peters & Waterman never empirically demonstrate or even cite the existence that what might fall into cell C, and if numerous firms were actually found not to populate cell C but rather cells A/D that would call the Peters & Waterman's entire theoretical framework into question.   In contrast Collins explicitly collected data regarding a control group of firms whose performance was 'good' (but not 'great'), and hence occupied low values on the Y-axis. Yet not a single member of the control group was found to have adopted the G2GSS, hence every single control dataset member was found to populate cell C with none populating D.

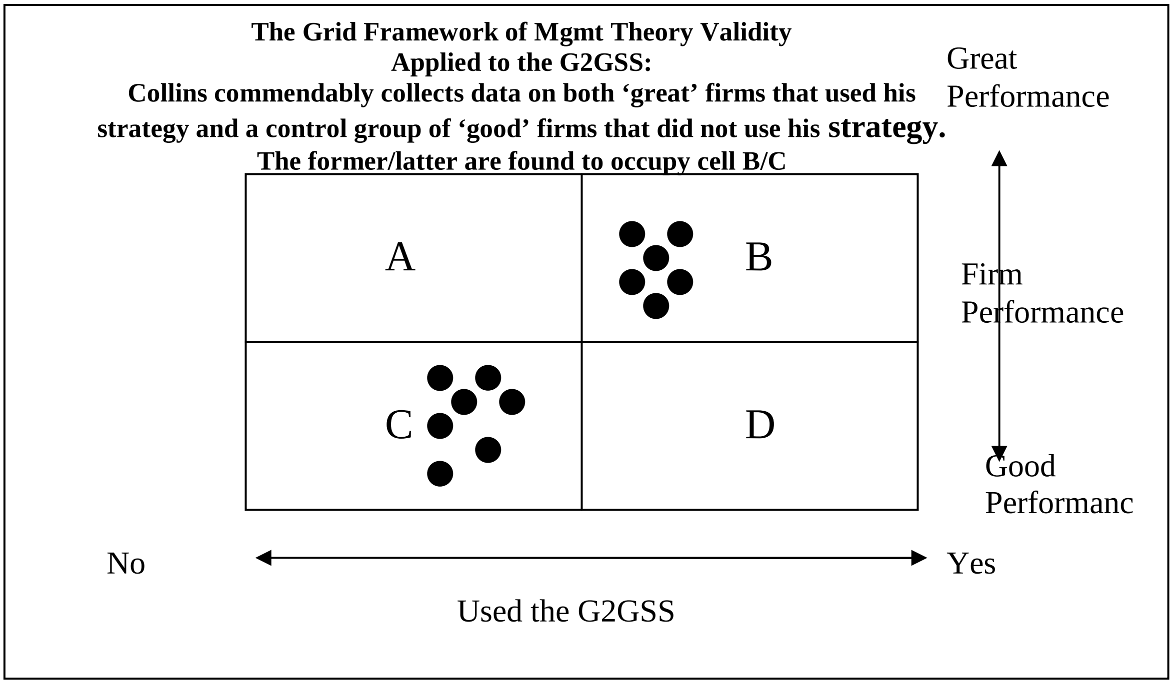


Figure 28

Did Collins therefore successfully validate his theory? The answer unfortunately is 'No.' Despite Collins' commendable usage of a control group, for Collins, like Peters & Waterman before him, never subjects his theory to risk. Recall that Collins explicitly stated that he induced the definition of the G2GSS from his dataset of good/great firms through a data-mining process. He therefore defined the categories comprising the X axis of the 2x2 grid framework as being that particular suite of strategy characteristics entirely shared by his set of great firms that were unshared by his control group of good firms.

Any strategy characteristics shared by both good and great firms – and that might therefore be indicative of datapoints in cell D – were simply not included in Collins definition of the G2GSS. Likewise, any strategy characteristics that that were not entirely shared by every member of his sample of great firms – and therefore perhaps might be indicative of datapoints in cell A - were likewise excluded from the definition of G2GSS. Hence, Collins datapoints neatly fell exclusively into cells B and C only because he defined them as such: his data-mining exercise allowed him to continually modify the precise definition of the G2GSS upon which the categories of the X axis are based until he found a G2GSS definition where his entire particular sample of great firms would necessarily entirely be classified into cell B and the good firms could likewise be necessarily entirely classified into cell C. [52] That Collins never placed his theory at risk by methodologically excluding any datapoints from appearing in cells A or D means that his theory has yet to be validated.

The following two figures invoking the grid framework illustrate the key problem. The first figure illustrates what Collins' methodology of first obtaining a dataset of good/great firms and afterwards mining that dataset to induce the specific categories of the X axis. In other words, he collected the data <u>first</u> and then determined what the salient categories ought to be <u>later.</u> Nothing is risked when you determine the categories to which the datapoints should belong only <u>after</u> knowing what the datapoints are. The second figure illustrates the hypothetical situation regarding what Collins might have found had he placed his theory at risk by <u>first</u> establishing the categories of the X axis and <u>afterwards</u>

[52] Peters & Waterman perform a similar data-mining exercise and their theory therefore suffers from the same flaw.

collected data to classify them into categories that had previously been defined. Under such a procedure, perhaps more datapoints would have been found in cells A/D than in cells B/C. But we would never know this as Collins methodology precludes any datapoints from populating cells A/D. A management theory is validated only if it endured the risk that disconfirming datapoints might be found.

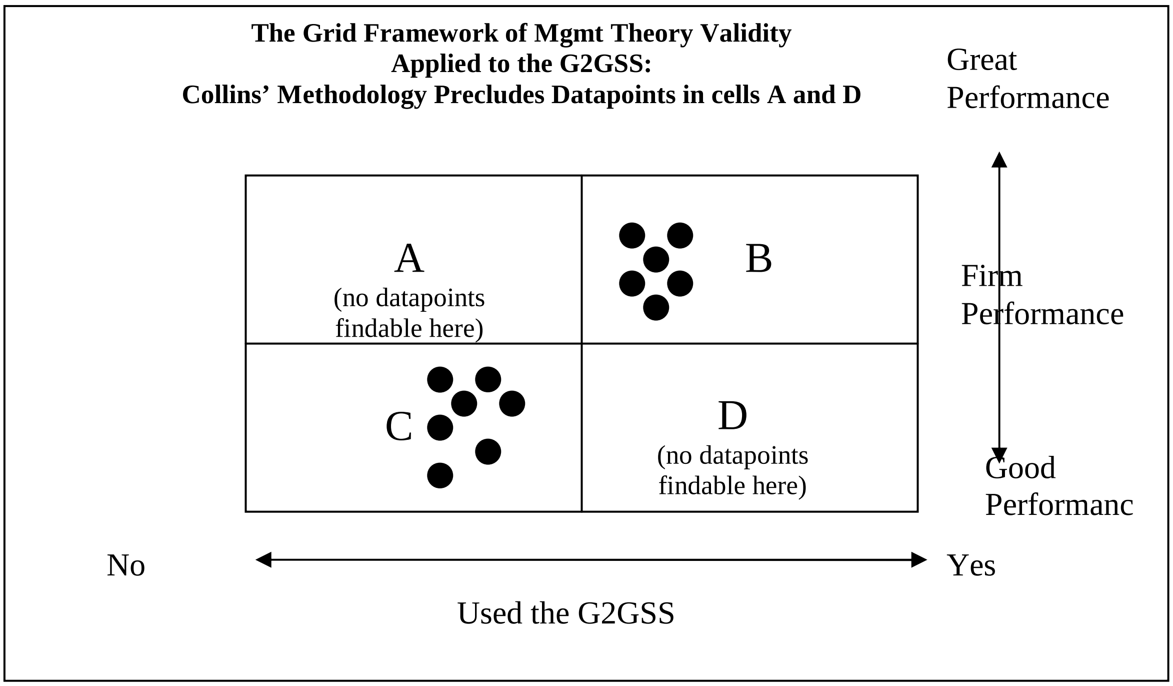


Figure 29

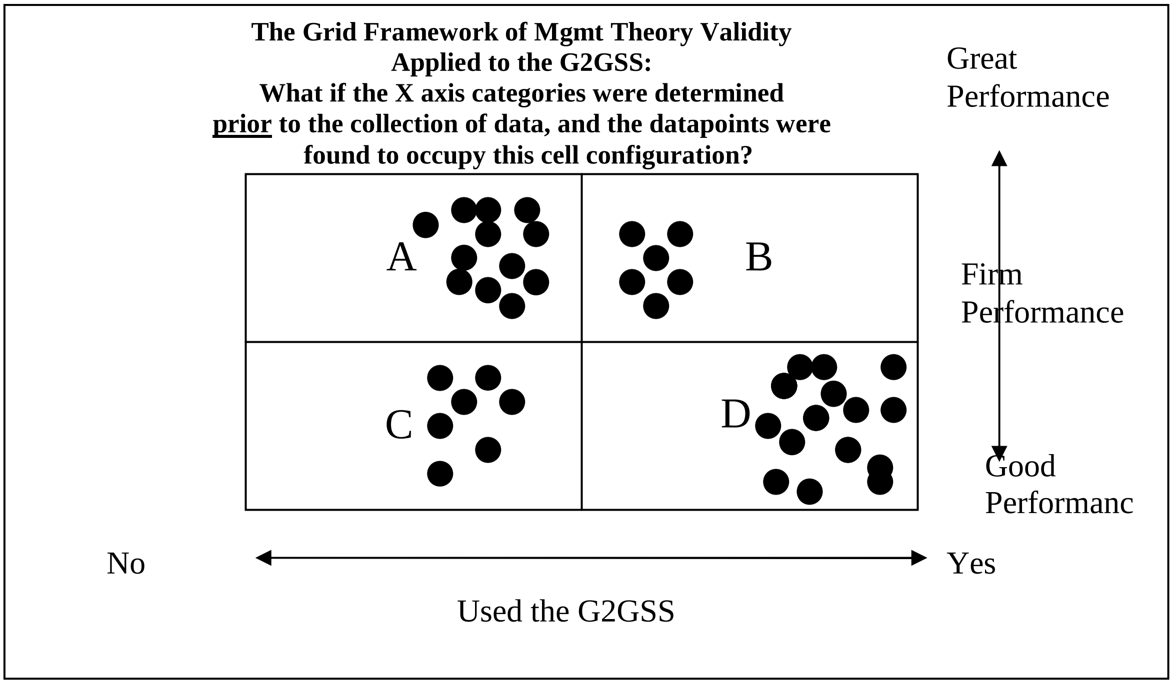


Figure 30

Now, to be fair to Collins, one could argue that he was merely engaged in a process of theory induction – a theory-generating process for which data-mining is a legitimate exploratory tool. The proper procedure to validate Collins' newly induced candidate theory would be to place it at risk of discomfirmatory data by first establishing the definition of the G2GSS (and therefore fixing the categories of the X axis) and then obtaining and categorizing <u>another</u> dataset comprised of an entirely new batch of good & great firms from which he did not induce the definitions of the G2GSS. That new dataset, if shown to predominantly occupy cells B/C rather than A/D, would then legitimately validate Collins' theory.

However, Collins argues in his book that such a step is unnecessary, claiming that his theory has been validated despite never obtaining a new dataset. Collins presents a novel statistical argument for why he believes this to be so. We now unpack his statistical argument to demonstrate why it fails to support proper validation for his theory.

4.2.1 Did Collins validate his G2GSS findings?

To be credit, Collins openly concedes that he cannot *deterministically* validate his theory. Collins states on his own website that "it's impossible to claim cause and effect with 100-percent certainty."[53] However he does offer the following statistical argument as evidence of the supposed validation of his findings by including the following statement (p.211) regarding a discussion he had with a leading statistician:

> "When we asked University of Colorado applied mathematics professor William P. Briggs to examine our research method, he framed the question thus: What is the probability of finding by chance a group of eleven companies all of whose members display the primary traits you discovered while the direct comparisons do not possess those traits? He concluded that the probability is less than 1 in 17 million. There is virtually no chance that we simply found eleven random events that just happened to show the good-to-great pattern we were looking for. We can conclude with confidence that the traits we found are strongly associated with transformations from good to great."

[53] http://www.jimcollins.com/books/research.html

To the layman's ears, the 1-in-17 million figure surely sounds intimidatingly impressive. An event of such infinitesimal probability sounds so tiny that it ought to be dismissed; therefore Collins would seemingly have statistically validated his candidate theory regarding great firm performance and the G2GSS. Alas, statistically speaking, Collins claim is far less meaningful than it may seem, indeed perhaps conveying no meaning whatsoever. As paradoxical as it may seem, low-probability events are surprisingly mundane. To understand why this is so, allow me to present two brief primers illustrating the counterintuitive nature of statistics. The first primer, regarding what statisticians have jokingly deemed the 'Law of Truly Large Numbers', illustrates the importance of carefully defining the *search space ensemble* under investigation. The second primer illustrates the crucial importance of the *timing* when a statistical calculation is conducted: either before or after the data has been analyzed.

4.2.2 "The Law of Truly Large Numbers"

Consider the following story. Less than a week after Thanksgiving of 2012, a miracle happened, statistically speaking – a miracle that renders the bold statistical claims of Jim Collins puny by comparison. Forget about the 1-in-17-million probability chance that Collins claims would have had to occur if the set of strategies he identified as differentiating the 'Good' from the 'Great' firms were merely a product of random luck. On November 28, 2012, an event occurred that the most eminent statisticians in the world would agree happens with only a 1 in 175 million chance – an entire order of magnitude of randomness greater than the claims of Collins. And the event showered vast fortunes upon its recipients. Why bother with 'Good to Great'; the recipients of this most recent miracle clearly have a more fortuitous tale to tell!

Even more remarkably, that 1-in-175 million chance event not only occurred on November 28, 2012, but has been occurring roughly every month for at least the last 9 years[54]. That's not to say that the event in question occurs each and every month like clockwork, for a few months may pass with no occurrence, whereas other months may witness multiple occurrences. But over the long run, the average rate of occurrence of that event seems to be about once a month: over 100 total events since recorded since 2003. The momentous event increases in total personal wealth precipitated by this unlikely string of events surely vastly exceeds whatever boosts in personal fortune any management could possibly enjoy by reading the works of Jim Collins or of any other guru. Shouldn't managers be far more interested in reading books that unlock the secrets of the Monthly Miracle? After all, what ludicrously improbable set of conditions could possibly entail a wealth-generating event of such absurd unlikeliness that it take a miracle to even occur at all, let alone every month for nearly a decade?

I am of course referring to winning the jackpot of the famed Powerball lottery. Winning Powerball requires first selecting five correct numbers from a randomly drawn set of white balls numbered 1-59, where the balls will be drawn in any order (in other words a draw sequence of balls numbered 1-5-10-12-45 is the same as 45-12-10-5-1). You must also select the correct 'Powerball' number from a set of red balls numbered 1-35, where the red ball numbers are segregated from the white ball numbers (hence, choosing the

[54] http://www.powerball.com/powerball/pb_stories.asp

white #35 does not qualify you to claim a red ball draw of #35). The odds of winning the jackpot purely by chance have been mathematically established as being roughly 1 in 175 million[55], and no credible evidence of exploitable nonrandomness within the Powerball drawing system has ever been unearthed. Yet the fact remains that not only did somebody recently manage to beat the dreadful odds of the Powerball jackpot, but that feat has been accomplished monthly for years.

What renders mundane the seemingly daunting odds of winning Powerball not just once but monthly– such that most managers and MBA students are (hopefully) not clamoring for books claiming to reveal the secrets of winning Powerball - is that while the aforementioned odds are indeed characteristic of choosing the winning Powerball ticket, those odds convey no information regarding the mountain of losing tickets. Yet for every winning Powerball ticket, hundreds of millions of losing Powerball tickets were purchased. The relevant question is therefore not: 'What are the odds of any <u>particular</u> ticket winning the lottery?' but rather: "What are the odds of a winning ticket being found amongst the <u>entire ensemble</u> of lottery tickets sold this month'? Because an average of ~175 million Powerball tickets are sold every month, finding one winner per month is a rather banal finding. Nobody can predict who the winner will be amongst that monthly set, but I doubt that anybody will be surprised to find that somebody within the search space will win.

Similarly, the more relevant statistical question regarding 'Good to Great' is not: 'What is the probability that the <u>particular</u> set of 5 strategic differences identified by Collins and

[55] http://www.powerball.com/powerball/pb_prizes.asp

his team had occurred purely by random chance (assuming that there is no true strategic difference between the good and great firms)?' The relevant question is: of the <u>entire ensemble of potential strategies</u> investigated by Collins and his team, what is the probability that <u>some</u> set of 5 strategic differences would be found? As Professors Bruce Niendorf and Kristine Beck pointed out: "This probability is far closer to 100% than to 1 in 17 million."[56] Let's harken back to the methodological procedure that Collins employed. As explained in his methodological section deemed 'Inside the Black Box':

> "It is important to understand that we developed all of the concepts in this book by making empirical deductions directly from the data. We did not begin this project with a theory to test or prove. We sought to build a theory from the ground up, derived directly from the evidence…When all was said and done, the total project consumed 10.5 people-years of effort. We read and systematically coded nearly 6000 articles, generated more than 2000 pages of interview transcripts, and created 384 million bytes of computer data…we would debate, disagree, pound on tables, raise our voices, pause and reflect, debate some more, pause and think, discuss, resolve, question, and debate yet again about "what it all means"…"[57]

Collins therefore explicitly states that he combed through a large ensemble of potential strategy sets. While nobody outside of Collins & his team knows exactly how many potential strategy sets were considered – and Collins has never published the details of his data analysis - devoting 10.5 people-years likely entailed the combing of a search space comprised of millions upon millions of potential strategy-set-combinations. In this light, the strategy-set ultimately selected has a probability of 1-in-17 million of occurring by sheer randomness therefore seems rather banal.

[56] Niendorf & Beck 2008.
[57] Collins, Good to Great, p. 10

Collins' proud declaration of the expansiveness and intensity of his research – given the large search space size that such expansive research entails – therefore not only provides the illusion of rigor[58], <u>it also statistically undermines the validity of his research</u>. Collin's good-to-great findings would actually be far more statistically impressive if his team spent, say, only a few person-hours rather than more than an entire person-decade! A tiny amount of time spent on research would necessarily imply a tiny search space ensemble – and if they nevertheless were able to discover a good-to-great strategy set despite the tiny ensemble, the claim of a statistically unlikely event would carry more weight. Instead his team built and plowed through a vast search space – it is little wonder that they found something—just like winning the Powerball.

Furthermore, even the provided figure of 10.5 person-years of effort fails to fully capture the dimensions of the ensemble under investigation. To invoke the time-worn idiom: whenever you look for something, you'll always find it in the last place that you look. That 10.5 person-year figure merely represents the final tally when Collins and his team ceased searching because they had ascertained a publishable strategy set. It seems unlikely that had Collins and his team still not discovered a publishable strategy set after 10.5 person-years of effort, they would have simply folded shop and declared failure. It seems far more likely that they would surely have continued searching until they finally discovered something, or had utterly exhausted their resources. Winning at a slot machine after only a few pulls is statistically impressive only if you had previously

[58] Such an illusion is akin to what Phil Rosenzweig has deemed the 'Delusion of Rigorous Research'. See chapter 5 of Rosenzweig (2007).

determined that, no matter what occurs, you would have ceased playing after those few pulls even if you had never won. It is far less statistically impressive to know that you would continue playing over and over until you finally win, even if you happened to luckily win after only a few pulls.

Even if Collins had exhausted his resources without discovering anything, he certainly would never publish anything enumerating the millions of strategy-set combinations they had considered that failed. If a future researcher (call him Kim Bollins) then chose to pursue the same research question that Collins pursued and failed - and Kim Bollins declared that he found a strategy-set with only a tiny probability of being a product of pure random chance – nobody would ever know that Collins and his team had previously checked millions of other strategy-sets and found nothing and therefore never published. By the same token, nobody knows if a researcher prior to Collins had likewise examined millions of strategy-sets and found nothing and therefore never published.

Finally, recall that the tiny 1-in-17 million figure as calculated by Collins' statistician presumes that the strategy set <u>specifically</u> includes five elements and <u>specifically</u> compares eleven great firms to eleven matched good firms. But who's to say that the strategy-set combination must consist specifically of five strategy elements? If Collins and Co. found a strategy-set combination consisting of, say, only four elements, or even only three, might that perhaps be sufficient? Does the book-publishing world enforce a rule of which I am unaware that requires that any strategy book must discuss no fewer than five strategy elements? Couldn't a man of such clearly prodigious writing talents as

Collins write a successful book based on only three or four strategic elements? Likewise, who's to say that Collins necessarily must include all eleven matched pairs of good/great firms? Surely including only ten or less would still provide sufficient material with which to publish a book. [59]

The proper statistical question that Collins asked is therefore not the aforementioned question Collins asked to his statistician: 'What is the probability of finding by chance a group of eleven companies all of whose members display the [five] primary traits you discovered while the direct comparisons do not possess those traits?". Rather, it consists of the far more expansive multi-part question:

[59] Granted, lowering the number of strategy elements and matched pairs would render the purported odds of randomness seemingly less impressive, but a calculated probability of "only", say, 1 in a million, or even 1 in a 100,000 would still surely dazzle the average Collins consumer who is untrained in statistical methodology.

Jim Collins Actual Statistical Research Methodology:

The entire search space ensemble consists of the following

A) The entire amount of time & resources that Collins and his team were ultimately willing to expend. (That is, not the 10.5 person-years that they did expend, but rather the total time that they would have spent until they declared failure.)
B) The flexibility to lower the number of strategic elements and perhaps change the number of company matched pairs as necessary, yet still provide enough material to publish a successful management book.
C) A correction factor to account for the possibility that somebody prior to Collins might have searched for a 'good-to-great' strategy set and found nothing and therefore never published, or similarly that Collins himself might find nothing and never publish, such that a future researcher who investigates the same question and does find a result would unwittingly believe that he was the only person who was searching.

Given the search space ensemble as defined above, what is the probability that Collins would have found some strategy set that is publishable?

Figure 31

The seemingly statistically unlikely but ultimately mundane findings of Collins – akin to the seemingly amazing Powerball miracle that nonetheless happens monthly - serve to demonstrate what statisticians Persi Diaconis and Frederick Mosteller jokingly dubbed "The Law of Truly Large Numbers"[60]. Their law stipulates that any random event, no matter how tiny the probability, not only might reasonably happen but indeed will almost surely inevitably occur, given a 'truly large' number of opportunities, just what Collins found.

4.2.3 The Importance of When a Statistical Calculation Was Conducted.

While witnessing the mundane 'miracle' of a Powerball victory would require waiting for about a month, why wait at all? Why not conjure up a miracle right now in your very home? Take some dice and roll one of them 15 times, recording the results of each throw sequence. Marvel at the miraculous sequence of dice-throws you just generated. Repeat to generate another miracle… and continue repeating to generate as many miracles as you desire.

What's miraculous about a series of dice throws, you may ask? Allow me to illustrate. While writing this very paragraph, I took some dice and rolled them 15 times, generating the following sequence:

[60] Diaconis & Mosteller 1989.

My dice throw: {1,6,2,5,5,2,6,4,2,3,1,1,4,2,5}

The statistical miracle is that, if we presume that my dice are fair, any particular outcome of a particular throw will occur with a probability of 1/6. Hence, the probability of the particular 15-throw dice-sequence I obtained would have occurred with a miniscule probability of $(\frac{1}{6})^{15}$ = 1 in 470 million – a probability so small as to make even Powerball seem banal by comparison. Nevertheless that sequence of 15 throws indeed occurred, infinitesimal as the probability may be.

Nor are such miracles confined to the probability-generating machinery of dice. Everyday life is inundated with statistical miracles. Whenever you drive, observe the license plate of the car in front of you, and marvel at the miracle of witnessing that particular combination of letters and numbers as opposed to any other potential sequence of letters and numbers you might have witnessed. Even a simple license plate sequence of only 5 entries, with each entry comprised of either a letter (hence 26 possibilities), or a number (hence another 10 possibilities) would result in a probability of $\frac{1}{36}^{5}$ = 1 in 60 million. Whenever you walk into a room where 10 or more people are sitting, consider the miracle of the particular seating configuration that those people took. Assuming that only 10 chairs exist, the first person has a choice of any of 10 empty chairs and so has a probability of 1/10 of choosing any particular chair, the next person has a probability of 1/9 of choosing any remaining open chair, etc. The probability of a particular seating configuration being taken is therefore $\frac{1}{10!}$ = 1 in 3.6 million. Clearly all of us are drowning in an ocean of statistical miracles!

It also bears mentioning that the aforementioned discussion regarding data mining does not apply. Bear in mind, we didn't roll the dice hundreds of millions of times in hopes of finding a particular dice-throw-sequence amongst the entire search space of rolls. We didn't search through millions of license plates to find a particular license plate. We didn't examine millions of rooms with seating configurations of 10 people to discover the configuration that we desired. We generated <u>one</u> sequence of dice-throws, we observed <u>one</u> license plate, we saw <u>one</u> seating configuration; nevertheless, we could document a miracle each and every time.

Ah, but surely you would object: those events hardly count as miracles at all, for while any 15-throw dice sequence certainly has only a tiny probability of occurring, certainly by throwing dice 15 times, <u>some</u> sequence would necessarily occur. Some outcome must necessarily occur, regardless of whether the statistical odds of that outcome are statistically miraculous.

Yet, consider your reaction if the statistical calculation was performed not after but rather <u>before</u> the event occurred. How would you react if I calculated the odds of a particular 15-throw dice-sequence, and <u>afterwords</u> proceeded to produce a dice-roll that corresponded to that very sequence? Or the probability of encountering a particular license plate alphanumeric combination before examining the car in front of me? Or the probability of a particular seat configuration before entering the room? Surely you would concur that something remarkable indeed has occurred.

What the above thought exercise demonstrates is how important it is <u>when</u> a statistical calculation is conducted. Statistical probabilities convey great meaning when calculated upon events that have yet to occur – they are far less meaningless regarding events that have already occurred. While we may be awash in an ocean of statistical 'miracles', the pertinent question is whether you are able to <u>predict</u> that such a miracle would occur – that is to say, prior to examining the evidence.

That Jim Collins discovered a strategy set and only afterwards calculating a 1-17-million probability is therefore no more statistically impressive than rolling a sequence of dice throws and then marveling at the infinitesimal probability of a subset of those rolls, say a string of 13 5's, resultant outcome. If Collins had calculated the odds of a specific candidate 'Good-to-Great' strategy set <u>prior</u> to examining the data, and <u>afterwards</u> validated that strategy set upon the data, he could legitimately claim that his strategy set is indeed unlikely to be the result of chance. But the infinitesimal odds that he cited from his statistician were calculated only *after* the fact, which renders them as unmeaningful as my aforementioned post-hoc calculation of the odds of a particular dice roll sequence.

4.2.4 Case Studies Conclusion

While 'In Search of Excellence' and 'Good to Great' do propose candidate theories, neither of them could be said to have been properly validated those theories, as none of those authors subject their theories to a fair risk of encountering and reporting disconfirming datapoints. Their methodologies generated theories directly from the datapoints they collected, and no fair validation test of a theory can be performed upon the very datapoints that generated the theory in the first place, for any theory induced from a set of datapoints must necessarily encompass those datapoints.

However, in fairness, it should be noted that neither Jim Collins nor Peters & Waterman lay claim towards being professionally trained social scientists and therefore possessing expertise in methodology, epistemology, or statistics. Neither do (most of) the other management gurus of their ilk. They are merely enthusiastic observers and students of business who believe in their research methodology and the accordant validity of their findings, and have likely never been taught to believe otherwise.

However, management academics – those holding PhD's and populating the tenure-track faculty ranks of the world's most eminent business schools – do indeed claim to be professionally trained social scientists. They do claim to possess expertise in methodology, epistemology, or statistics (and often times all three). Indeed, the entire legitimacy of their careers and of business schools proper is predicated upon such claims.[61] It is they, rather than management gurus, who therefore should understand that

[61] Whatever legitimacy that tenure-track business school professors may possess certainly does not stem from their expertise, or even personal interest, in real-world management. Harvard Business School Professor Rakesh Khurana noted in 'From Higher Aims to Hired Hands' (2007, p.311) that: "Many of the discipline-trained scholars joining business school faculties were not intrinsically interested in business…few were motivated in their research by a desire to examine the real problems that managers faced." In an infamous piece in Harvard Business Review, Warren Bennis and James O'Toole (2005) stated that: "Today it is possible to find tenured professors of management who have never set foot inside a real business except as customers" and that "while many conscientious researchers take it upon themselves to learn about the practice of business after they are tenured, there are few incentives for them to do so."

the theory validation process necessitates a fair risk of discovering disconfirming data. I therefore now turn to the validity of the research produced by management academia.

4.3 The Statistical Validity of Management Academia

If peer-reviewed articles in A-level journals are the heart of the modern-day academic management research enterprise, then surely statistical techniques are the heart of the empirical validation process of said enterprise. Whether statistical tools have gained favor for their ability to handle large datasets, to capture the inherent uncertainty of the business world, or merely to lend the veneer of authenticity to the craft of management as an academic discipline, one cannot deny the pervasiveness of statistical tools as the currently dominant empirical methodology. Large-n observational studies nowadays are the subject of the burgeoning statistical arms-race with researchers continually engaged in a game of econometric one-upsmanship. Smaller-n experimental studies likewise must sport appropriate statistical p-values to improve their odds of publication. Case studies and other qualitative empirical papers are increasingly becoming endangered species; the management academic literature has become increasingly inundated with "statistical fetishism"[62]. Therefore, the validation of the bulk of management academic papers must necessarily incorporate the tenets of statistical validity.[63]

4.3.1 Statistical Anomalies of 'Significant' Results Across the Literature

[62] Davis 2010.
[63] Although, to be sure, I also examine the validity of qualitative empirical papers in the following chapter.

To ascertain the statistical validity of the management literature , I gathered a dataset of 150 randomly selected empirical papers published in the last 30 years that relied upon statistical methodologies, all published in the 'usual suspects' of A-level journals: Academy of Management Journal, Administrative Science Quarterly, Strategic Management Journal, Organization Science, Management Science. I then took the hypotheses from each of those papers, and either took the p-values if they were explicitly listed (which rarely occurred), or back-calculated the p-values from the listed coefficient and standard error values. Papers that did not provide sufficient information to calculate a particular p-value – as some papers listed only p-value ranges, and others listed coefficient values without standard errors - were replaced with other randomly selected papers. However, only seven such papers were so replaced and therefore dropping those papers should not bias my results much. Upon encountering 'dual-headed' hypotheses pairs– where one hypothesis would predict one direction of a correlation, and an adjoining hypothesis predicts the exact opposite direction - I dropped one hypothesis of the pair, as those hypotheses-pairs are obviously mutually exclusive. I also counted sub-hypotheses (for example, Hypotheses that were enumerated as 1A, 1B, 1C, etc.) as individual hypotheses. If a paper reported multiple 'results' in support of a particular hypothesis, I conservatively took the result corresponding to the lowest p-value. I then plot the p-values in the following histogram and ask the question: <u>What's wrong with this picture?</u>

P-value histogram drawn from 150 randomly drawn empirical papers from A-level

journals (~1000 total hypotheses)

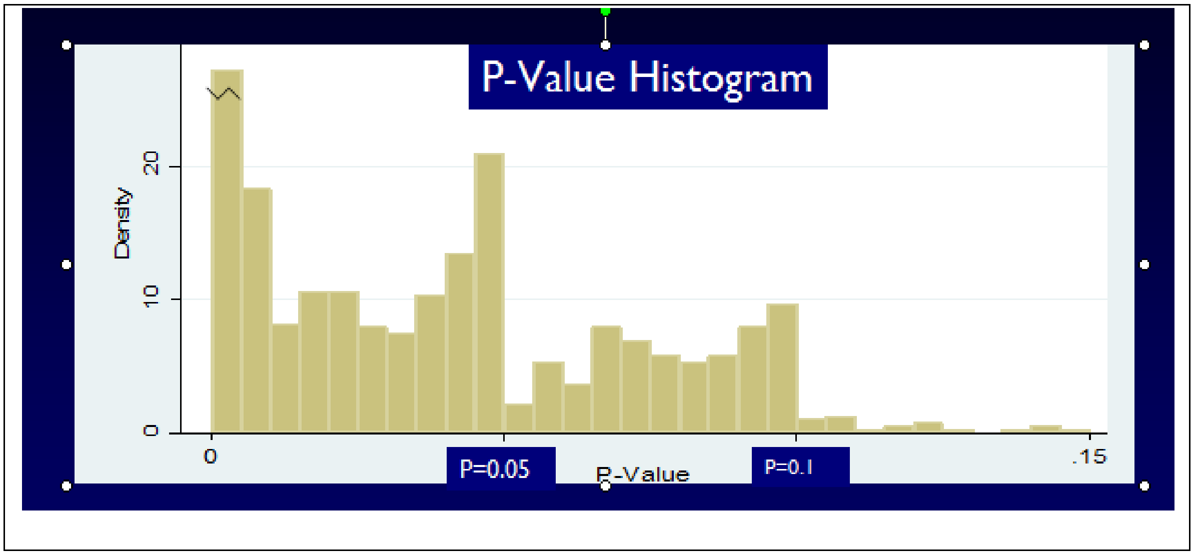


Figure 32

The statistical analysis of the above graph, and a bevy of other anomalous findings within the management literature are available upon request, which I think that technically minded readers would find especially entertaining. Nevertheless, even the non-technical reader can appreciate the above graph's denotation of the conspicuously large number of academic findings with p-values that barely attain the cut-off levels of statistical significance necessary for publication of 0.05 and 0.1. Much validation within the published literature is therefore questionable.

4.4 Conclusion

The upshot is that much – perhaps even most- of the academic management literature, in spite of its claims, has never really been validated in the sense that it has never been subject to a truly fair risk of disconfirmation. Rather, most theories that abound in the management literature are no more than candidate theories for which true validation is still lacking. What evidence of theory validation that does exists is provided by the original theory's authors themselves – for which the P-value histogram and Chekhov's Gun tropes renders such self-validation suspect.

However, one way that such theories might attain greater credibility is through a retraction process – where researchers continually revisit a particular finding with new data and new tools and retract those findings that fail to pass muster – along with a replication process where other researchers will conduct the key replication step. We now turn to these topics in the next chapter.

5 Causality

At this juncture, allow me to provide a review of the previous chapters. Properly validated theories must provide empirically falsifiable predictions that have been subjected to rigorous testing and validation. Tautologies such as the resource-based view or principal-agency theory cannot be empirically tested at all. The Peters & Waterman theory regarding the effect of the 8-themed strategy-set they identified upon 'excellent' performance could be tested in principle. But such testing was never properly conducted despite Peters & Waterman's assertions because they never invoked a proper control group. Similarly the strategy-set theories identified by Jim Collins in his entire oeuvre – not merely Good to Great but also Build to Last, Great by Choice and the remainder of his bibliography – have been openly admitted by Collins himself to be derived from the data itself.[64] Therefore the books of Collins, whatever their merits as potentially inductively drawn proto-theories, have never been subjected to even a single proper validation, let alone the rigorous replication that a validated theory requires. The same could be said for the numerous academic theories. Not only have numerous academic theories been tested only once, but such testing usually has been conducted by none other than theory's proponents themselves, with such testing having been characterized by certain scholars as displaying "eerie accuracy"[65]. Chekhov's Gun is evidently just as viable of a narrative trope within academic management theories as it is in Hollywood screenplays.

[64] See the methodology sections in the early chapters and/or appendices in each of Collins books.
[65] Oxley, Rivkin Ryall (2010).

Having said that, allow us to consider two particular hypotheses regarding myself. Those statements follow the tenets demonstrated above in that they provide empirically falsifiable predictions that can and have been validated repeatedly. Not only that, but I am confident that they will continue to pass any replication tests in the future with flying colors. The two statements are:

1) My daily waking time is positively correlated with the probability that the sun has risen.
2) My age is positively correlated with the size of the universe: the older I become, the larger the universe.

The above statements generate clearly falsifiable empirical predictions regarding my waking patterns compared to the location of the sun, or my age compared to the size of the universe. While you will have to trust me on this point, I can assure you that almost every day of my life, whenever I have awoken, the sun is indeed up. I have no reason to believe that that pattern will cease in the future. Similarly, one can surely validate that for every year of my life until today, the universe has expanded. Furthermore, the universe will almost certainly continue to expand for the rest of my life. Those statements have therefore survived repeated past empirical replication, and will likely survive continued future replication. If we also assume the non-obviousness of those statements[66], then those statements could be viewed as *candidate theories*.

However, few people – least of all myself – would assert that those statements are *causal* (or I'm a far more powerful celestial entity that I ever imagined!) To assert that my awakening actually *causes* the sun to rise is absurd. If any causality is occurring at all, it

[66] I would suspect that the expanding state of the universe might be non-obvious to those unfamiliar with modern astronomy.

occurs in the reverse direction: the rising sun awakens me. Similarly, surely nobody believes that my aging actually *causes* the universe to expand. The confounding variable of time both drives me to old age and causes the universe to expand.

Yet as fallacious as we may find any personal assertions of astronomical causality, much highly influential modern-day management guru discourse relies upon the same logical fallacy. Correlation is not causation, except apparently in the world of management gurus. Managerial gurus generally feel little compunction about asserting causality without demonstrating it. Perhaps more nihilistically, the managerial *consumers* of guru ideas seldom demand convincing evidence of causality.

Consider the 2013 Thinkers50 list of the top management thinkers as determined by a consortium of top business schools, firms, and leading business publications.[67] Roughly half of the Thinkers50 list consists of current or former tenure-track academic faculty; academia does indeed maintain high standards of causality, a point that I discuss further at the end of this chapter. However, the other half of the Thinkers50 list consists of management gurus or practitioners[68]. Earnest and, in some cases, as laden with work experience as they may be, not a single one of them has presented causal evidence within any of the theories espoused within the works that won them recognition with membership on the Thinkers50 list.[69] To the extent that they offer any evidence at all of

[67] The sponsorship of Thinkers50 can be found at http://www.thinkers50.com/about/

[68] The management practitioners and gurus of Thinkers 50 consist of: Martin, Tapscott, Goldsmith, Collins, Pink, Hewlett, Gratton, Buckingham, Hamel, Lafley, Friedman, Erickson, Liu, Heath, Sandberg, Haque, Goleman, Chowdhury, Trompenaars, Zook, Kakabadse, Wiseman, Ready, Wang. Full list available at http://www.thinkers50.com/t50-ranking/2013-2/

[69] I define those works to be those listed within the blurbs within the Thinkers50 description of those gurus/practitioners.

their assertions, that evidence is merely correlational in nature, with no clear causal interpretation.

Nor is the Thinkers50 list particularly unusual in this respect. An examination of every article written by gurus (hence non-academics) promoting a theory within the Harvard Business Review published from 2009 to 2013 reveals that not a single such article provided clear evidence of causality supporting the espoused theory, whether within the body of the text or within any supporting research that they referenced. The same is true for a random selection of the top management guru books as ranked by popularity within the business section of Amazon.com. Evidence of causality seems to be as rare as hen's teeth.

To be fair, demands for causality sometimes serve as rhetorical disguises for sheer disbelief of an espoused claim. For example, suppose that somebody proposed the theory that gender diversity improves profitability. Yet the evidence proffered is merely correlational data that highly gender-diverse firms tend to be highly profitable than vice-versa. It would then be entirely correct to declare that such correlational data by itself fails to establish causality. Yet those who are implacably opposed to the linkage between diversity and profits may simply invoke the call for causality as nothing more than a cheap way to discredit the claimant, under the unspoken notion that the demand for clear causality would be impossible to answer. This is not only unfair to the person proposing the theory, but more importantly is also an etiological fallacy. While causality may be impossible to *irrefutably* prove within the social scientists and perhaps even within the natural sciences, methodologists and epistemologists have nevertheless devised a number of techniques with which one can provide some evidence of causality. A more fair criticism would be to point out that the correlational evidence linking diversity and performance does not demonstrate causality from diversity to performance, *but then propose some methods by which such causality could be established*. That then leads to the question: how can causality be established anyway? We are now at the following underlined line of the Christensen-Sundahl pyramid.

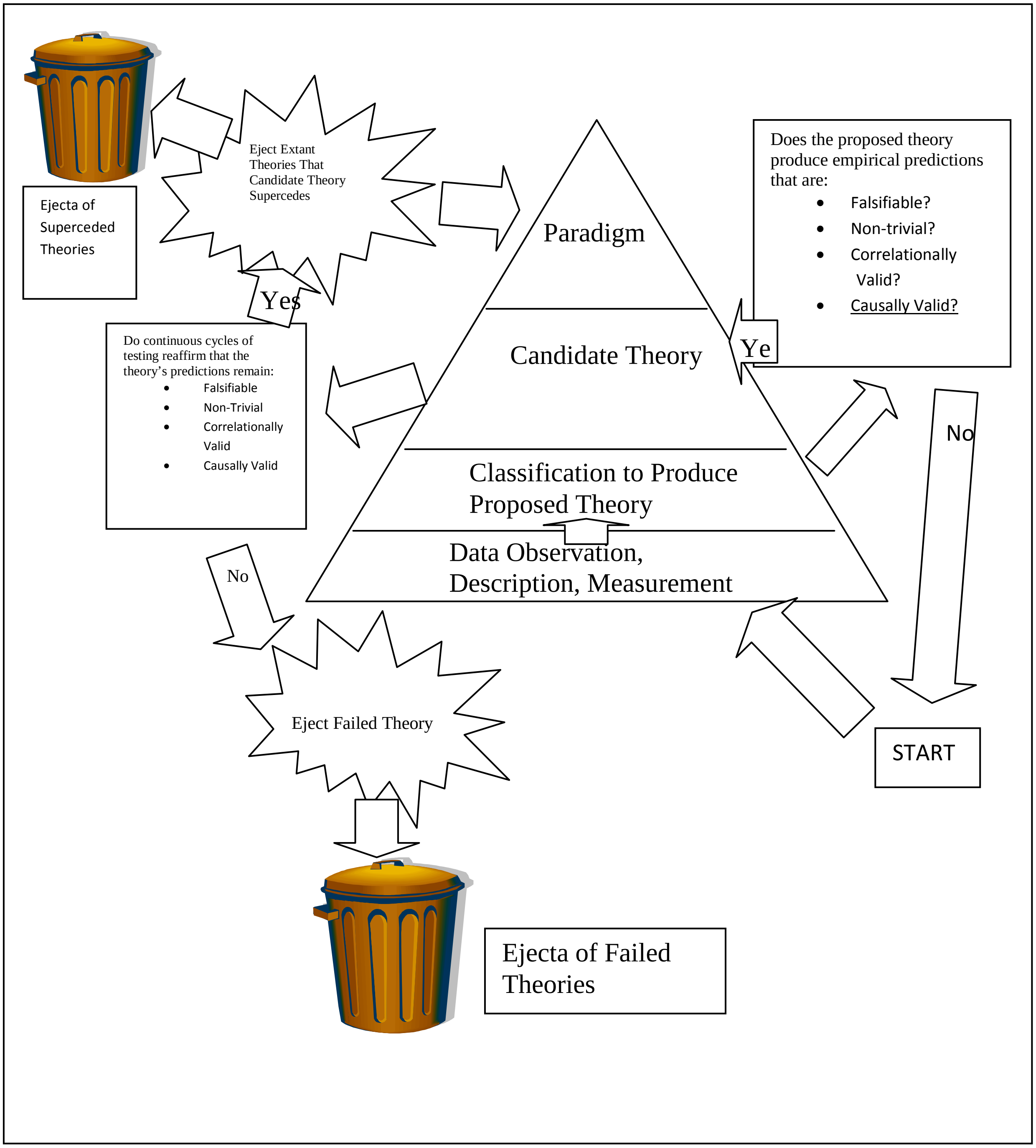

Ejecta of Superceded Theories
Eject Extant Theories That Candidate Theory Supercedes
Yes
Paradigm
Candidate Theory
Classification to Produce Proposed Theory
Data Observation, Description, Measurement
Does the proposed theory produce empirical predictions that are:
Falsifiable?
Non-trivial?
Correlationally Valid?
Causally Valid?
Ye
No
START
Do continuous cycles of testing reaffirm that the theory's predictions remain:
Falsifiable
Non-Trivial
Correlationally Valid
Causally Valid
No
Eject Failed Theory
Ejecta of Failed Theories

Figure 33

5.1 Why is Causality Important?

Before delving into the details of how causality is established, one might reasonably ask – why is causality important in the first place? After all, philosophers of science place little emphasis upon causality per se. A proposition that proved to be highly reliable and non-obvious might well satisfy the Popperian definition of a scientific truth, even if the underlying causality of that proposition remains obscure. Indeed, most of the greatest scientific discoveries known to mankind are causally opaque. To this day, nobody knows what 'causes' gravity, relativity, quantum mechanics, or electromagnetism. Yet shouldn't the mere fact that objects fall down without necessarily needing to understand gravity's underlying causal mechanism be sufficient? Similarly, might the fact that certain variables within the realm of management are reliably correlated be sufficient without necessarily needing to understand their underlying causality?

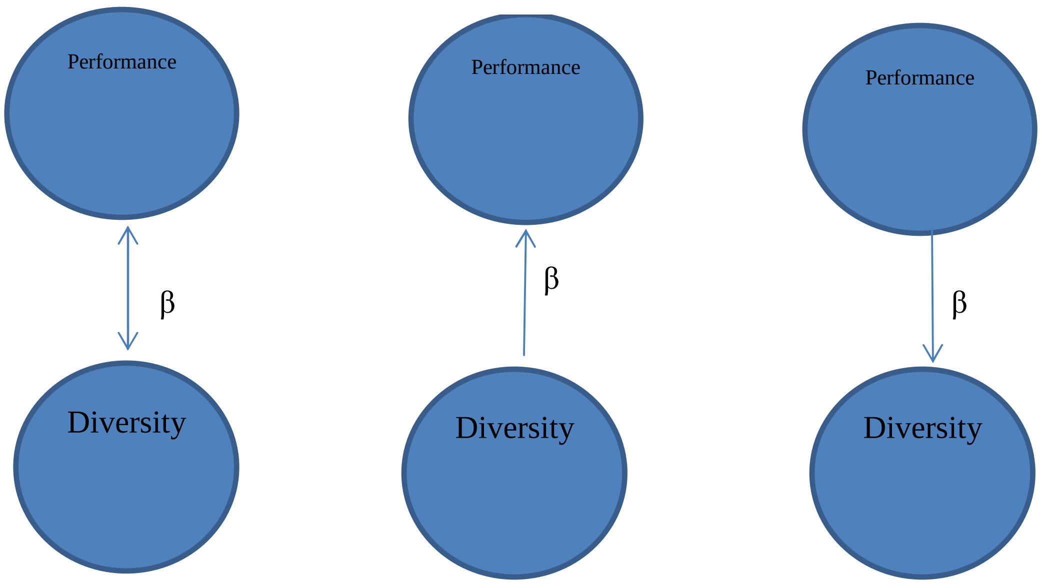
Such a state of knowledge might indeed be sufficient if the field of management was satisfied with being merely a pure science. However, I would argue that the field of management is more than a science, but rather akin to a field of engineering. Just as engineers are unsatisfied with simply understanding the electromagnetic force but also want to manipulate it to design computers and smartphones, so too do managers want not only to understand managerial variables but also want to manipulate them to improve organizational performance. Doing so requires an intimate understanding of causality. For example, it is not particularly salient to discover that certain managerial variables are merely *correlated* with improved performance if those variables do not actually *cause* improved performance. Certainly MBA students and management practitioner's community – both the audience for and ultimate financial supporters of management research – would be unimpressed by mere correlations alone. Management academia ultimately is ultimately judged by its ability to deliver not only Popperian-style reliable & not-obvious correlations, but also causal effects.

5.2 The Establishment of Causality Through Time Sequences and Leads/Lags

To be sure, causality is a most elusive quarry. The determination of causality requires additional painstaking steps taken after a correlation has been determined. Given, say, an established correlation between gender diversity and firm performance, how would one establish causality? Perhaps gender diversity indeed drives firm performance, or perhaps highly-performing firms can afford to boost their diversity. Perhaps each causes the other. Disentangling the two variables might therefore seem to be a hopeless task. Must we then be resigned to merely documenting correlations only, with no hope of ever establishing any evidence of causality?

To clarify the discussion, the following figure presents three 'directed graphs' that displays correlational relationship regarding the data that we have compared to the causal relationship that we might want to know. The term 'directed' refers to the usage of arrows to designate either a correlational or causal relationship. The diagram on the left represents the mere correlation between gender diversity and firm performance, where the connection between the two is represented by a double-headed arrow that indicates a two-way correlation with an unclear direction of causality. The other two diagrams represent the two potential causal scenarios between the variables, with single-headed arrows indicating the direction of causality. The middle diagram indicates that diversity is causing performance, whereas the right diagram represents vice versa. Note that the left diagram by itself implies that either of the two other diagrams might be true; we have no method to distinguish between the two at this time.

Figure 34

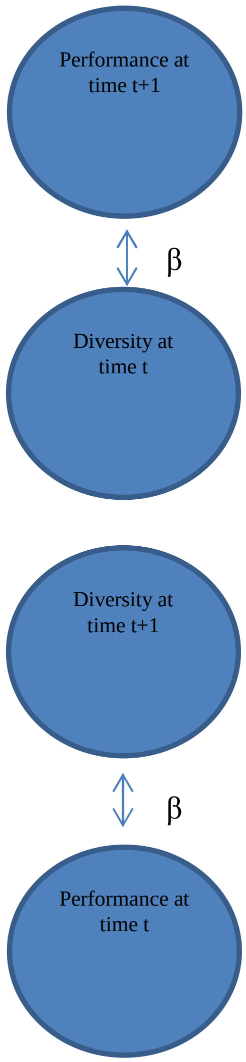


Left Diagram Indicates a Known Correlation of value β between Performance and Diversity. Middle and Right Diagrams Each Represent a Potential Causal Direction of value β. How Can We Infer From Our Knowledge of the Left Diagram to Either the Middle or Right Diagram?

The above directed graph indicates a potential solution towards establishing the proper direction of causality. Rather than examining correlations between diversity and performance *contemporaneously, a* technique that provides no information about the direction of causality, one can instead investigate correlations between measurements of those variables obtained at different times. For example, if *prior* levels of high diversity are correlated with *subsequent* levels of high performance, then one might interpret that as causal evidence that diversity does indeed cause performance.[70] Likewise, if prior levels of high performance are correlated with subsequent levels of diversity, then one might interpret that as evidence that performance may cause diversity. One could then calculate the correlation β between the lagged and leading variables; β would have the interpretation that a unit change of the lagged variable would cause a change of β units of the leading variable.

Granted, one still cannot be entirely sure whether the direction of causality truly points in the proposed direction as further explained in this chapter. Nevertheless, ‘lead/lagged’ correlations provide some evidence regarding the proper direction of causality.

[70] The underlying assumption is that anticipatory effects are not at play: that current performance does not change *in anticipation* of a future change in diversity. Such anticipatory effects could be accounted for by the use of proper psychometric control variables as discussed in the next section or through the use of sufficiently long leads/lags. Anticipatory effects are also generally important only when dealing with market valuation variables and are unlikely to be important in most internal firm operational variables.



If a statistical correlation β is found between *lagged* diversity and *leading* performance, then one might interpret that as diversity causing performance by a value of β.

If a statistical correlation β is found between *lagged* performance and *leading* diversity, then one might interpret that as performance causing diversity by a value of β.

Figure 35

Two natural questions one might ask are what duration and what units should the time increments be between the variables? This question is important only if you desire to obtain a precise answer regarding the dynamic effects of *how quickly* a particular causal effect is felt. One can investigate correlations between, say, diversity at a particular time t with performance at various later times: t+1, t+2, t+3, etc. The units of time could be in days, months, quarters, years, or whatever other time unit is believed, whether based on theory or on practical experience, to be the most appropriate time frame towards capturing the causal effect. For example, any causal impact of diversity upon performance will almost surely not make itself felt within a matter of hours or days, but might plausibly do so within a matter of months or quarters. One could then investigate correlations between today's diversity a performance measured in subsequent months/quarters.

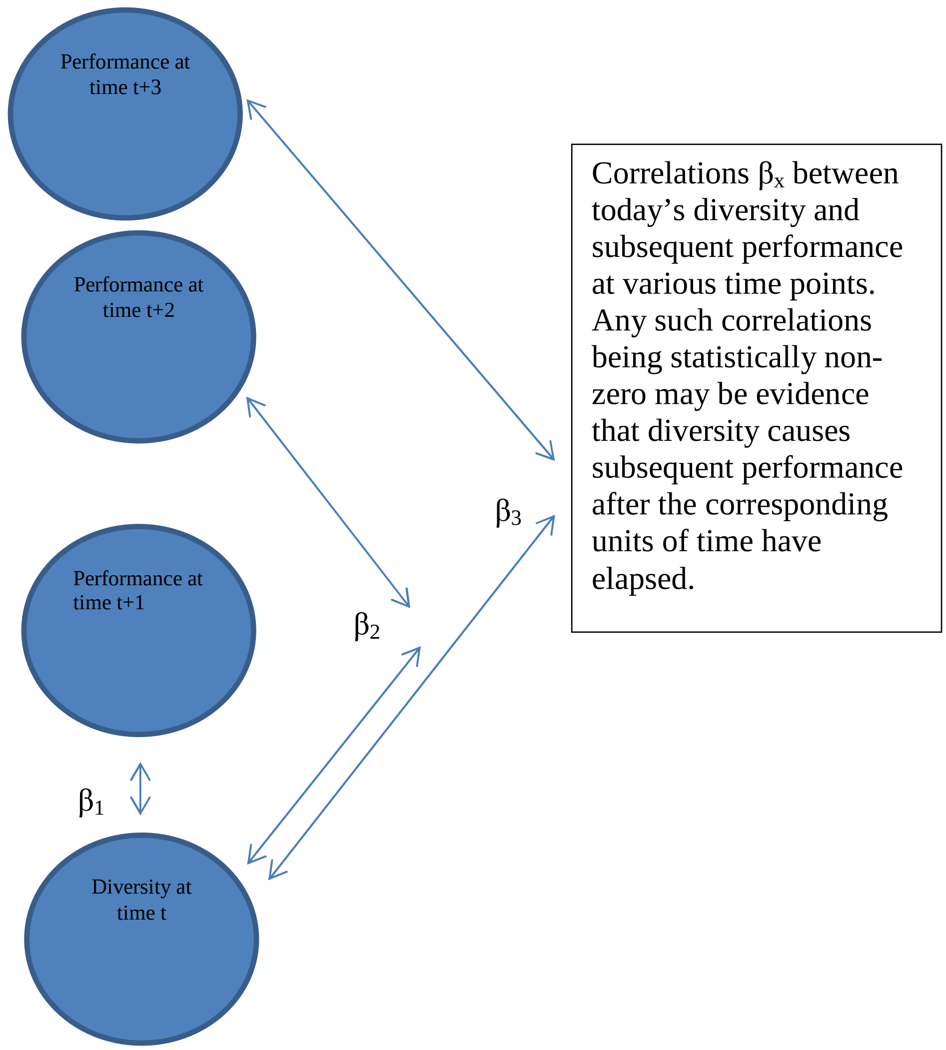


Figure 36

Granted, such inferences of causality through the leveraging of correlations between leads/lags between two variables obviously requires collecting data regarding those variables over time. Any single-shot datasets that Gallup and other polling agencies collect are therefore uninformative. But most polling agencies do not collect mere single-shot datasets but instead collect numerous polls periodically over time. It is therefore trivial for researchers to correlate past values of certain variables collected in prior polls with subsequent values of other variables collected in later polls. The simplicity of this task makes it all the more surprising that more management gurus do not conduct this analysis in an effort to infer causality, or why more consumers of management guru literature do not demand that they do so.

5.3 Control Variables

To be sure, the entire above analysis assumes that reverse causality is the only potential threat to causal interpretation that must be addressed That, however, is seldom true. The more common threat is that of a *confounding variable*: a third variable that has a causal effect upon both of the two target variables in question, hence driving a spurious correlation between those two variables that have no causal interpretation. For example, perhaps well-educated managers (e.g. those attending top MBA programs) have a causal effect upon current company diversity at time t by instituting more outreach and recruiting programs and those well-educated managers also have a causal effect upon subsequent firm performance at time t+1 through savvy management tactics. The positive correlation between current diversity and subsequent firm performance is then confounded by that third variable - the correlation β has no causal interpretation. Manipulating diversity alone would not modify subsequent performance by a value of β. Indeed, manipulating diversity might not cause any change in subsequent performance at all, as perhaps the entire correlation between diversity and subsequent performance is driven by managerial education.

Education of managers is driving both diversity and subsequent performance. Any measured correlation β between diversity and subsequent performance is therefore spurious and has no causal interpretation.

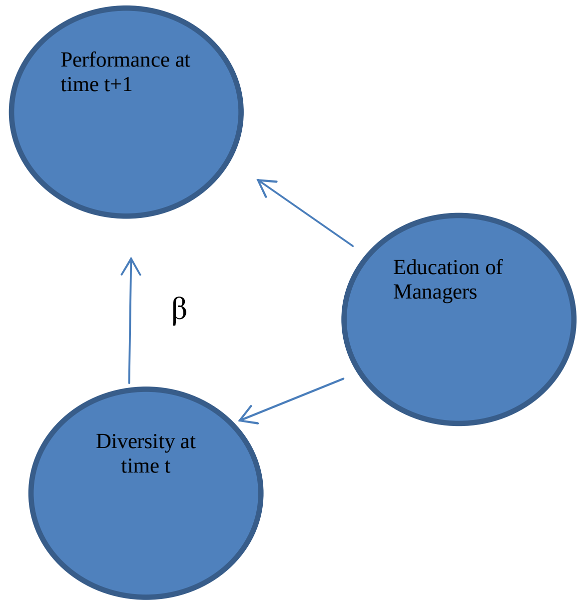


Figure 37

The effect of that third variable can be removed by *controlling* for it. I illustrate such control by placing a box around the third variable. If a correlation β is established between diversity and subsequent performance after controlling for the education of management, then one may be able to interpret that correlation causally.

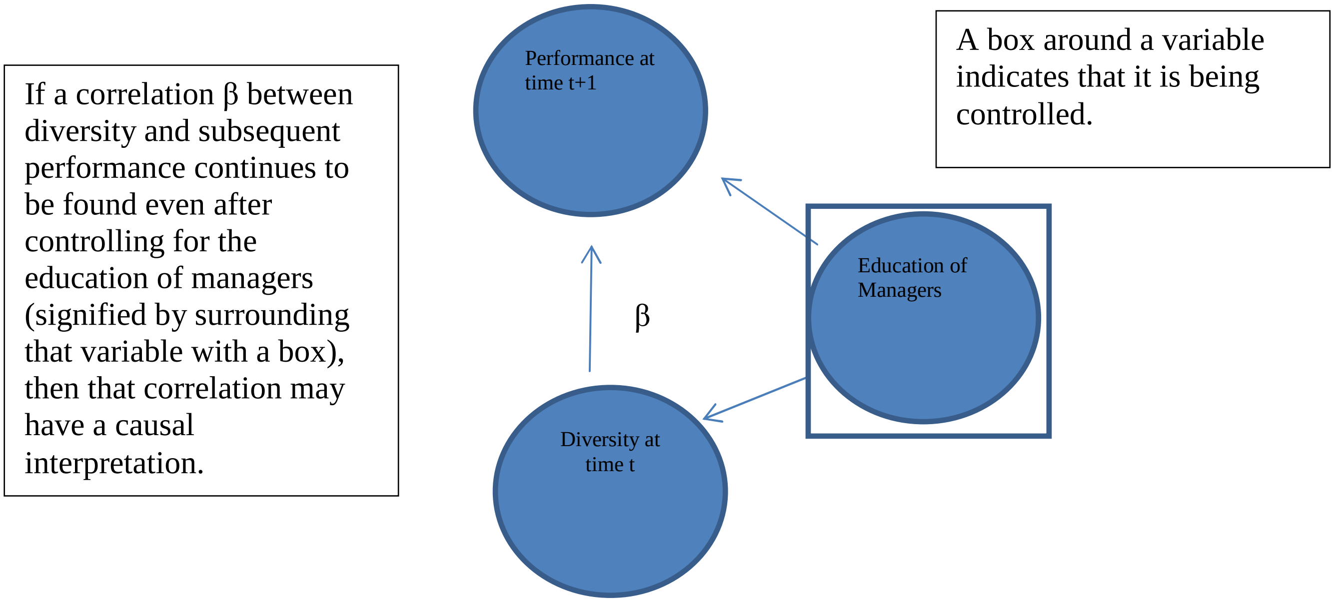
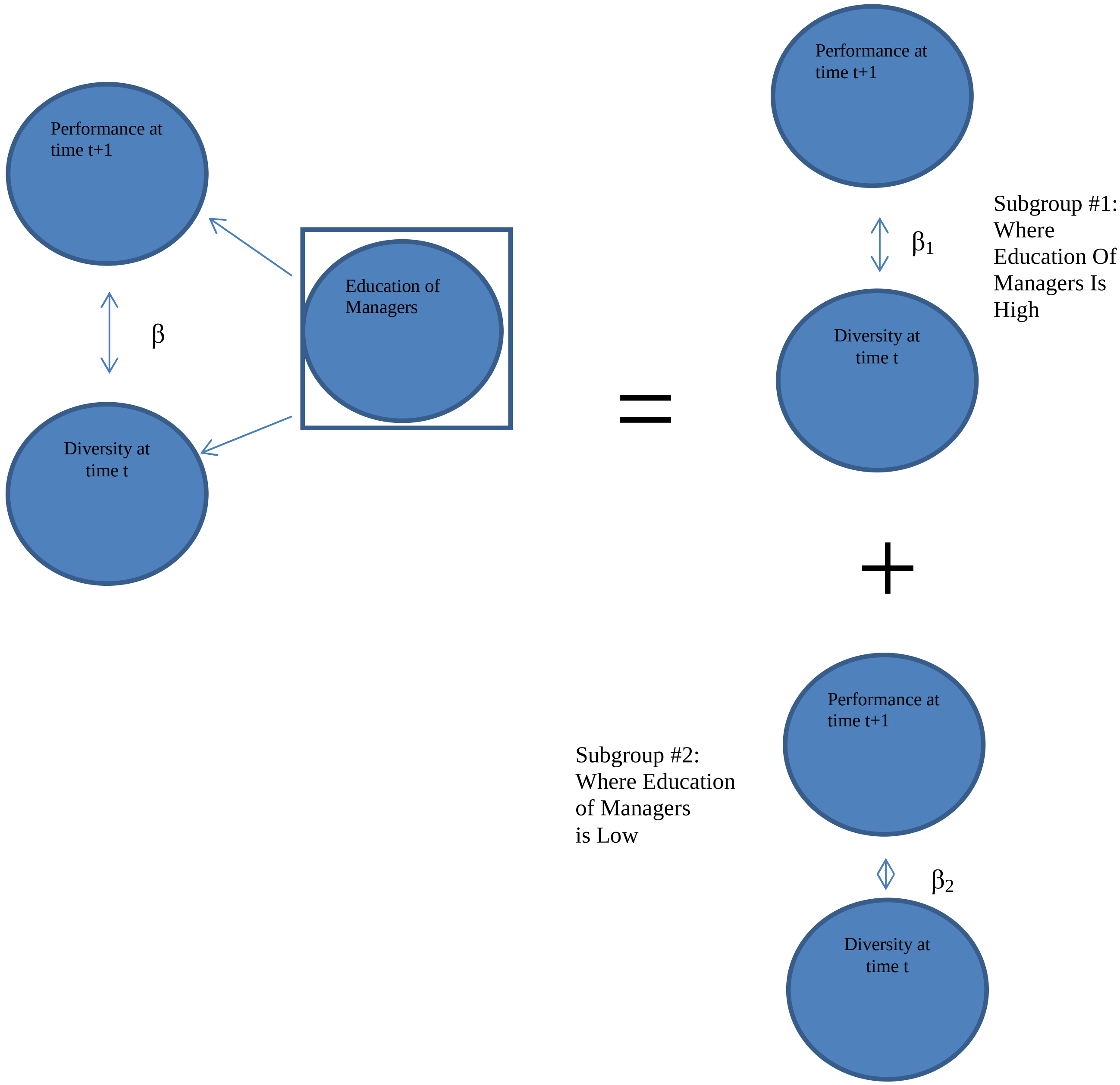


Figure 38

Controlling for variables is a well-worn technique that abounds throughout the research literature. Yet the pedagogical meaning of controlling for a variable is often times obscure, particularly to the layman. Econometric texts might characterize controls as act of subtracting off the effects of the confounding variable from both the dependent and independent variables in question[71]. Such descriptions likely foster confusion rather than shed light, not least because one might reasonably wonder how/why certain effects would need to be subtracted at all, as well as any potential interpretation regarding whatever has been subtracted. Those texts also invariably assume that one is conducting regression analysis – for which the subtracting of effects indeed is mathematically logical – rather than matching analysis, when matching and regression are largely homologous.[72]

[71] See Wooldridge 2010, Kennedy 2009.
[72] Angrist & Pischke.

Hence, the most intuitive interpretation of ‘controlling for a third variable’ is that one can imagine investigating each of the correlations between the two variables of interest at specific values of the third variable. For example, if we assume that the Education of Managers variable is dichotomous (that is, can be either “High Education of Managers” or “Low Education of Managers”) then we calculate two subcomponent correlations: that between diversity and subsequent performance for “High Education of Managers” and that between diversity and subsequent performance for “Low Education of Managers”. The two correlations are then blended together, usually through a weighted average (with the weights assigned by how many data observations are in each of the subsets of High/Low management education) to obtain an overall correlation between diversity and subsequent performance. That overall correlation may have a causal interpretation. I illustrate this below.



The left diagram, where education of managers is controlled, is equivalent to investigating the subgroup correlations between diversity and subsequent performance within each of the groups (high/low) that education of managers be. The two calculated subgroup correlations β_1 and β_2 are then blended (usually through a weighted average) to obtain the overall correlation β between diversity and subsequent performance. <u>That overall correlation may have a causal interpretation</u>.

Figure 39

The efficacy of this technique stems the fact that education of management is being held constant, while each of the two sub-component correlations are calculated between diversity and subsequent performance, which means that education of management cannot be driving that subcomponent correlation. If those two subcomponent correlations are then blended together to form an overall (statistically) non-zero correlation, then one may interpret that as the *overall* causal effect of diversity upon subsequent performance across constant values of management education.

Similar logic applies if management of education is not dichotomous but rather trichotomous or follows any other discretization scheme. The same logic also applies for continuous control variables, although one probably might then invoke the econometric textbook logic of subtracting off the effect of management education. However, the same intuitive logic expressed above continues to apply: one can simply imagine calculating and blending subcomponent correlations calculated at each of the various values of the control variable.   Similar logic also applies with multiple control variables: one simply has to imagine controlling for the levels that each of the control variables can take, calculating subcomponents correlations for each, and then blending them all to obtain the overall correlation.

5.4 Exogenous-Shocks

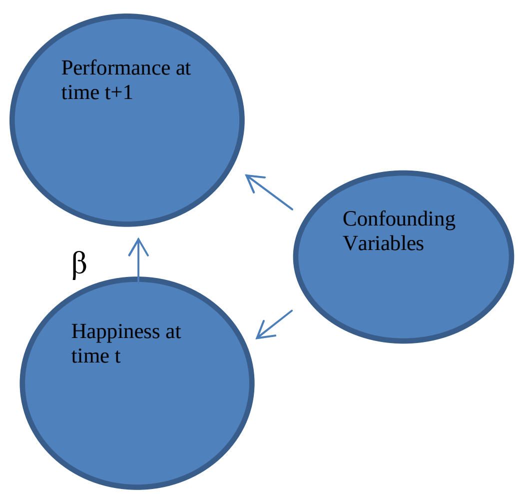
The savvy reader might already have ascertained the core problem with relying upon control variables to establish causality. Managerial phenomena are seldom causally established through one or even a handful of control variables. Many of them are likely have dozens, even hundreds of control variables impinging confounding any causality. Worse yet, many such control variables may not be directly observable to the researcher or may not even be currently known such that the researcher could possibly control for them.

Hence the rise of the instrumental-variable/exogenous-shock/natural-experiment/randomized-experiment, hereafter shortened as just the 'instrumental variable methodological strategy - in stark contrast to the requirement of blocking all confounding pathways that impinge upon both the cause and effect variable as the control variable strategy requires – instead requires leveraging a variable that affects only the purported causal variable while having no direct effect upon the effect variable except through the causal variable. Such an variable is deemed an 'instrumental variable', and they are often times discovered through investigating exogenous shocks and natural experiments: sources of randomness that are applied not directly through researcher manipulation, but rather through outside forces that the researcher did not control but which the researcher can nevertheless exploit to ascertain causal effects[73]. Note that truly randomized experiments, whether conducted in the field or in the lab, can be considered a subset of the instrumental variable framework where the instrument consists of a randomized treatment directly applied by the researcher.

Given that the above paragraph is undoubtedly confusing, an example is in order. Suppose that we wish to establish the well-trod causal effect of employee happiness upon performance. Such a relationship is arguably confounded. Perhaps performance actually causes employee happiness, or perhaps other variables such as economic booms or better management boost both employee happiness and performance. One could attempt to identify the causal effect by correlating happiness with subsequent performance after controlling for all confounders. However, you may not even know what some of those confounding variables are, let alone have the ability to observe and control for all of them.

Figure 40

A causal interpretation between happiness and subsequent performance could be ascribed to β if all confounding variables could be controlled. However, such control may be implausible, for doing so presumes that you know what all confounders are, and have the data to control for them.



[73] Certain esoteric technical differences also exist regarding the terminology amongst instrumental variables, exogenous shocks, and natural experiments, none of which impinge upon the discussion here. Interested readers are directed to Pearl 2009, Angrist & Pischke 2005, and Morgan & Winship 2004.

Alternatively, one might leverage the ‘instrumental variable’ of a state lottery. Assuming that happiness is correlated with winning the state lottery, one might observe employees who won the lottery (but with insufficient winnings to incentivize them to quit their jobs, hence we might investigate some of the minor lottery winnings of a few thousand dollars), one could then track the subsequent productivity of those lucky employees over time, and compare them to the productivity of employees who also played the lottery and lost. Under the assumptions that lottery winners are indeed determined randomly and that winning the lottery has no direct effect upon employee performance one could infer the causal effect of happiness – as instigated by lottery winnings – upon performance. Winning the lottery therefore serves as an instrumental variable as finding the correlation between winning the lottery and employee happiness ($\beta_{instrument}$) then allows you to solve for the ‘unconfounded’ portion of β.[74]

[74] The necessary equations can be found in any standard econometrics textbook.

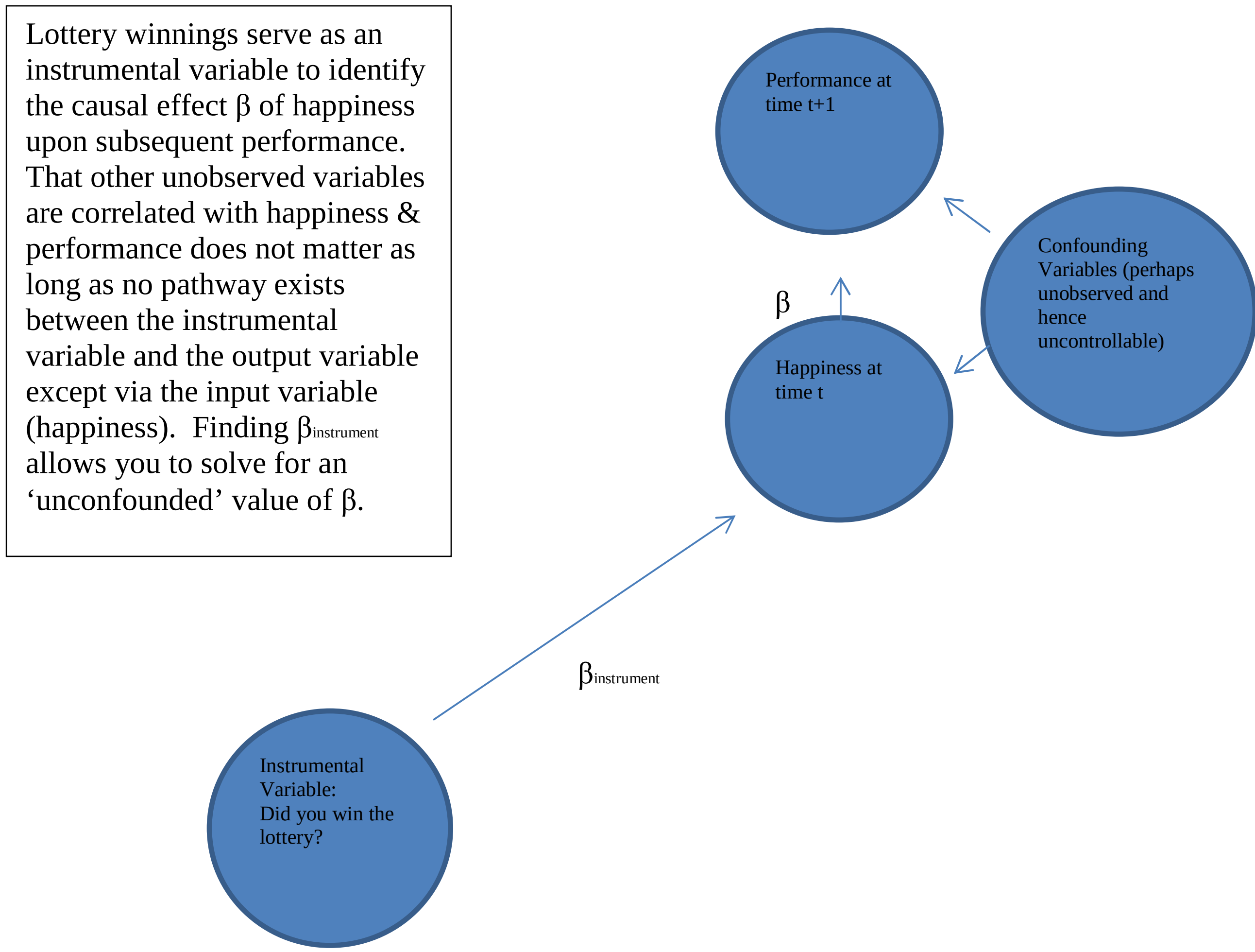


Figure 41

To be sure, the assumptions of instrumental variables are most stringent indeed. Any alternative causal pathway between the instrument and the effect variable other than through the causal variable will invalidate the instrument.

If an alternative pathway exists either between winning the lottery and performance, or between the determinants of who win the lottery and performance (i.e. if winning the lottery allows employees to purchase training or education which then improves their future performance), then the instrumental variable strategy is invalidated. The calculated value β has no causal interpretation.

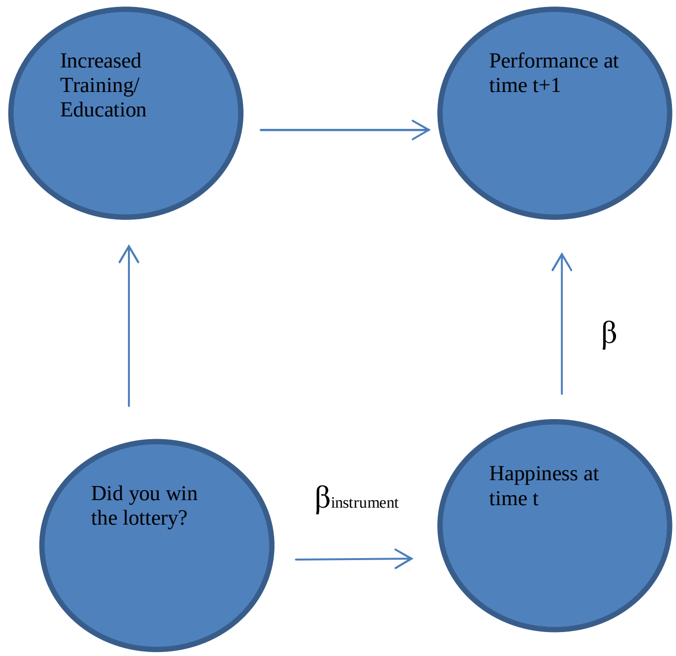


Figure 42

Granted, if an instrument's alternative pathway is known, then one could invoke the aforementioned techniques of controlling for that pathway.   However, one should note that instrumental variables may have multiple pathways to the outcome variable and all of them would need to be controlled.   Otherwise, the requirements of the instrumental variable strategy are violated and β has no causal interpretation.

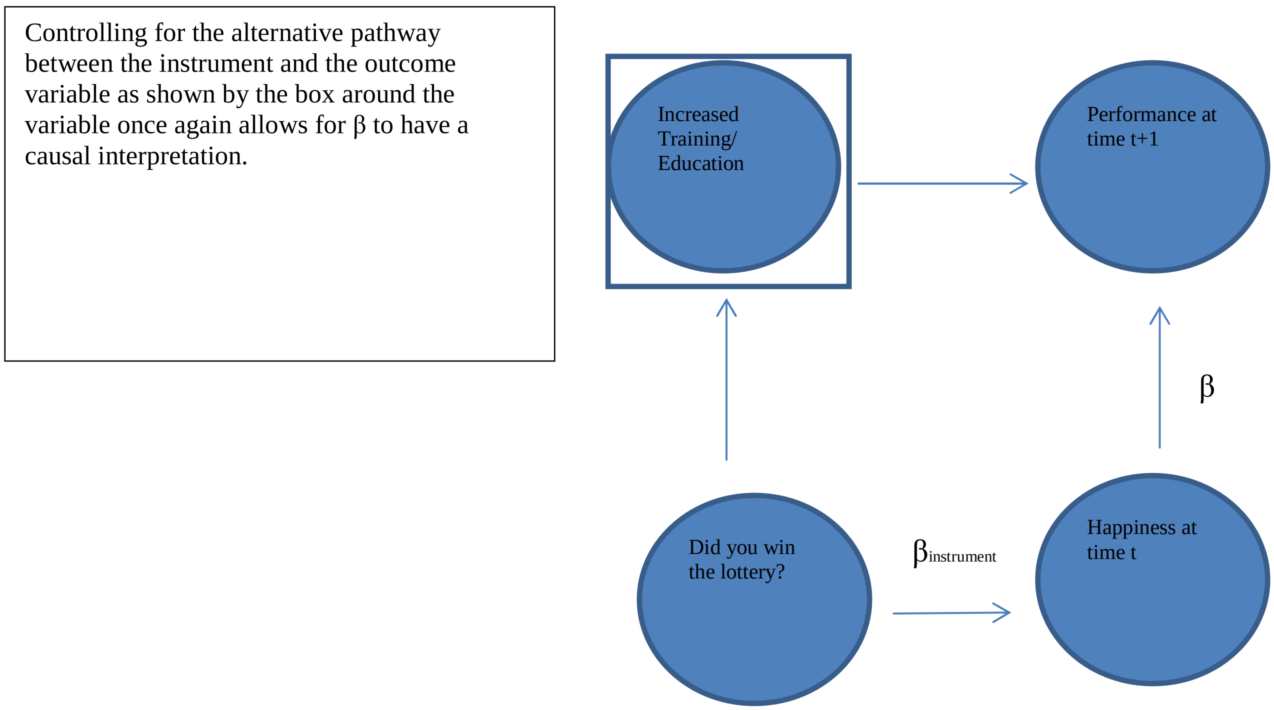


Figure 43

The stringent assumptions necessitated by the instrumental variable strategy have thus far reduced the usage of the strategy to niche purposes. However, that niche is growing as the burgeoning availability of data provides more opportunities to leverage new instruments, either by unearthing formerly unknown variables which can serve as instruments or by converting known variables whose status as instruments were vitiated because of alternative pathways into usable instruments by uncovering sufficient data to block said pathways. Instruments may therefore arguably the most promising empirical causal strategy in management academia going forward.

5.5 How 'Causal' is the Published Literature?

This chapter has merely scratched the surface of the vast realm of literature regarding empirical causality that has erupted in recent years. Indeed, the 'Causality Revolution' within management academia – and the social sciences in general- has taken hold only during the 15-20 years. Causality was arguably first established as a legitimately rigorous philosophical concept from a social-science standpoint through the work of Judea Pearl via his publication of a series of papers in the 1990's and culminating with his masterpiece 2000 theoretical tome fittingly entitled Causality. Pearl was followed by series of eminent scholars - Josh Angrist, Jorn-Steffen Pischke, Guido Imbens, Christopher Winship, and Stephen Morgan just to name some of the more prominent members – who continue to provide key practical methodological tools with which management researchers can calculate and elucidate causality. A comparison of the methodological tools used by researchers before and after the Causality Revolution took hold is akin to night and day. No longer must researchers be resigned towards stumbling through the fog of correlation lacking any semblance of causality that was typical of pre-Revolutionary research. Today's researchers can avail themselves of a smorgasbord of techniques such as forward/backward-instruments, regression discontinuities, bounds analyses, dynamic panels, and other methodological exotica.

Yet despite the vertiginous methodological changes that have occurred within management academia, one aspect of management academia has not changed a whit.

Such stasis is reflected in my dataset: of the 150 A-level management papers, how many conceded that their proposed theory failed to demonstrate causality?

Answer: None

That is, not a single time did the authors concede that their findings failed to establish causality for their own theory. In some cases, causality was never mentioned at all - even if it was heavily implied. In other cases, causality of the authors' theory is always proven, or at least never disproven. To the extent that causality is ever challenged at all – a conspicuously rare outcome – that lack of causality is always regarding somebody else's theory, never their own. Regardless of which methodological tools any researcher uses, causality of their own theory is apparently never disproven.

To be sure, the establishment of causality is necessary to convert a proposed theory to a candidate theory as delineated by the Christensen-Sundahl pyramid. However, it is rather curious indeed that so many theories published in A-journals – indeed, apparently all of ones comprising my dataset - manage to traverse that step. Not a single time did authors concede that their theory's causality is unsupported. Moreover, couple that evidently perfect causality with the findings of the previous chapters that no published paper ever concedes that their theories are empirically untestable, are obvious, or have not been properly empirically validated – despite much evidence to the contrary – and apparently every single publication represents a brand-new candidate theory. Indeed, management academia is churning out candidate theories like clockwork.

Granted, management academia's prolific theory generation might not be a problem – and indeed might be healthy – if theories that are later found faulty are then removed. It is to this topic that we now turn.

6 Continuous Testing

The past chapters demonstrated that management research – at least as expressed within the A-level journals – never seems to concede that theories are tautological, trivial, invalidated, or not-causal (despite evidence to the contrary). Yet one might argue that that is an entirely appropriate role for the journals to play. The purpose of the journals is to serve as a filter; they rightfully refrain from publishing findings that are tautological, trivial, invalid, or non-causal.

Yet one must remember that science progresses through a process of continuous testing. Theories gain credence not merely because they survive an initial set of tests sufficient for initial publication, but only insofar as their ability to survive future repeated attempts at falsification. In Popperian terms, theories must be subjected to merciless and perpetual reverification not only with respect to the methodological tools and data of the past, but also with respect to future data and methodological tools. Indeed, theories gain greatest credence when they demonstrate empirical compatibility with datasets and tools hitherto undreamed of by the original proponents of the theory.

We can likewise apply that level of scrutiny to management theories that have attained candidacy status simply by re-applying the criteria that allowed them to achieve candidacy in the first place, but confronted by new evidence and tools. Do those candidate theories remain empirically testable – that is, have they become tautological in light of new evidence? Should they now be rejected for producing results that are now understood to be obvious, even if they were not understood to be so at the time of publication? Do those theories remain standing under the withering fire of new datasets and methodologies? Do the *causal* tenets of those theories remain standing?

To be clear, the answers to each of those questions remain necessarily subjective. I would not presume to be so omniscient as to know whether any particular management theory has lost its candidacy theory and should therefore be cast into the dustbin of history. The community of management researchers as a whole must decide whether particular candidate theories ought to be discarded. That community must determine that certain theories are no longer useful for providing reliable, non-obvious, causal results and therefore should no longer be utilized by both academics and practitioners alike. Such rejection also serves as a key indicator as to the progress that the management field – or any scientific field for that matter – is achieving. Scientific progress is popularly conflated with the successful testing of a particular theory. That is merely one side of the coin. Scientific progress also requires the *failure* of theories. History is replete with thousands of scientific theories that had been once believed and passed initial testing, only to be subjected to continued testing and found wanting. It was once widely believed that the universe revolved around the Earth, that the heart was where a person's mental capabilities resided, that vision was effected by light waves emitted from a person's eyes, that heavier objects fell faster than do lighter objects. We are at the stage marked in underline of the Christensen-Sundahl pyramid.

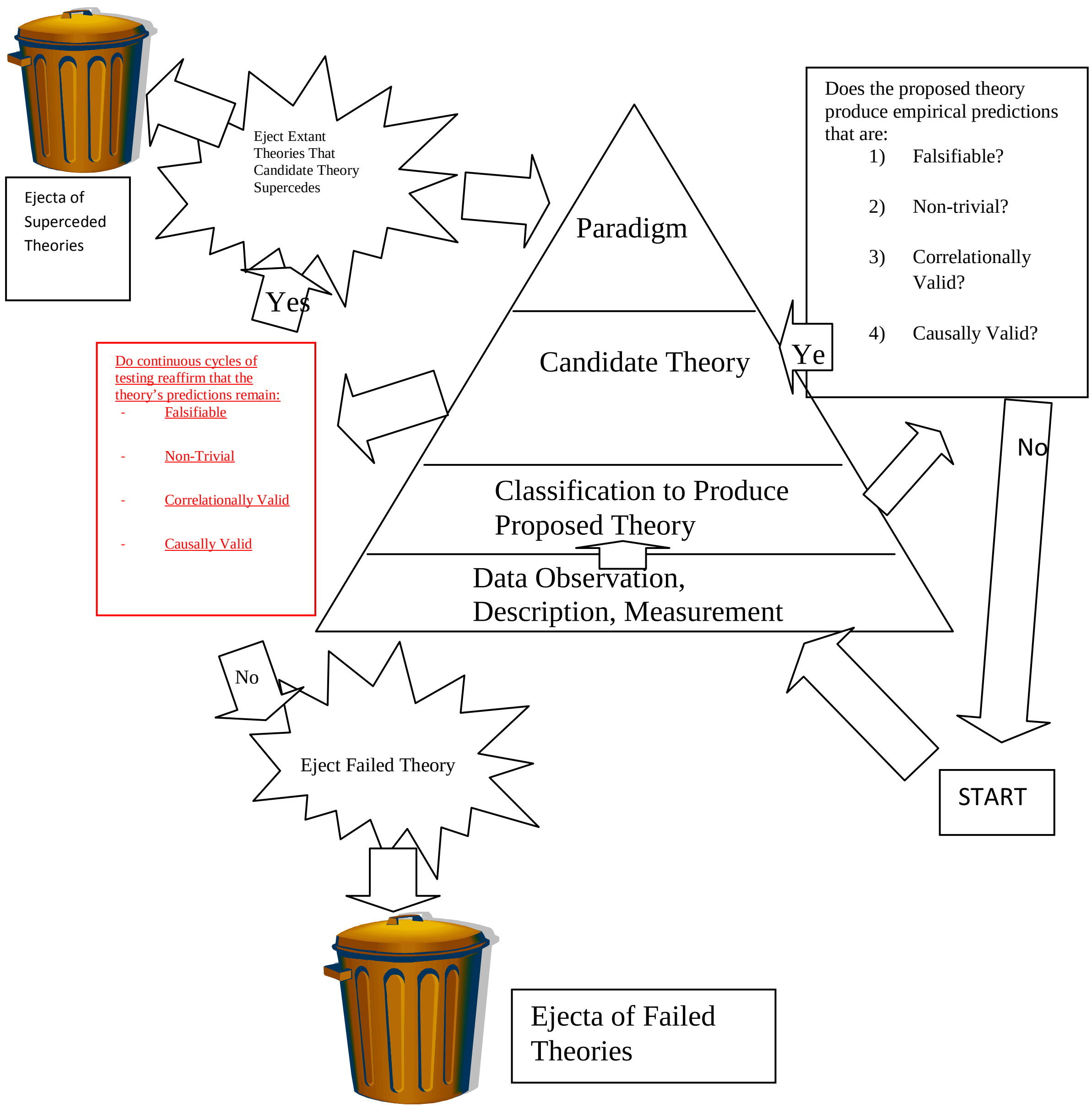

Ejecta of
Superceded
Theories
Eject Extant
Theories That
Candidate Theory
Supercedes
Yes
Paradigm
Candidate Theory
Classification to Produce
Proposed Theory
Data Observation,
Description, Measurement
Does the proposed theory
produce empirical predictions
that are:
1) Falsifiable?
2) Non-trivial?
3) Correlationally
Valid?
4) Causally Valid?
Ye
No
Do continuous cycles of
testing reaffirm that the
theory's predictions remain:
- Falsifiable
- Non-Trivial
- Correlationally Valid
- Causally Valid
No
Eject Failed Theory
START
Ejecta of Failed
Theories

Figure 44

6.1 Are Management Theories Ever Later Ejected for Being Tautological?

A discussion of the history of science is in order. Consider the discovery of electromagnetism. Since the days of antiquity, man has interacted with electricity, even if only to be mystified by its nature. Ancient philosophers pondered the seemingly divine and destructive nature of lightning. Ancient mariners routinely encountered the coronal phenomena now known as St Elmo's Fire, the ancient Greeks believing that its appearance was a blessed sign from the legendary divine/mortal twins of Castor and Pollux. The ancient Egyptians were well aware of the powerful electric shocks delivered by the electric eel, even incorporating them into their religious rituals. Later, systematic experimentation upon electricity was conducted upon amber's ability to hold static electricity to the development of the Leyden Jar and Benjamin Franklin's famed kite experiments to harness the electricity of lightning. By the same token, scholars extending to the days of Aristotle contemplated the mysterious nature of magnetic lodestones. The ancient Chinese harnessed the lodestone to develop the magnetic compass. The earliest lenses and hence the first scientific investigation of light can be traced back to as far back as the ancient Assyrians of 600BC.

Throughout history, scientific connections were periodically drawn amongst the relationship between electricity, magnetism, and light. Nevertheless, those three fields were widely considered to be three entirely separate physical phenomenon studied.

However, during a period of only a few decades of the 1800's, all three phenomena were demonstrated to be entirely equivalent to each other. In a span of 12 years, the seminal publications of Hans-Christen Orsted, Andre-Marie Ampere, Joseph Henry, and Michael Faraday demonstrated that not only could electricity induce magnetism and vice versa, but that they were actually the same force, stemming from the same source of charged particles. No longer did it make sense to investigate electricity's "effect" upon magnetism or vice versa when that relationship was now understood to be tautological. A few short decades after the unification of electricity and magnetism, James Clerk Maxwell proved that light was simply a special case of electromagnetism and the waves it generates, light consisting of electromagnetic waves of a constrained range of frequencies (that being known as the 'visible spectrum'). Again, no longer would it make sense for scientists to continue to investigate correlations between electricity or magnetism and light after they are all understood to be the same force.

Electromagnetism is but one of the bevy of scientific phenomena that were once widely considered to consist of entirely separate if related forces but were later shown to be unified and therefore tautological. Scientific progress hinges upon discovering that certain topics of empirical investigation are actually tautological in nature and therefore no longer worthy of empirical study. Einstein famously unified the concepts of energy and mass through his celebrated equation E=mc-squared: the basis of nuclear energy. Einstein through his theory of special relativity also unified the concepts of space and time by demonstrating that they both occupy the same 4-dimensional continuum known as space-time. Publishing a new scientific paper that demonstrates that space and time are correlated would be no more interesting that declaring that 1+1=2.

Management academia would likewise progress by determining that certain empirical relationships are actually measuring the same concept and are therefore tautological in nature. As Thomas Powell disdainfully stated: “[Management] research is burdened with an immense body of obfuscatory grammar – [such as] ‘intangible-invisible assets,’ ‘causally-ambiguous causes” ”[75] One might reasonably infer that perhaps some of those myriad terms encumbering the management literature are actually measuring the same concept which renders any empirical analysis of their relationship superfluous. One struggles to think of precisely what the difference is between ‘organizational creativity’ and ‘organizational innovation’ or ‘social capital’ and ‘network capital’. What precisely is the difference between ‘classical’ organizational institutional theory and ‘new’ organizational institutional theory? Future management research would ideally collapse at least some of those concepts into a unified idea. Indeed, today’s management research should be collapsing popular management concepts of the past by demonstrating that they are mere components of the same overarching entity, in the same manner as the formerly discrete research communities of electricity, magnetism, and light all had to concede that they were investigating the same topic.

Yet has that ever happened in management? Of the 150 paper comprising my dataset, not a single time did the authors later concede that the findings within are tautological.

[75] Powell 2002.

I therefore propose the Tautology Challenge: **Can anybody name a single published A-level management paper where the authors themselves agree that the conceptual relationships within were later demonstrated to be tautological/unified in light of new evidence, hence rendering the paper to be superfluous and hence ought to be ejected?**

To be sure, there should be no shame in making such a concession. Earlier electricity researchers did not realize that their field was tautologically connected to magnetism and vice versa. But their concession meant that their standalone discoveries became tautological ejecta – that no such standalone theory of electricity, magnetism, or light exists - with such ejecta indicating scientific progress. Management academia would likewise progress by generating tautological ejecta where researchers concede that their former discoveries are later found to be unified and hence tautologically equal to other management discoveries. Has there been even one such example in management?

6.2 Are Management Theories Ever Later For Being Rendered Trivial?

Let's recall that a key value-add of any scientific theory is that it provides non-trivial propositions. The standard discourse of scientific progress is that a bevy of seemingly non-trivial findings are published at a particular point in time. Additional vigorous testing will reveal that some of those seemingly non-trivial findings are found to actually be trivial after all, often times through error of the data analysis or other such new information that renders the finding mundane. For example, the supposed recent discovery of non-trivial faster-then-light-speed motion was later found to be merely a trivial issue of improper clock synchronization. The supposed recent discovery of arsenic-based DNA – which would represent an entirely novel and non-trivial type of genetic coding – was likely nothing more than a trivial outcome stemming from laboratory contamination. By a similar logic, a newly proposed (yet untested) scientific theory that provides a mix of both non-trivial and trivial propositions, might carry value as long as some of the non-trivial propositions do survive later empirical testing. However, if the only propositions of that theory that do survive are the trivial ones, then that theory does not provide much intellectual value and should be ejected.

Yet as demonstrated in past chapters, at least within the A-level journals, management researchers never seem to concede that their findings are trivial despite the fact that they evidently are, as judged by a panel of untrained participants. That might not be problematic if those researchers later conceded the triviality of their findings. But to my knowledge, that apparently never seems to happen either. Of the 150 papers comprising my empirical dataset, not a single time did the authors later concede that their findings are trivial.

I therefore propose the Triviality Challenge**: Has any management researcher ever conceded that their published A-level paper is trivial in light of new evidence, hence rendering it ejected?**

6.3 Are Management Theories Ever Ejected Empirically Invalidity?

The Titius-Bode Law of astronomy was a once-formerly held hypothesis that predicted the locations of objects of the solar system. It was first formulated using the known positions of the six known planets at the time- Mercury, Venus, Earth, Mars, Jupiter, and Saturn – as well as their contemporaneously known moons. The Titius-Bode Law gained great credence when it successfully predicted the hitherto undiscovered planet of Uranus in precisely the location that the Law said that Uranus would appear, and gained even more credence with the discovery of the planet Ceres (whose status was later downgraded to an asteroid). However, the discoveries of Neptune and Pluto failed to conform to the Law and the further discoveries of the Kuiper Belt and the dwarf-planet Eris entirely discredited the Law. Modern astronomy now treats the Titius-Bode Law as a mere historical footnote.

The moral of the story is that initial publication of any theory should be viewed as merely tentative in nature, forever subject to revision and ejection in light of new data. A theory that fails to explain newly collected empirical data is not an empirically valid theory and should be ejected as such.

That implies that management academia should be ejecting theories constantly in light of the recent veritable explosion of available data and computational power. Much management research published as recently as even ten years ago were based upon relatively sparse datasets using primitive analytical tools. Yet today, not only have many years' worth of data been collected since that time, but much of it exists in readily accessible formats, available for easy computational consumption. As Gerry Davis quipped: "Undoubtedly, our current undergraduates will soon be downloading data to their cell phones and running sophisticated fixed-effects regressions on every company listed on the New York Stock Exchange in the 20th century".[76] Furthermore, formerly closed nations such as China and Russia who had heavily censored their data have now opened their economies and provided far greater transparency into their managerial practices. A candy-store of unparalleled data awaits the eager management researcher of today.

One would therefore think that management researchers would be constantly revalidating past publications. Even a simple tactic such as retesting theories upon every years' worth of new data – possibly even appended to the extant dataset – would suffice. Perhaps more fruitfully, researchers could test their theories upon entirely new industries or new countries as they become available. 20 years ago, the entire Internet/E-commerce industry didn't exist at all, whereas now it spearheads the cutting edge of technology innovation. One might think to retest, say, theories of innovation of a generation ago to examine whether they remain valid upon this entirely new industry. One would then report those theories that no longer hold predictive power as being empirically ejected.

[76] Davis 2010.

But that evidently never seems to happen. Of the 150 empirical papers in my dataset, not a single time did the authors ever later report that their data failed to replicate upon new data, whether that replication was conducted by themselves or others. Indeed, in the vast majority of cases nobody seemingly ever bothered to replicate the original study at all (or if it was, that fact was never reported).

Hence my Validation Challenge**: Name me a single A-level management paper where the original authors conceded that their work failed subsequent replication attempts and therefore ought to be ejected.**

6.4 Are Management Theories Ever Ejected for Non-Causality?

As mentioned before, the 'Causality Revolution' took hold within management academia only within the last 10-15 years. Prior to that time, entire bookshelves full of management articles and books were written that demonstrated little if any supporting evidence of causality. Indeed, many of the most influential management theories in history were published during the late 1960's to early 1980's[77] with little if any supporting causal evidence. Many such works consisted of little more than correlations at best, with the actual causal directions amongst variables left unstated and obscure. But perhaps more importantly, even in the midst of the Causal Revolution, empirical methodologies have been progressing rapidly. Researchers have been invoking ever-more-exotic causal tools - split-sample instrumental variables, dynamic panel data

[77] Davis 2010.

models, doubly robust estimators just to name a few – that I've now personally heard on multiple occasions how people lament that papers published even a few years ago could never be published today.   One only need compare a randomly selected journal issue of, say, Strategic Management Journal published five years ago to one published today and marvel at the vastly increased rigorousness of causality required.

But yet again, what apparently never happens is researchers revisiting their past published work that revisited either by themselves or other researchers using the most topical causal empirical methodologies of the day, and declaring that the causality of their prior published result is no longer supported.

As a specific example, recall that the proper usage of an instrumental variable requires that there be no alternative unblocked pathway from the instrument to the outcome variable. A decade ago, few papers invoking instrumental variables actually checked for potential unblocked pathways, perhaps because the methods for doing so were largely unknown outside of the statistics community. Nowadays, such techniques are far more commonplace.   Furthermore, management academia presumably continues to gather new knowledge about, such that subsequent research might identify a hitherto unknown unblocked pathway.   Yet despite this bevy of new information, apparently no researcher ever returns to their original paper and declares their prior instrument to be suspect.
I therefore issue my Causality Challenge: **Name me a A-level management paper whose authors later conceded that subsequent research has undermined the causality of their paper such that it should be ejected.**

6.5 Conclusion

A properly functioning academic discipline would be ejecting candidate theories like clockwork. Initial publication would be treated as little more than a basic (and flawed) screen; candidate theories who survived that screen would then be subjected to a merciless battery of tests to ensure that they remain robust. Most candidate theories would be expected to fail these tests.

However, apparently that's not what happens in management academia. Firstly, initial publication in management academia is apparently the entire ball game: a single A-level publication alone being sufficient to land a tenure-track job at a well-regarded business school, and a string of A's is generally sufficient for tenure. Whether any of those publications actually survives the battery of ejection tests is irrelevant. Those A-level publications garnered the academic job offer or promotion, and for most subjects in question, that's what ultimately matters.

But more importantly, management academia lacks a culture of replication. Indeed, many management journals evidently outright discourage replication.[78] What should be a relentless barrage of retesting is nothing more than a mere whimper. Once somebody publishes a particular managerial finding, nobody is ever really allowed to challenge that

[78] Hambrick 2007.

finding ever again – and certainly the original authors of that finding never concede that that finding was successfully challenged.

Nevertheless, one might argue that the above problems speak to the structural careerist issues of management academia. As long as the incentive system fails to reward replication and ejection, little of either will be produced. Yet obviously not all management theories will be ejected. Perhaps certain researchers have indeed uncovered a management theory that has indeed proved to be robust, having survived the battery of falsification tests and demonstrated that the theory in question remains non-tautological, non-trivial, empirically valid, and causal. Better yet, could a dedicated team of researchers who eschew the standard careerist pressures generate and replicate a robust theory by themselves? That is, might a certain subset of management theories be progressing, even if the rest of the field is not? We now turn to this question in the following chapter.

7 Have Any Management Theories Become Paradigms?

7.1 Might Individual Theories Be Progressing Even if the Field is Not?

Let's harken back to the evidence used to demonstrate my claim that the management field (consisting of both gurus and academia) is unscientific because it violates basic empirical methodological rules regarding proper validation. A savvy reader might argue: hasn't my empirical methodology itself violated that very same rule? That is, by demonstrating that management gurus such as Peters & Waterman sampled on the dependent variable to bolster their hypothesis, couldn't one argue that I too sampled on the dependent variable by selectively choosing the particular example of Peters & Waterman to bolster my hypothesis? Similarly, couldn't my examination of Collins' Good to Great, the Resource Based View, or the various other examples that I utilize likewise be viewed as other instances of sampling on the dependent variable by only choosing those examples that bolster my hypothesis? Hence, couldn't one then argue that my claim that management is unscientific is itself unscientific because it itself is not properly validated because I specifically chose only examples that bolster my hypothesis – the very same crime committed by the management field?

The crux of the counterargument hinges upon whether you believe that the appropriate level of analysis consists of the entire management field as a whole or rather of individual theories that comprise the management field. If you subscribe to the former viewpoint, then it doesn't matter whether the statements I've made regarding the field are based upon selective data or not. After all, the world has only one lone field of management. I am making no inferences towards other fields of management existing in parallel universes. Therefore as long as the statements I am making about the lone extant field have the evidence to support them, it doesn't matter how that evidence was found. As an analogy, if my car has a blown gasket, it doesn't matter how I discovered that my gasket is blown because I am not making inferences about other cars. What matters is that a central component of my car is broken which renders that particular car inoperable.

However, perhaps you believe that the appropriate level of analysis should consist of the individual theory – and not just any individual theory, but certain 'preferred' theories. That is, even if the management field as a whole may be unscientific, perhaps certain individual theories developed by individual scholars are indeed progressing scientifically. Then one might argue that it matters not whether the field as a whole – or even most individual theories that comprise the field – are not progressing scientifically, as long as your favored theory by your favored scholars are doing so. Might individual theorists be able to quarantine themselves from the maladies that afflict the field as a whole?
Were that to be so! Unfortunately, I'm afraid that it's not that simple. While I have no doubt that certain individual management scholars are indeed attempting to make true scientific progress, unfortunately, that progress is thwarted by the activities of other researchers who are not making scientific progress and by the management field that permits such thwarting. As ASQ Editor Gerry Davis acidly observed: "Given the diverse predictions of the many paradigms in [management] theory, it is almost always

possible to find a theoretical rationale for a result." [79] Whatever scientific predictions are being produced to answer a particular research question are obscured by any conflicting unscientific unreliable predictions regarding the same research question – and those conflicting predictions almost always exist as Davis argues.

Let's consider what scientific progress is, which stems from the true societal purpose of science. Scientific progress occurs not simply when individual researchers acquire increased knowledge but when <u>society</u> is provided with increased knowledge. Indeed, a baseline norm of the academic community is that to receive scholarly credit for a research finding, it must be disseminated through public conferences and publications rather than through coded documents and secret correspondence, with the underlying purpose being that society is provided with the finding. [80] If I discover the Higgs Boson but hide that discovery away and never inform society about it, I would say that scientific progress has not actually occurred. As the philosophical saying goes, if a tree falls in the forest and no-one is around to hear it, does it make a sound?

On the other hand, let's say that I do rigorously and scientifically discover the Higgs Boson and also do publicly communicate that finding to society. However if other researchers who are not being scientific nevertheless claim that they too have discovered the Higgs Boson – and their findings stand in direct opposition to my findings with no consensus method to adjudicate veracity - whatever scientific progress has been made

[79] Davis 2010

[80] Patents likewise are granted only after the inventor fully discloses the workings of his invention, where the contents of the patent are available to anybody for inspection. Society is essentially trading knowledge about an invention in return for a temporary monopoly on the profits of that invention.

has been attenuated. To be sure, it is better for society to have a mixture of correct and incorrect Higgs Boson results rather than having only incorrect Higgs Boson results. Nevertheless, in this case, the research community as a whole has not provided society with more knowledge regarding where the Higgs Boson is because society does not know which researchers to believe. If a clutch of trees simultaneously falls in the woods, can you elucidate the sound of any individual tree?

Hence, whether we like it or not, the efforts of individual management scholars in achieving scientific progress are mediated by the social norms of the management field as a whole. Scientific progress occurs not only simply through the act of individual scholars discovering and disseminating reliable results to society, but also through the complementary act of unreliable results not being disseminated to society. The most beautiful guitar solo in the world is ruined if other musicians are simultaneously playing a cacophonous dirge. Unfortunately, management academia is riven with theoretical disputes with no means to resolve them and therefore no agreed-upon method with which to remove obsolete theories. As Gerry Davis remarked: "although dozens of studies have purported to provide critical tests to adjudicate between theories, the contests always seem to end as a draw—a Stanley Cup playoff that never ends."[81]

Indeed, one shining example of a theoretical dispute that should have concluded long ago but apparently has devolved into the interminable Davis-ian endless Stanley Cup playoff is the well-worn conflict between Resource Dependence Theory (RDT – not to be confused with RBV) and Transaction Cost Economics (TCE). Both theories were

[81] Davis 2006.

initially subjected to highly scientific inquiry in terms of their empirical falsifiability and their validity. Each theory also provided clearly non-obvious predictions, at least relative to each other - a crucial point that I shall explore further. Yet somewhere along the lines, each of the theories took a non-scientific turn: the playoff series that should have ended years ago nevertheless plays on, with no winner in sight for the foreseeable future. In this next section, I demonstrate how each theory climbed the initial steps of the Christensen-Sundahl pyramid, but how each ultimately failed to climb the final steps to generate an ejecta. I dub this case study: A Tale of Two Theories:

7.2 A Case Study of Conflict: A Tale of Two Theories

It was the best of theories, it was the worst of theories, it was the wisest of theories, it was the most foolish of theories, it was the epoch of belief, it was the epoch of incredulity. Such has been the rhetoric held forth by both the supporters and detractors of Resource Dependence Theory (RDT) and Transaction Cost Economics (TCE): two schools of thought that not only have each become foundational pillars of the management theoretical landscape with citations easily numbering in the 5-figures, but whose epic rivalry has raged ever since their very inception. Oliver Williamson, winner of the Nobel Prize for developing TCE, dismissively characterized RDT as “a pied piper whose enticements are better resisted in favor of more mundane efficiency concerns”[82] and that the value of RDT is incremental at best - that it will merely “sometimes add detail”[83]. Jeffrey Pfeffer, the father of RDT, retorted that TCE presents a “very benign view of

[82] Williamson 1981
[83] Ibid.

social organization"[84] that – to the extent of holding any truth value whatsoever – does so only because TCE has become a self-fulfilling prophecy that shapes managerial behavior according to the very tenets it espouses[85]. Indeed, Pfeffer further characterizes TCE is a prime exemplar of the primacy of faddishness over truth within management academia, describing TCE as an archetypal illustration of "Murray Davis's argument that great theories in social science attain their status <u>not because they are true but because they are interesting</u> and engage the attention of their audience of experts and practitioners."[86]

I invoke the seemingly interminable conflict between RDT and TCE as a case study that examines how the management field as a whole <u>should</u> have developed compared to how it actually developed. I describe how both RDT and TCE ascend the initial rungs of the Christensen-Sundahl pyramid as each theory provides a litany of falsifiable, non-obvious predictions that have been subjected to empirical validation. I then describe how the process of continual revalidation and retesting of each theory – particularly in light of the existence of the rival theory – presents a golden opportunity for the generation of theoretical ejecta, which I had previously denoted as a crucial indicator of a healthy academic discipline. But that opportunity has been squandered. Despite the conflicting assumptions and empirical predictions of each theory, neither theoretical camp has conceded even an inch. Each theory continues to attract supporters who then denigrate the importance of the rival theory. The inability to adjudicate between rival theories and eject one in favor of the other characterizes the very heart of the dilemma of the management field.

---

[84] Pfeffer 1996.
[85] Ferraro Pfeffer Sutton 2005.
[86] Ibid.

7.3 TCE and RDT: A Quick Overview

The complete tenets of both TCE and RDT lie far beyond the scope of this work: indeed, each theory has spawned hundreds of shelves full of books and journal articles. Interested readers of TCE are directed to Oliver Williamson's seminal 1975 book 'Markets and Hierarchies' and his updated 1996 work 'The Mechanisms of Governance', while those interested in RDT are directed to Jeffrey Pfeffer's and Gerald Salancik's original 1978 edition of 'The External Control of Organizations' and its 2003 updated edition. Those books, along with the works that cite them, provide a complete overview of the traits and nuances of each theory.

However, in a nutshell, TCE purports to explain the age-old question of why organizations exist in the first place. Why does the world have organizations with employees that require oversight by managerial hierarchies; why not instead simply contract for everything they require through the open market? For example, instead of an organization employing secretaries directly upon its payroll, why not simply outsource all secretarial tasks, perhaps as temp work, whenever necessary? On the flip side, if the direct hiring of employees is always preferable to outsourcing, then why is there not a single organization in which everybody is an employee?

Transaction Cost Economics (TCE), fittingly as per its name, explains such behavior as a manner by which organizations <u>minimize transaction costs</u>. Open-market transactions may be expensive to complete: not only necessitating search and information costs in simply finding what you require (i.e. finding qualified and available secretaries whenever secretarial work must be completed), but also bargaining/negotiating costs in procuring desired products at an acceptable price and policing/enforcement costs in ensuring that work requirements are fulfilled.   However, open market transactions provide cost advantages of their own, as the market forces of competition and innovation may cause certain outsourced transactions to be cheaper than in-house projects. While most organizations directly hire employees to perform secretarial work, few organizations would directly hire their own carpenters and welders to construct their own office furniture rather than simply outsourcing that task by buying furniture through the market. TCE predicts that insofar as the direct hiring of employees allows an organization to minimize costs, then internal hiring will be more prevalent than market transactions and vice versa. TCE therefore generates predictions about organizational boundaries regarding make/buy decisions: what organizations choose to make in-house versus what they choose to buy through the market.   Such decisions are reflected not only in the internal operations of the organization, but also in the products/services they provide. Apple, for example, outsources the manufacturing of the IPhone while retaining the design work of the IPhone in-house – a prediction that TCE would readily support assuming that the outsourcing of manufacturing is less costly for Apple than is maintaining in-house IPhone manufacturing capabilities, with the reverse being true regarding IPhone design work. This is an assumption that I shall later revisit.

Resource Dependence Theory (RDT), unlike TCE, doesn't attempt to explain why organizations exist, but, like TCE, does attempt to explain the behavior of those organizations. RDT posits that the behavior of organizations, rather than merely serving to maximize profit or minimize cost, is driven by social context and power. Indeed, Pfeffer states that organizations routinely engage in behavior "regardless of considerations of profit or efficiency".[87] Rather, organizations behave in a manner that maximizes their relative power within their social context. To the extent that other parties possess necessary resources – which could be capital goods, labor, investment capital or anything else which the organization depends– the organization will attempt to gain control over parties that provide those resources. For example, if a restaurant requires a particular rare vegetable that is only available at only one farm in the world to produce its signature sandwich, the restaurant will attempt to gain control over that farm that provides that vegetable, perhaps by merging with the farm, or enacting a strategic alliance with it. The restaurant may also attempt to weaken the power that the farm holds over it by diversifying its sandwich offerings by marketing a different sandwich that does not require that vegetable. Until and unless the restaurant is able to do so, then the restaurant is <u>dependent</u> upon the <u>resources</u> provided by that farm. That restaurant may in turn possess resources upon which other organizations depend. For example, if the restaurant happens to provide a convenient location and ambience for local business deals to take place, then the restaurant will hold power over local businesses who want to close deals. Those local businesses may in turn attempt to weaken the power that the restaurant holds over it. The socially-embedded power derived from resources that

[87] Pfeffer 1987.

organizations depend upon is therefore the chief driver of organizational behavior according to RDT.

To be sure, some of the predictions of TCE and RDT coincide. However, many of them – in particular those stemming from the baseline assumption over whether the key driving force of organizational behavior is either efficiency-driven cost minimization or socially-driven power maximization – directly conflict with each other. That represents a golden opportunity for the generation of a theoretical ejecta as both theories ascend the Christensen-Sundahl pyramid: one theory should have been ejected in favor of the other. Or so it would seem. I now walk the reader through how each theory successfully ascended the initial steps of the pyramid on their quest for paradigmatic status such that one theory <u>should have ejected</u> the other, and yet how that last step never seemed to happen.

7.3.1 Step 1: Falsifiability of TCE/RDT

Recall that the first step of the Christensen-Sundahl pyramid is regarding whether a prospective theory generates empirically falsifiable predictions – predictions that are not true strictly by definition. Both TCE and RDT clearly do so. TCE's independent driving variable is 'transaction cost', whereas the dependent variable is the resulting action of the organization: whether the organization chooses to merge/acquire another organization or not, to engage in a strategic partnership/alliance, to spin-off a division, to outsource a particular task, etc. TCE predicts that those organizational actions associated with the lowest transaction costs will be preferred. What makes TCE falsifiable is that transaction costs are not defined to be organizational actions. Organizations are not required to act in a manner that minimizes transaction costs purely by definition. For example, organizations could conceivably consistently engage in mergers despite those mergers not actually minimizing transaction costs, which would represent a direct threat to the validity of TCE.

Similarly, the independent driving variable of RDT is the power derived from possessing/lacking key organizational resources. That independent variable predicts the outcome of the dependent variable representing resultant organizational action. RDT predicts that organizations will engage in actions that maximize their relative power. Again, RDT is falsifiable because power/resources are not defined to be organizational action. For example, if organizations consistently chose to engage in spin-offs in a manner that actually weaken their power, that would be a direct threat to the validity of RDT.

7.3.2 Step 2: (Non)Obviousness of TCE/RDT

To traverse the second step of the Christensen-Sundahl pyramid, a prospective theory must generate non-obvious predictions. That raises the question – by whom should obviousness be judged? In an earlier chapter, I had proposed that obviousness of a set of randomly selected management predictions be judged by baseline set consisting of people who have no professional or academic management background. While I could do the same regarding the obviousness of the predictions of TCE/RDT, I believe that a simpler method avails itself. Given the directly conflicting predictions of TCE/RDT regarding the driving force of organizational actions (cost-based efficiency vs. resource-driven power), the predictions of TCE will not be obvious to the proponents of RDT and vice versa. The prediction that organizations will consistently behave in a manner that reduces their transaction costs even at the expense of power is non-obvious to the proponents of RDT, for to repeat Pfeffer's quote, organizations generally behave in a manner "regardless of considerations of profit or efficiency."[88] Likewise, the prediction that organizations behave in a manner that consistently increases their power even at the expense of incurring higher transaction costs is not obvious to the proponents of TCE. RDT/TCE also can provide non-obvious predictions when the other theory is inconclusive. For example, if an organization is faced with a set of choices that all have the same impact upon transaction costs but differing impacts upon power, TCE is unable to predict the choice the organization will make, but RDT can. Hence, the prediction of RDT in this case will not be obvious to the proponents of TCE.

Therefore, in contrast to my previous methodology that relied upon volunteers with no management background to judge obviousness, in this case, obviousness is being judged by a group of management scholars – the proponents of either TCE or RDT – who have extensive backgrounds in the literature. Put more starkly, the predictions of RDT are apparently not obvious to a Nobel laureate such as Oliver Williamson in light of his dismissive comments about RDT, and the predictions of TCE are likewise not obvious to a giant of the management field such as Jeffrey Pfeffer. I would therefore argue that

[88] Pfeffer 1987.

both RDT and TCE clearly surmount the requirement for obviousness. Any theory that provides a set of predictions that directly contradict the predictions of a highly learned group of scholars is clearly an important contribution – assuming that those predictions are empirically validated, to which we now turn.

7.3.3 Steps 3&4: (Initial) Validation and Causal Analysis of TCE/RDT

Both TCE and RDT have indeed been subject to a legion of empirical causal validation. Enumerating the now thousands of empirical studies that have tested various aspects of TCE/RDT are clearly beyond the subject of this text: interested readers are directed to the summaries presented in Pfeffer & Salancik (2003), David & Han (2004), Scott & Davis 2007 and the citations that they present. Suffice it to say that both theories have been successfully subjected to extensive (initial) validation. Hence, TCE and RDT can rightfully be promoted to candidate status.

7.3.4 Step 5: Retesting and Revalidation: Falling off the Pyramid

Recall that <u>initial</u> validation alone is insufficient for a theory to attain paradigmatic status. The theory must be subjected to a cycle of extensive retesting and revalidation, incorporating the latest theoretical and methodological advances to ensure that the predictions of the theory continue to hold. Those theories that survive initial validation but fail to survive the cycle of continuous revalidating/retesting should be discarded as ejecta.

Such revalidation and resultant ejecta should have happened in the case of the conflict between TCE and RDT. The proponents of each theory clearly know about the existence of the other theory - so much so that they aren't shy about tossing derogatory comments at each other. The proponents of TCE should therefore have retested their predictions using RDT as a potential counter-explanation, and vice versa regarding the proponents of RDT. For example, the proponents of TCE could have re-run their regression models but using resource-driven power as an additional explanatory variable to see if that new variable provides more predictive power than the transaction-cost variable. They could have conducted surveys and interviews to ascertain whether managers are more apt to respond to considerations of costs or to power. They could have conducted a myriad of other potential studies to retest the notion that transaction costs rather than power are indeed what drives organizational behavior, and should concede defeat if it does not. But apparently, this never happened. Or at least, nobody conceded defeat and relegated their own theory to the status of ejecta. Each theory's proponents continue to hold forth that their theory is correct while the other is wrong.

If anything, much of the empirical evidence regarding mergers and acquisitions seems to consistently disfavor TCE, as firms consistently engage in merger/acquisition activity that fails to improve cost structure, but are perhaps more readily explained by considerations of power. As stated forcefully by Scott & Davis (2007):

> "On the face of it, the weight of evidence would seem to favor resource dependence, as most acquisitions either do not increase organizational performance or actually decrease it – share prices of acquiring firms frequently decline upon the announcements of acquisitions (Morck, Schleifer and Vishny 1990) suggesting that the stock market generally views them as a bad idea. The verdict on diversifying mergers is especially negative: 'The evidence that corporate diversification reduces company value is consistent and collectively damning' (Black 1992:903) and Porter (1987) finds that firms that diversified ended up disposing of three-quarters of their acquisitions…" [89]

Yet TCE not only lives on, but thrives despite not only the lack of retesting that incorporates RDT as a counterargument, but also the consistently negative empirical evidence provided by corporate merger/acquisition behavior. Now, granted, one might argue that a few minor phenomena that TCE cannot explain should not cause the entire theory to be rejected. But merger/acquisition behavior can hardly be dismissed as a 'minor phenomena', but rather one of the central activities that any organization will conduct. Nevertheless, only a few short years after Scott & Davis published the above quote regarding the consistent inability of TCE to explain merger/acquisition activity, Oliver Williamson was awarded the Nobel Prize in Economics for his work in TCE.

[89] Scott & Davis 2007.

That is not to say that RDT hasn't had empirical problems of its own. Indeed, RDT has had extensive problems with explaining the cycles of mergers and divestments that have occurred over the last century. A central prediction of RDT is that corporations should seek to become conglomerates to maximize their power vis-à-vis suppliers and customers, yet the large conglomerates of the 1970's had almost completely disappeared during the subsequent decades' wave of divestment and leveraged buyouts. During the 1990's, organizations pursued outsourcing with gusto, readily handing away manufacturing and production capabilities to third-party contractors while retaining the (so-called 'core competency') thin layer of design, marketing and distribution. RDT can explain this divestment trend only by a corresponding trend of power/resource shifts corresponding to those divested assets, yet such a shift has never been clearly established. Why Apple now outsources almost all of its manufacturing today when it used to keep manufacturing in-house only a few decades ago is not something that can be clearly explained by RDT but might well be better explained by TCE.

But the upshot is that, given the directly conflicting predictions provided by TCE and RDT, they should have both been subjected to continual retesting and revalidation until one of them was ejected. Either that, or at least they should be currently subject to a program of testing that will eventually eject one theory in favor of the other, such that the proponents of one theory will concede that their theory is untenable. While nobody can predict the future with certainty, the chances of that happening seem rather bleak. Management theories never seem to die; we instead seem to be stuck with the "Stanley

Cup playoff that never ends" that was alluded to earlier. As stated trenchantly by Gerry Davis and Chris Marquis:

> "One sign of progress is that weak theories are selected out, as researchers favor progressive theories capable of accounting for observed regularities while making novel predictions . We are willing to assert that organization theory as a discipline has no history of such selection and little prospect for it in the future."[90]

The conflict between TCE and RDT presented a golden opportunity for management academia to generate an ejecta – an opportunity that heretofore has been squandered. Each theory produces sharp predictions that directly contradict each other. Therefore both theories cannot be true. At best, the theories should be integrated together contingently such that TCE is applicable in certain contingent conditions, whereas RDT is applicable in different contingently conditions. Each theory could then be said to be 'contingently ejected'. But as far as I know, that hasn't happened either. The proponents of each theory maintain that their theory is dispositive while the other theory is incorrect. The management field cannot progress as a whole while these warring camps perpetually refuse to surrender any ground to each other.

I present this step underlined in the following diagram:

[90] Davis & Marquis (2005).

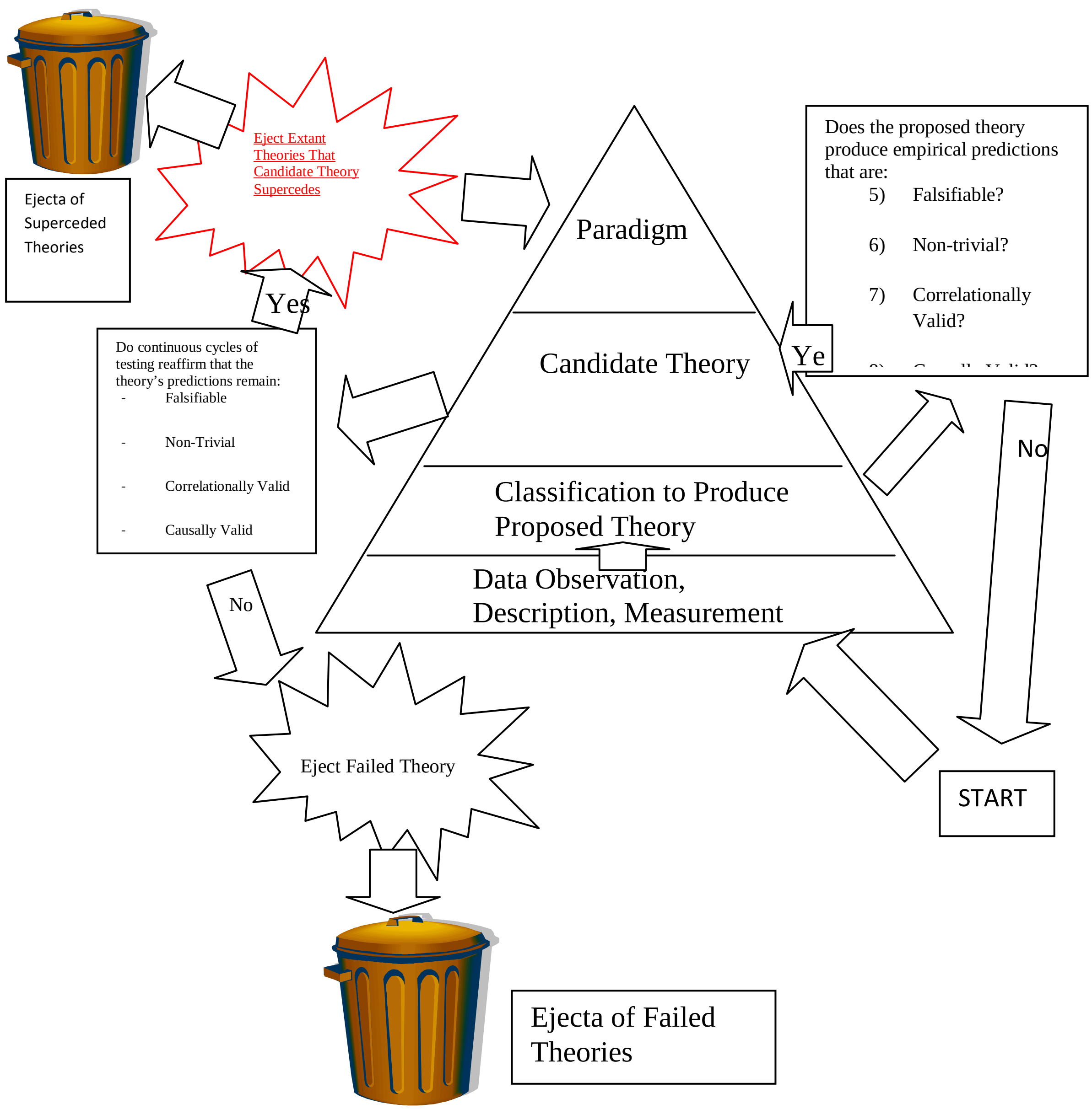

Ejecta of
Superceded
Theories
Eject Extant
Theories That
Candidate Theory
Supercedes
Yes
Paradigm
Candidate Theory
Classification to Produce
Proposed Theory
Data Observation,
Description, Measurement
Does the proposed theory
produce empirical predictions
that are:
5) Falsifiable?
6) Non-trivial?
7) Correlationally
Valid?
Ye
No
START
Do continuous cycles of
testing reaffirm that the
theory's predictions remain:
- Falsifiable
- Non-Trivial
- Correlationally Valid
- Causally Valid
No
Eject Failed Theory
Ejecta of Failed
Theories

Figure 45

7.3.5 Summary and Recommendations Regarding Theoretical Dissensus

Harkening back to the argument that I had data-dredged the case studies of the previous chapters to demonstrate my argument that the management field is unscientific, one could similarly argue that I data-dredged the TCE/RDT conflict to demonstrate my argument that the management field is unscientific. Yet the conflict between the TCE-RDT dyad seems to pervade the management field as a whole. Indeed, such a theoretical dissensus has been remarked upon periodically by numerous scholars within the management field.[91]

For example, a central prediction of the influential population ecology theory is that firms are afflicted with 'organizational inertia' and are generally unable to successfully change their existing organizational form (e.g. from a top-down command and control type of organization to one that relies upon organic teamwork and employee consensus), and those that try generally increase their odds of bankruptcy. Overall changes to the organizational forms of a particular domain are effected only by deaths to existing organizations that used the organizational structures, which are replaced by newly founded organizations that utilize the new organizational structure. However, another highly cited theory - structural contingency theory - predicts precisely the opposite: that not only can organizations indeed successfully change, but indeed should do so to custom-fit their organizational form in accordance with changes to the environment. For example, if customer tastes shift from initially wanting only black Model-T Fords to now wanting different car designs in different colors, then Ford could readily shift from manufacturing only one car design in one color to offering multiple designs and multiple colors despite the costly change to Ford's organizational design.

Similarly, proponents of New Institutional Theory (NIT) predict that organizations tend to adopt the same general organizational features as other organizations within the same industry by a process that NIT calls 'institutional isomorphism'[92]. As Gerry Davis put it:

[91] Also see McKinley Mone 1998.
[92] Powell DiMaggio (1983) colorfully depicts this as the 'iron cage' of institutional isomorphism.

"[NIT's] imagery of relentless pressures for conformity is appealing in a world in which every strip mall in every town is populated with the same retailers, restaurants, and latte parlors."[93] However, one of the central prescriptions of the eminent strategy scholar Michael Porter is that organizations survive by differentiating themselves from competitors in the same industry.[94] In particular, Porter proffers that such differentiation tends to happen by two key mechanisms: either low-cost leadership or through product innovation and (perceived) quality. Porter therefore maintains that organizations should attain a state of <u>bipolarity</u> rather than the <u>unipolarity</u> predicted by proponents of NIT, – for example, the restaurant industry should bifurcate into becoming either McDonald's or Ruth's Chris Steak House.[95] Still other strategy scholars predict that a spectrum of equilibrium points on the cost-quality spectrum are feasible and that industries should attain a state of multipolarity [96] (e.g. Outback Steakhouse might occupy a rung in the middle of McDonald's or Ruth's Chris) – an even farther cry from the unipolar, isomorphic world predicted by proponents of NIT.

Such disagreements seem to be not the exception but rather the rule. As stated by Gerry Davis: "Given the diverse predictions of the many paradigms…it is almost always possible to find a theoretical rationale for a result – either before or after the results are known."[97] Jeffrey Pfeffer noted that "the domain of organization theory is coming to resemble more of a weed patch than a well-tended garden. Theories . . . proliferate along with measures, terms, concepts, and research paradigms. It is often difficult to discern in

[93] Davis 2010.
[94] Porter 1980.
[95] Ibid
[96] Miller 1992.
[97] Davis 2010.

what direction knowledge of organizations is progressing."[98] He later noted that "there is virtually no disagreement about the <u>fact</u> of more…[theoretical] proliferation"[99] Indeed, that dissensus that pervades management has been rife that Pfeffer proposed that that very dissensus itself become a topic of research to find out why the field of management cannot generate the greater theoretical agreement enjoyed by even the social sciences, let alone the natural sciences with their strong paradigmatic consensus regarding their core theories (biology having evolutionary theory and genetics, chemistry having molecular bonding theory, physics having quantum physics and relativity, etc.) [100] To date, as far as I know, nobody has taken up Pfeffer's offer to conduct such a study.

Hence, given the wide diversity of theory and predictions that populate the management landscape, any theory that successfully traverses the pyramid to its apex can do so only through the production of ejecta of prior theories that predict opposite results. For TCE to summit the pyramid and become a true paradigm, then RDT must be wrong and should be ejected, and vice versa. Likewise, for NIT to summit the pyramid to become a true paradigm, then the theories of Porter and the other strategy scholars that predicted a bipolar/multipolar world must be wrong and therefore should be ejected, and vice versa. Yet, as evidenced by my challenge questions, such ejection never seems to happen; nobody can think of a clear example of an ejection. To requote Gerry Davis and Chris Marquis: "To our knowledge, no organizational theory has ever been [definitively] "rejected'".[101]

---

[98] Pfeffer 1993.
[99] Pfeffer 1997.
[100] Pfeffer 1993.
[101] Davis & Marquis 2005.

That lack of ability to eject even a single theory lies at the heart of the broken machinery that is the management field. Nobel laureate George Stigler once noted that: "There is no obvious method by which a [field] can wholly rid itself of once popular theories, logical error aside (and even this may not be a true exception)."[102] Pfeffer likewise noted that:

> "**There is little apparent agreement about how to resolve the controversies among competing [theories]-not only disagreement about which one is correct or useful, but disagreement about how to even go about figuring this out. Because of these fundamental disagreements, debates about basic epistemological issues, even though useful at one level, never seem to produce much resolution. Rather, they are repeated periodically, often covering the same ground."[103]**

While I would never claim to be of the same philosophical caliber as that of Stigler or Pfeffer, I would submit that another modification of the Christensen-Sundahl pyramid that might go some way towards squaring this circle. <u>Every theory should be accompanied with clear empirical conditions of falsification/ejection,</u> such that if those conditions are met, then the proponents of the theory will publicly agree that the theory must be ejected. Such conditions can be probabilistic in nature e.g. that a certain percentage of future studies must find results that clearly contradict the central predictions of the theory. Those conditions might be relative to another theory: that if a series of future retests indicates that if theory A produces more accurate predictions than theory B (even if neither theory is perfectly accurate), then B should be ejected in favor of A. Those conditions might even be contingent in nature: that if theory B provides

[102] Stigler 1978.
[103] Pfeffer 1993.

more accurate predictions within certain contexts than theory A, then theory A should be ejected within those contexts. {For example, it is highly plausible that TCE is applicable only for profit-seeking organizations but cannot explain the organizational behavior of non-profit organizations such as government agencies or NGO's, and so the use of TCE should be ejected from those contexts.} But whatever those conditions are, those conditions must be clearly stipulated in writing. If the theory is new, then those falsification conditions would have to accompany initial publication, perhaps within an online companion. The authors/proponents of extant theories should be encouraged to state clear falsification conditions for their theories, and if they refuse, then that refusal should be taken into consideration when future scholars choose whether to adopt those theories.

With clear falsification conditions in hand, future scholars will have a set target by which they can conduct the revalidation/retesting step, either to eject their own theory or to eject somebody else's. For example, as mentioned before, for either TCE or RDT to be promoted to true paradigmatic state would require continuous retesting of each theory until one reaches its ejection condition. The continuous retesting of NIT would result in either the (possibly contingent or relative) ejection of either NIT or of the theories of Porter and other strategy academics, depending on whether the true optimal state of organizations is unipolar or not.

To be sure, the recommendations that would remedy what ails management academia reach beyond merely clarifying the pervasive theoretical dissensus, helpful as that would be. I segue to these recommendations in the final chapter.

8 Final Conclusion: Proposals, Recommendations, and Final Thoughts

The past chapters have documented the litany of problems that management academia faces. The Popperian stance regarding the philosophy of science holds that theories generate value when they produce reliable, non-obvious predictions. For the most part, management academia currently generates neither. Rather, as I have shown, many theories are inherently tautological and therefore unfalsifiable. Others that are indeed falsifiable in principle prove not to be falsifiable in practice, either because the proponents of the theories engage in p-value data-dredging and and/or rummage through the armory of Chekhov's guns to avoid theory disconfirmation, or simply outright ignoring disconfirming evidence and persist in citing and building upon ostensibly falsified theories anyway.

Furthermore, replication is an endangered species, outright retraction is as extinct as the dinosaurs, and non-obvious predictions are similarly as rare as hen's teeth. The 'Exogeneity Revolution' held the promise of enforcing standards of causality and rigor upon purely-correlation (and hence unactionable) management theories of decades past. Yet the Exogeneity Revolution proved to be a revolution for which nobody - or at least none of the well-established management scholars - showed up. Instead, it has merely degenerated into a publishing encumberance that only the current generation of untenured management scholars must endure. If anything, the 'Exogeneity Revolution' has ironically been subverted into a 'counter-revolutionary' force that serves only to reinforce extant theories by established scholars while squelching new theories by young scholars from rising to prominence. Curiously and conveniently, not a single one of the influential management theories born during the pre-revolutionary golden age of the late 1960's-early-1980's[104] [105] has later been found to lack exogeneity. Finally, even regarding those rare instances where a community of scholars does seem to support a set of reliable predictions, progress is stymied when another community supports a set of opposing reliable predictions, with no clear way to adjudicate the conflict.

After perusing this parade of horribles, one might therefore reasonably wonder what – if anything – is to be done for management academia to progress beyond the epistemological quagmire in which it currently finds itself. In essence, the management field needs a Scientific Revolution of its own, similar to what the natural sciences experienced in centuries past. The question then becomes how best to ignite that Scientific Revolution.

To be clear, the predictive success of any such suggestions is necessarily speculative, for at this time, how/why the Scientific Revolution of the natural sciences occurred is still one of the greatest mysteries in history. Nevertheless, I first consider three popular suggestions for management academic reform - a re-emphasis upon qualitative researchers, practitioners, and of 'dedicated cadres' - regarding how management academia might progress. I reject each of them for being deficient in certain regards. I then propose what I believe to be the suggestion with the most promise, and which also happens to be aligned with the topics raised in prior chapters, with an appeal to history

[104] Davis 2010
[105] Colquitt 2005.

regarding why I believe that my proposal would enjoy the greatest probability of success. However, pace Pangloss, I fully recognize that my proposal stands a high chance of failure as well. I then turn to a 'interim proposal': perhaps the value that management academia can provide to the world is not through teaching management theories that are riven with dissensus, but rather teaching both the dissensus itself and also teaching research methodologies through which students can ascertain truths by themselves. Such a proposal could not only provide immediate and sizable pedagogical value to students, but might also serve to lay the groundwork for theoretical breakthrough for management academia in the future.

8.1 One Potential Proposal: How About More Qualitative Research?

A natural and commonplace reaction to the p-value graphs of Chapter 4 is that the state of quantitative research is suspect. As previously discussed in that chapter, quantitative researchers evidently have great freedom to generate the levels of statistical significance that they need to publish through model specification adjustment, subsample-analysis, or simply by collecting additional data until they obtain the result that they desire. Perhaps more commonly – as dictated by the Rosenthalian 'File Drawer Problem'[106] - quantitative researchers can simply refuse to submit findings that fail to produce significant results, and journal referees can likewise reject papers that lack significant findings: hence the dearth of findings with p-values just above the levels of significance. Indeed, Gerry Davis himself expressed the sentiments likely held by many others that the published literature exhibits "almost certainly widespread data fishing"[107]. A natural

[106] Rosenthal (1979)
[107] Davis (2010).

response would then be to place less emphasis upon quantitative research in favor of more qualitative research.

To be sure, certainly few people – least of all I – believe that rigorous qualitative research will continue to be an integral component of the management research toolkit. Such qualitative research is not only necessary when the variables of a particular phenomenon are so poorly characterized that a quantitative research project would be impossible to launch,[108] but also can provide a crucial sanity check when a phenomenon is sufficiently characterized to support a quantitative research enterprise. Qualitative research can elucidate new causal pathways while revalidating established ones, and have proven to be especially useful in ascertaining the beliefs and goals of individual managers whose decisions ultimately are the microfoundational basis for all of management academia.

However, to assert that management researchers should shift its focus towards more qualitative research simply because of the quantitative data-mining concerns of Chapter 4 is to take matters too far. In particular, doing so would be to exhibit strong sample-selection bias. While quantitative research certainly is riven with data-mining concerns, I have no reason to believe that qualitative research is not likewise riven with the same such concerns. Indeed, data-mining within qualitative research might well be more ubiquitous, simply because no comparable concepts to the p-value or statistical significance exist in the world of qualitative research, affording little opportunity to conduct the type of forensic quantitative analysis akin to my histograms.

[108] Edmondson & McManus (2009).

Furthermore, the land of qualitative research seems to be as bristling with Chekhov's Guns as is the realm of quantitative research. Pray tell, when has a qualitative paper ever proposed its own theory only to find that theory to be unsupported by its own data? When has a qualitative publication (published in a top journal) in the entire history of management academia ever been voluntarily retracted? When has a formerly heavily-cited qualitative management theory ever been abandoned/ejected for lack of empirical support? Therefore, from a prime-facie standpoint, there seems to be no reason to believe that qualitative research is any more reliable than is quantitative research. Hence any effort to replace quantitative research with qualitative research seems unlikely to allow management academia to become a truly mature discipline.

8.2 How About Bringing Back the Practitioners?

Another common refrain is to simply abolish management research by academics and instead revert back to the practitioner-centric business school faculties reminiscent of the state of affairs prior to the publication of the Ford/Carnegie Foundation reports of the 1950's that [109] A central impetus behind this proposal is that if academics – whether quantitative or qualitative – are perpetually driven to care only about publishable research findings, and the academic publication process is seemingly interested only in publishing papers that support a purported theory, then perhaps the wise course of action might simply be to hire faculty who are less subject to that publication pressure. Practitioners - who seldom participate in the academic publication process – might be a source of such faculty. Practitioners might furthermore serve as a repository of practical managerial

[109] Khurana 2007.

wisdom that could serve as a salutary counterbalance to the abstraction academic theories.

Akin to my feelings about qualitative research, I would also never deny the value that practitioner's experience can provide. Practitioners can undoubtedly provide insights into causal pathways and managerial decision-making of which pure academics are unaware. Practitioners may also be aware of subtle yet crucial variables that even qualitative researchers might not uncover with the tool available to qualitative researchers at this time. Perhaps more importantly, such practitioners might redirect the research questions of academic researchers towards truly business-centric topics rather than the pure social science research that is often times their wont. As the noted management academic critic Rakesh Khurana lamented: "Many of the discipline-trained scholars joining business school faculties were not intrinsically interested in business. Few were motivated in their research by a desire to examine the real problems that managers faced**."**[110] Practitioners' experience also provides key pedagogical credibility when teaching MBA and (especially) executive education students alike; indeed, this fact alone might be a crucial reason why more management academics should develop some practical experience before being allowed to teach.[111]

[110] Khurana 2007.

[111] It is indeed one of the great ironies in all of academia that management academia – in strong contrast to the panoply of other professional programs - law, engineering, medicine, education, architecture, nursing, etc. – tends to be taught by faculty that not only lack practical experience in the field which they are teaching, but seldom even hold any academic degrees in the program that they are teaching. Most management faculty at the top-ranked programs lack MBA degrees or even undergraduate business degrees, and many even lack doctorates in management (but rather hold PhD's in a pure social science), yet are somehow expected to teach business students and care about issues pertinent real-world management.

However, the hiring of more practitioners by business school faculties would likely not be wholly beneficial; key drawbacks befall the business school that replenishes their faculty with practitioners. The hiring of practitioners would likely be subject to strong sample selection towards success: few if any 'failed' managers would be deemed fit to be hirable by any top business school. Perhaps more importantly, even those practitioners who are hired would – like anybody in any field – be more privy to discussing their professional successes rather than their failures. Yet to discuss only/predominantly business successes would be to sample on the dependent variable: failures are just as instructive as successes.

Perhaps more crucially is the obvious point that few if any practitioners are trained researchers. They generally are not conversant in basic methodological concepts such as control groups, construct validity, causality, identification, data-mining, and the like. To be fair, the utility of such methodological knowledge for a practitioner may be limited, hence explaining why few current real-world managers currently possess such knowledge. Yet why those particular practitioners *who then join business school faculty* seldom develop such knowledge is far less obvious, particularly given the ample opportunities that the academic setting provides to develop such skills. As perhaps the archetypal example, consider Jim Collins, who after formerly working as a practitioner at McKinsey and Hewlett-Packard, later served on the Stanford GSB faculty for years before embarking on his celebrated career as a management guru. One might logically think - given Stanford GSB's reputation as not only one of the world's most prestigious B-schools, but that also heavily stresses academic rigor – that Collins should have known fully well that the research methodology he promulgated in ‘Good to Great’, ‘Built to Last’ and the rest of his oeuvre involved intensive data-mining. His remonstrations to the contrary regarding the statistical validity of his work - recall his infamous quote: “the probability [of randomly finding the ‘Good to Great’ findings] is less than 1 in 17 million”[112] - are therefore a case of Collins doth protesting too much. (Granted, the cynic could argue that Collins knows fully well that his research methodology is flawed and simply doesn’t care as long as it sells books. The onus to learn methodology would then be upon the *consumers* of such guru books – a point that I revisit later.)

One might also think that practitioners who later became business-school faculty might want to develop research knowledge to determine whether the decisions they made as a practitioner that they thought were correct at the time turned out to be wrong. For example, a practitioner who decided to, say, engage in a particular merger, and who later joined a business-school faculty might then determine through his own research that his former merger decision was inadvisable. Yet I struggle to think of even a single practitioner-turned-faculty who has actually declared through research that his past decisions and beliefs were wrong. That’s despite the fact that the much of the entire point of research is to demonstrate that what you once believed to be true is not actually true after all (pertaining to Popper’s point regarding the value of theory in providing non-obvious, yet reliable predictions).

Yet perhaps the most serious indictment of all regarding the value of practitioner’s knowledge is its sheer diversity of opinion. Regarding this standpoint, the sweep of

[112] Collins 2001.

practitioner's beliefs might well rival that of management academia, and therefore be as equally muddled. One need only glean the opinion pages and interviews published by business magazines such as Businessweek, or the roomfuls of biographies about business managers to find conflicting opinions regarding the best ways to manage a company. Should managers adopt an attitude of kindness towards their employees, or should they invoke an attitude of intimidation and psychological manipulation – perhaps even mining the dark depths of sociopathy – in the infamously ruthless style of Steve Jobs[113]? Should a company devote its attention to pleasing its current customers or inexorably seek new markets to conquer? It seems as if regardless of whatever opinion you might hold, you can find a practitioner who will support you.

[113] Isaacson, W. 2011. Also consider Steve Jobs' own quote to Robert X. Cringely: "Sometimes I can be an [expletive deleted]" at http://www.cringely.com/2011/10/05/steve-jobs-is-dead/

8.3 How about 'Dedicated Research Cadres'?

Another proposal might be to simply form an insulated, dedicated cadre consisting of a small team of researchers. The team members might comprise an eclectic mixture of quantitative researchers, qualitative researchers, and practitioners. Such a cadre would then conduct research in a rigorous manner while insulated from the pressures of the extant academic career ladder and the concomitant requirement to publish 'significant' results. That cadre would also be free to challenge its own ideas, even retracting research as necessary. Through dedication to the principles of the scientific method while shielded from distortionary academic publication pressures, such researchers might discover truly reliable, non-obvious theory that Popper denoted.

However, the success of such a cadre generally founders upon various practical concerns. First off, who would fund such a cadre given its insulation from the rest of the academic apparatus? Universities are understandably reluctant to support such an initiative, if for no other reason than institutional jealousy (if other faculty members must subject themselves to the academic publication process, why should others be exempt?) Perhaps more importantly, its success rests upon the assumption that it would indeed truly be rigorous regarding the research that it develops. The epistemological challenge is palpable - as an outsider, how would you know whether a particular cadre is indeed being rigorous? Indeed, it is entirely possible that such a cadre might actually be *less* rigorous than the standard academic process, despite all of its shortcomings. Opening your research to inspection and criticism from the greater academic community, at least in principle, increases the odds of finding errors and tempering unwarranted research claims. A small, closed group of researchers might arguably be more likely to promote spurious findings. Jim Collins, for example, launched a dedicated cadre consisting of a hand-picked group of researchers to develop his statistically dubious research findings of Good to Great and his other books. If he had shared his research with the greater academic community, perhaps the statistical problems of his research would have been revealed to him (although, again, the cynic might argue that Collins wouldn't have cared anyway, as his incentive is simply to sell books and propagate his fame).

But even assuming that such practical concerns could be overcome, perhaps far more importantly, to insulate oneself from the academic system means to close off channels of communication with academia and hence the ability to *change* academia. Akin to the philosophical quandary of the proverbial tree falling in the forest with nobody around to hear it, it likewise makes little difference to academia to discover a rigorous new finding if the rest of academia never knows about it and, more importantly, never chooses to adopt it into its corpus of knowledge. As a stark case in point, while the research by Clark Gilbert, Clay Christensen and their acolytes may indeed be highly rigorous and even celebrated in many circles - the term 'disruptive innovation' being one of the most widely used practitioner terms developed by management academia – the unfortunate fact is that relatively few modern-day management academics continue to develop that stream of research. Retrospectives regarding the most influential *academic* (as opposed to practitioner-oriented) management theories seldom mention even the work of

Christensen[114], let alone Gilbert or other scholars of that vein . Not to put a fine point on this topic, but I personally struggle to think of a single young scholar who is actively promoting Christensen-esque theory of disruptive innovation by actively publishing in top journals today, in stark contrast to the teeming hordes of young scholars who are, say, New Institutionalism or Principal-Agency Theory scholars.

8.4 Whither Ejection?

None of the aforementioned ideas are new. Indeed, each and every one of the aforementioned ideas to change management academia has been extensively tried. For example, qualitative researchers used to be predominant in business schools in the past and continue to be hired today, albeit in admittedly in ever-smaller numbers. Practitioners almost exclusively staffed business-school faculties in the days prior to the Ford/Carnegie Committees of the 1950's-1960's and still populate business-schools today. Indeed, it has become an increasingly common trope of world-famous managers upon retirement to serve a stint at a top business school – Jack Welch heading to MIT Sloan being a famous recent example. And dedicated research cadres such as Jim Collins Boulder Colorado management lab, along with the research divisions of countless consulting firms, exist to discover rigorous management research findings (although, to be sure, their impartiality is uncertain). Similarly, numerous business schools have also attempted various flavors of insulated cadres by launching discrete divisions and research centers – the Desautels Centre of Integrative Thinking at the University of Toronto being a rather prominent example.

[114] Examples would be Davis 2010, Scott 2009, Colquitt 2005. Miner 2003.

Perhaps most strikingly – at least until recent decades - Harvard Business School itself arguably proffered the most famous example of an insulated cadre on a grand scale. Until lately, HBS faculty members were notably less heavily invested in the academic publication process compared to their colleagues in peer schools.[115] Indeed, many HBS faculty were (and still are) neither considered to be pure academic researchers at all, nor was HBS considered to be a traditional academic business-school in the mold of a Wharton, Chicago-Booth, Stanford GSB, or MIT Sloan. The HBS DBA program was formerly not considered by the greater academic community to be a true research degree but rather primarily an ad-hoc training program for aspiring new HBS faculty – a stigma that lingers to this day. Many HBS faculty were entirely content to avoid academic publishing entirely, instead opting to publish cases and/or in practitioner-oriented articles in the Harvard Business Review. People could therefore spend their entire academic careers – doctoral training, initial jobs placement, tenure promotion, publishing – in HBS-controlled training and faculty positions, publishing journal articles and cases promulgated by HBS Publishing, and many did exactly that. If any business school possessed the vaunted prestige necessary to change the rest of management academia, surely it would have to be HBS.

Nevertheless, the proof of the pudding is in the eating. Perhaps the most compelling reason of all to abandon the notion that any re-emphasis upon qualitative research, practitioners, or dedicated cadres would work is the evidence of history. Even if qualitative researchers, practitioners, dedicated cadres, or some combination of all three

[115] This is a point that Khurana (2007) discusses at great length.

could indeed develop better theory than what currently exists today, it frankly wouldn't matter unless they could also then convince the rest of the academic world to abandon existing theories. Yet the fact remains that management academia continues to be plagued by a dearth of consensus regarding a set of reliable, non-obvious predictions that can be delivered to practitioners. More importantly, as per my aforementioned challenge, I struggle to think of a single formerly-widely-believed management theory that was later discredited as a matter of scientific consensus and therefore been ejected from the field. Such ejection is part and parcel of academic progress. What progress might be made by an academic field that never seems to reject any extant ideas?

As a stark counterexample, consider the watershed historical example of the Copernican Revolution that regarding the heliocentric model of the universe. While Copernicus's heliocentric theory indeed generates a set of truly non-obvious yet reliable predictions regarding when and where various astronomical events would occur that pales in comparison to Copernicus's <u>true</u> contribution. After all, heliocentric theories of the universe had been proposed since the days of antiquity – Aristarchus of Samos having proposed such a model in the 3rd century BC; and Aryabhata and other Indian astronomers having built a well-developed heliocentric model by the 5th century AD. Indeed, Copernicus's model actually provided <u>less</u> accurate empirical predictions than the extant astronomical models of his time. Copernicus's true contribution – which fully validates Copernicus's status as one of the most important people in world history - was in *convincing others to eject* the abandon the geocentric model of the universe that had been predominant until his time.

Yet as important as Copernicus's contribution may have been to convince astronomers to eject the extant geocentric model in favor of heliocentrism, even that specific contribution alone does not capture the full scope of his contribution to the world, for that specific contribution would be of interest only within the field of astronomy. The true scope of Copernicus's contribution was metascientific in scope: instituting a skeptical mindset towards all prior scientific knowledge and thereby laying the groundwork for ejection across the <u>vast sweep</u> of the scientific landscape, akin to the initial falling domino instigates a long sequence of subsequent dominoes to collapse. For example, Galileo's refinement of Copernicus's heliocentric model sparked the design of his famed falling-objects experiment, thereby ejecting the extant Aristotlean theory of gravity. Similarly, Newton, while building upon the work of Copernicus and Galileo, demonstrated that the predominant Aristotlean Theory of Impetus/Inertia was wrong, replacing it with Newtonian mechanics. The discrediting and ejection of Aristotle's physics of motion also incited greater skepticism and the ultimate ejection of Aristotle's theory of continuous matter in favor of atomism – which held that all matter is composed of tiny but discrete particles - ultimately instigating the ejection of the pseudoscience of alchemy in favor of the science of chemistry. That in turn paved the way for the ejection of ancient views of biology and physiology in favor of modern-day biology and medicine. Copernicus's true contribution was therefore knocking over the initial domino that eventually toppled/ejected several millennia of received scientific knowledge. Therefore the truly important question is regarding what might actually incite all of management academia to become more skeptical towards and eventually abandon extant theories. Given the plethora of extant management theories and their diverse – even mutually exclusive - predictions, some management theories must surely be wrong and therefore should be ejected. Yet clearly identifying wrong theories and then convincing management scholars to stop using them has proven to be a most Sisyphean task. Nevertheless, given the fact that qualitative researchers, practitioners, and dedicated cadres have all been tried repeatedly yet none of them have evidently succeeded in generating an ejection, a reprise of those ideas strikes me as folly. To repeat the old nostrum, the definition of insanity is doing the same thing over and over again while expecting different results.

8.4 You Say You Want a Revolution? Follow the Principles of Research

If the aforementioned strategies of more qualitative researchers, practitioners, and insulated cadres have all been tried and unfortunately been found wanting, then perhaps a more promising avenue would be for management academia to invoke a strategy that has yet to be tried. The most straightforward such strategy would be to actually adhere to the rigorous tenets of reliable, non-obvious research that management researchers have always claimed to have followed, yet the previous chapters have demonstrated otherwise. I would therefore 'merely' be requesting that management researchers actually follow their purported principles as enumerated in the previous chapters. I proffer the following suggestions that would aid the management academia system to either produce theories that successfully ascend the pyramid to its apex, or that generate ejecta.

Step 1: Identifying and Ejecting the Tautologies:

Dispending with tautologies would seemingly be elementary: the proponents of any particular management theory ought to provide an addendum stipulating that if the independent and dependent variables of a particular prediction of that theory are found to be equivalent, then that prediction should be declared to be tautological and therefore ejected Hence, the proponents of a theory should be able to stipulate a clear thought exercise demonstrating that the independent and dependent variables can in principle vary independently of each other such that they every single box of the aforementioned 2x2 matrix could potentially be populated. To that extent that the all of a theory's predictions are found to be tautological, the entire theory should be declared to be ejected.

Granted, such a step is easier said than done as the longevity of the surely-tautological Resource-Based View Theory (RBV) has proven. Journal editors should therefore openly challenge the proponents of RBV to clearly demonstrate the thought-exercise of, say, a firm enjoying long-term competitive advantage despite lacking any resources, or else concede that RBV is not an empirically testable theory and hence should be ejected. The same challenge should be extended towards other established and well-cited management theories that may indeed tautological.

Step 2: Ejecting the Obvious

The proponents of each management theory should clearly enumerate what they believe to be the non-obvious predictions of that theory along with a rationale for why they believe such to be non-obvious. For example, they could invoke my methodology and demonstrate that a randomly selected group of participants indeed found that set of predictions to be non-obvious. Alternatively, they might demonstrate that their predictions conflict with that of other, well-established management theories, indicating that the predictions in question are non-obvious to the supporters of those other theories. Importantly, the proponents of each management theory should be willing to concede that if their theory fails to generate predictions that are not obvious to anyone, then their theory should be ejected.

Furthermore, while every non-obvious prediction need not be tested immediately – many of Einstein's most non-obvious predictions regarding the theory of relativity remained untested until only recently – management theories should be rendered suspect until such time as their non-obvious predictions are empirically validated/replicated. The proponents of the theory themselves should be willing to actively temper academic enthusiasm for their theory until such validation has occurred. For example, they might publish an addendum warning actively discouraging other scholars from using that theory as a base upon to construct other theoretical contributions until such time as a certain set of non-obvious predictions have indeed been validated.

Step 3: Ejection Via Validation/Revalidation

To encourage the production of ejecta via the generation of non-significant results and the concomitant reduction of data-dredging and Chekhov's Guns necessarily requires tackling the issue of publication bias, an admittedly Herculean task. Numerous academic disciplines have grappled with publication bias with varying rates of success. The root cause of data-dredging/publication-bias is that referees invariably prefer to find statistically significant results that support the theory in question. They have also been demonstrated to prefer results that conform to the referee's own personal predilections, and tend to reject papers not for lack of rigor but simply because the results fail to conform to their own beliefs[116] – quite the epistemological problem within a field such as management academia that is characterized by such diversity of beliefs. Given that referees have proven themselves to be untrustworthy judges of results, one relatively easy reform that could readily be incorporated within today's publication process is a <u>*'Blinded-Results´*</u> peer-review process. Referees would be allowed to inspect the entire front-ends of submitted papers – including methodology, description of the dataset with summary statistics, and even the structures of the tables themselves, but not any results themselves. Any discussion that pertains to those results would similarly be redacted. Referees would judge whether the paper should be accepted or not based only on the sections that they are allowed to inspect and nothing more. Once a paper is accepted would the full paper be unblended and published. The incentives for scholars to dredge their data for significant results would then be eliminated because they would know that referees are barred from viewing those results and therefore cannot affect the acceptance/rejection decision.

[116] Mahoney 1977.

If such a publication reform proves to be too abrupt, one might imagine a '*Two-Stage'* refereeing process. The first stage would be akin to the aforementioned process: referees would make a decision to accept conditionally without being allowed to inspect the results. However, if such a conditional acceptance is granted, then the process would move to a second step where referees would be allowed to now view the results. While referees could then choose to reject at that second step, they would then have to justify why had been previously willing to accept the paper only to now object only upon knowing the results. Hence, the second-stage burden of proof would rest squarely and strongly upon the shoulders of the referees rather than upon the authors. Editors would be strongly encouraged to veto any such second-stage rejections unless referees can present impeccable reasons for their change of heart.

Furthermore, given the unreliable state of the current literature base, if I had my druthers, I would enforce a moratorium, or at least a vast reduction, for several years regarding the publication of new articles in the top journals in favor of replications of existing articles. The most 'important' articles – as judged by either citation count, survey that measured influence or other such instrument – would be targeted for a series of replications. Such replications would not be confined to being mere reproductions of the original study using the original data, as valuable as such an exercise might be. Rather, they would be conducted upon new data derived from new settings. Those theories that fail to replicate would be ejected. Furthermore, such replications would be conducted using the latest methodological techniques, especially including techniques that had yet to be invented at the time of original publication of the theory. After all, if current papers are forced to survive the gauntlet of modern-day empirical testing, prior theories should do likewise. Regarding the seemingly interminable conflicts in predictions – the aforementioned conflict between the Resource-Dependency Theory school versus the Transaction-Cost Economics school being a preeminent example - journals could invoke the notion of 'adversarial collaboration'. Each school's proponents would be required to agree upon a set of (ideally, non-obvious) predictions regarding future data for which one school would be declared victorious over the other which should be rejected. Future data is, by definition, data for which nobody has access and therefore for which nobody can dredge and overfitted model. The predictions of each school would then be compared to future outcomes, and the school that consistently fails to generate reliable predictions relative to that of the other school will be ejected. If one particular school refuses to participate at all, that fact itself should be interpreted as a strong indication that that school should be ejected. After all, what epistemological value does a theory truly have if its very own proponents forsake it for generating predictions upon future data? However, insofar that neither school generates reliable predictions, or, more likely, that the two schools cannot even mutually agree upon a set of victory/defeat conditions in the first place, then both schools would be rendered highly suspect. What epistemological value does a theory have if its proponents cannot agree on the conditions of defeat?

Step 4: Ejection for Lack of Causality

Given my druthers, I would propose that some of the top journals should launch a moratorium upon new articles in favor of clear replications, using causal methodologies, of highly influential extant theories.   The proponents of such theories should be required to present a clear causal path diagram that elucidates what they believe their key causal pathways to be - something that for many older theories has to date never clearly been performed. Each of the resulting pathways should then be rigorously analyzed and confirmed with the correct causal direction, along with a relevantly large magnitude size. For example, regarding Powell & DiMaggio's New Institutionalism theory – arguably the most influential theory in management history[117] - does the adoption of 'isomorphic' organizational structures truly cause greater legitimacy and therefore superior performance as the theory holds, or is the arrow of causality reversed: superior performance actually causes the adoption of isomorphic structure?   Or consider Contingency Theory, which holds that firms that adopt organizational structures that align with (and hence are contingent with) their environment causes superior performance. Perhaps the arrow of causality is similarly reversed: superior performance causes firms to adopt organizational structures that align with their environment. Alternatively, perhaps those phenomena are characterized by simultaneous causality in which case the magnitudes of the forward and reverse causal pathways should each be carefully disentangled and estimated.   Perhaps a confounding variable – whose presence was unknown during the days of publication of the theory in question but is known today – is ultimately driving the theory's predictions. . Given causality's central role in providing practitioners with levers to effect outcomes, those theories whose causal

---

[117]

content is found wanting may be ejected or at least relegated to being merely correlational in nature.

Therefore, ideally and to the extent feasible, journals would establish and oversee a series of randomized controlled trials that would test as many extant theories as feasible. Naturally, the opportunities to run true randomized controlled trials are rare. Therefore, if randomized controlled trials are not forthcoming, then at least extant theories should be subjected to the same types of rigorous causal methodologies – matching, difference-in-difference, regression discontinuity, instrumental variables, etc. – to which current papers are forced to endure.

One might also view such a policy from the standpoint of 'generational equity'. If the work of the current generation of management scholars should be subjected to the excruciatingly strenuous standards of modern-day methodological rigor, then why should the work of prior generations of scholars be exempt? The entire premise underlying the Exogeneity Revolution is that rigorous causal methodology provides crucial epistemological insight regarding whether manipulation of the independent variable would indeed cause the dependent variable to vary, hence exposing a key lever that practitioners can potentially manipulate. If that is true, then prior theories should be subjected to the same level of causal epistemological scrutiny. On the other hand, if such causal methodologies fail to provide epistemological value such that older theories need not be examined for rigorous causality, then new theories ought not to be subject to such rigorous methodologies either. What's fair is fair. What scientific progress would be possible if older theories are forever exempted from modern-day methodological scrutiny by a de-facto grandfather clause?

8.5 Two Immediate Deliverables: Dissensus and Pedagogy

The reforms discussed in the previous section would surely strike the more jaundiced management academic scholar as ingenuously optimistic, indeed Panglossian in scope. Most of the suggested reforms would require the direct intervention of the editors of the A-level journals to enforce replication and revalidation. Perhaps more importantly, it would require the acquiescence if not the outright cooperation of eminent scholars in agreeing to subject their own theories that established their status to a crucible of testing to prove that those theories indeed generate reliable, non-obvious, causal results on pain of ejection. Both journal editors and established scholars have little to gain and much to lose from such reform, for what if foundational theories underpinning those journals and the status of established scholars are found to be ejected? Younger, unestablished scholars whose theories would replace the ejected older theories would naturally be more receptive towards such reforms. But such scholars, because, they are young and unestablished, lack the institutional power to implement any reforms. I readily concede that I wouldn't hold my breath waiting for such reforms to be enacted.

If that is the case, then that raises the natural question: What shall we do with an academic field that never seems to generate any progress? What to do with a field that never ejects any theories, but rather whose theoretical landscape only ever seems to expand monotonically? What to do with a field whose track record is one of continual disappointment in delivering not the consensus reliable, non-obvious, useful prediction rules that its target audience desires, but instead delivers nothing but dissensus? Pity the poor practitioner who simply wants to know the best leadership style, the best competitive strategy, and/or the best way to manage personnel, yet who is then confronted with the bewildering array of management theories that support any number of a myriad of potential answers.

However, therein perhaps lies hope. **Perhaps that dissensus itself, at least for now, could be one key deliverable that management academia can deliver to the world.** Indeed, that very dissensus might satisfy the tenets comprising a theoretical paradigm laid forth within this dissertation. The dissensus certainly seems to be reliable: evidence of such dissensus being easily ascertained and even "replicated/revalidated" by periodically drawing random samples of the management academic literature and perusing the resulting conflicting theoretical predictions.[118] Furthermore, such dissensus is arguably non-obvious to practitioners, particularly given the enduring popularity of management gurus whose nostrums are predicated upon the illusion of consensus. Would management guru works such as Good to Great or In Search of Excellence truly have enjoyed the record sales that they did if practitioners were actually aware of the fundamental dissensus regarding the efficacy of the strategies proposed by those books – or regarding the efficacy of any business strategy for that matter? Color me doubtful.

[118] For example, I recently drew several random samples of 50 papers each from A-level management journals and found that inevitably each sample would include some papers whose theoretical bases conflict with the theoretical bases of other papers within the same sample (Transaction Cost Economics vs. Resource Dependency Theory, New Institutional Theory vs. Contingency Theory, etc.} Indeed, sometimes those conflicting theories bases would be invoked <u>within the very same paper.</u>

True wisdom is knowing what you don't know, Confucius once postulated. Right now, despite the efforts of legions of intelligent and diligent management scholars and the ever-increasing availability of datasets and methodological tools, not only do we still frankly know very little about management, we seem to lack the wisdom to admit that lack of knowledge to ourselves, something that the management academic community should immediately remedy. In the absence of reliable knowledge, epistemological humility and skepticism therefore seems to be the most appropriate stance to take.

The most direct implication of such a philosophical stance is that management gurus, consulting firms, and others who proffer strong claims regarding how to improve management/strategy practices ought to severely temper their claims or should be confronted with healthy, harsh skepticism – and the more grandiose their claims, the more skepticism that is warranted. Extraordinary claims require extraordinary evidence. Given that management academia, despite its cutting-edge methodological techniques and disciplinary training, has still not been able to arrive at anything approaching a clear consensus regarding optimal management/strategy techniques, it seems rather unlikely that any guru or consultant could accomplish such a feat. Those like Jim Collins who continue to insist that they have indeed uncovered such a technique ought to be confronted at every opportunity with a challenge to either submit their findings to a peer-reviewed journal or – better yet - a public set of predictions that would effective validate their models upon future data, should they dare. Perhaps the most aggressive option of all would be that gurus and consultants who persist in promoting strong claims should be publicly challenged to submit their evidence to a bona-fide academic journal for formal inspection and peer-review. If they truly believe their findings to be robust, then their findings ought to survive whatever academic review process that the journals would have them undergo. Either that, or they should publicly explain why their supposedly robust insights need not undergo such review, while academic claims must.

Indeed, a related contribution that management academia could provide immediately is to **provide a public website and data repository that serves to debunk, fact-check, or (in rare cases) perhaps even confirm the numerous prominent management claims continually invoked by gurus, consultants, practitioners and the general business press**. As an analogy, highly trafficked websites such as Snopes.com and TruthorFiction.com serve to discredit the widely circulated stories regarding drugged travelers having awoken in ice-filled bathtubs with their kidneys stolen by organ thieves or of movie-goers contracting AIDS from bloody needles that were deliberately implanted in their seats. Similarly, shows such as Mythbusters have debunked the claim that one can synthesize 99% pure crystal methamphetamine that nevertheless has a distinctive blue tint as the TV show Breaking Bad would have you believe. In other words, society apparently is willing to exert tremendous effort in debunking horror-movie-style urban legends and claims presented in fictional TV shows. Yet in stark and ironic contrast, the strongly provocative claims of dubious veracity routinely advanced by management gurus and consultants not only routinely go unchallenged, such claims serve as the ingredients for best-selling management books and lecture-circuit fees. Surely business-schools could deliver immediate value to the world by launching a 'Business-Snopes.com'. I could imagine that the first three entries of Business-Snopes.com might be regarding the three guru examples discussed previously in this dissertation: that 'In Search of Excellence' had no control group, that the 'Core Competencies' practitioner theory is tautological, and that Jim Collins dredged his data.

However, regarding the issue of epistemological humility and the public debunking of strong managerial claims, a reasonable question might be: wouldn't that vitiate the vast sweep of the teaching at elite MBA and executive-education as we know them today? After all, the underlying premise behind top-flight business schools is that their professors offer insights that average business schools cannot. Yet the adoption of epistemological humility would necessitate publicly admitting that management academia – top B-schools included – still has yet to deliver a consensus body of theory that generates reliable, non-obvious prediction rules. Furthermore, many of the professors at the top B-schools have lately been behaving, quite frankly, as little more than glorified gurus. Their grand ideas therefore would likely be publicly discredited by any Business-snopes.com site. What would business schools now be left to teach? What now would justify the notion that elite business schools offer better teaching than do lower-ranked schools?

Allow me to propose a method that might square that circle. While <u>management theories and grand ideas</u> themselves might well be of questionable epistemological value, the <u>underlying methodologies</u> seem to have attained widespread consensus.[119] For example,

[119] What methodological debates do exist within management or the sciences as a whole generally tends to revolve around <u>under what conditions</u> is a particular methodology is useful or when it is not, particularly given practical considerations However, assuming that those conditions hold, there is little dispute regarding which methodologies should be utilized. For example there seems to be little if any dispute

there seems to be widespread agreement, not just within management, but within all of academia, that the ideal standard to identify clear causality is through a randomized controlled experiment. The other methods for ascertaining causality - exogenous shocks, matching/regression, panel data, etc. - comprise ways to approximate the ideal randomized controlled experiment. All of these research techniques require that independent and dependent variables not be defined in a tautological manner. They also demand that a control group not only be invoked, but should also be justified as to why the control group in question is appropriate. Without a randomized control trial, one must be constantly aware of threats to validity such as potential reverse causality or confounding variables. If one is making a generalizable claim, then one needs a representative sample. If one isn't even sure what the variables or the potential causal pathways are in the first place, then one would probably want to use a qualitative field study, perhaps even an ethnography, to explore the environment.

I therefore propose that business schools teach a course entitled 'Research Methodology & Evidence-Based Management' to MBA and Exec-Ed Students. Such a proposal would be akin to teaching a hungry man how to fish rather than just providing him with fish. Rather than handing students a panoply of grand business ideas of dubious veracity, we would instead be providing them with a set of tools with which they could rigorously develop business knowledge for themselves. Perhaps just as importantly, they would learn how to critically assess the business ideas proposed by others. Some of those students will likely become hobbyist business debunkers

---

regarding whether, say, the FDA should continue to insist upon randomized controlled trials as a gold standard when assessing drug candidates.

themselves, contributing new entries to Business-Snopes.com, in the same manner that amateur sleuths today debunk urban legends as a hobby. Such training would immunize those students from the pseudo-rigor offered by the management guru or consulting firm. It might even spark some of them to participate in business academia, either by reading academic journals now that they are armed with the necessary training to understand them, by perhaps trying to publish in them, or even sparking interest in some of them to join business academia. Given that most MBA's and (especially) exec-ed students have extensive practical experience, their participation in business academia might just be the spark that management academia needs to resolve its dissensus and mature into a true science.

To be clear, methodological training for MBA's and exec-ed students does not necessarily mean that they must learn heavily quantitative econometrics. While I certainly have no objections towards them learning that if they so wish, the core concepts of causality and methodology can readily be conveyed in non-mathematical format through use of the Directed Graphs that I invoked in previous chapters. One could imagine an MBA methodological course that walked students through a series of popular guru books such as Good to Great or In Search of Excellence that exposed their methodological flaws, and then demonstrated in graphical format how one might actually go about designing a methodology to properly validating the claims of those books. A final class project could consist of students taking a current guru book or otherwise widely held management idea and then proposing a rigorous methodology to test that idea.

An interesting feature of this proposal is that it would not undermine the current prestige-stratification system of the extant B-schools, but would instead likely replicate it. After all, the most prestigious B-schools tend to be the ones who conduct the most research. Hence, they would tend to be populated with faculty who are the most capable of offering

8.6 Finale

An academic field generates progress insofar as it produces a consensus opinion of empirically testable, reliable, non-obvious results. Unfortunately, the field of management has yet to succeed on any of those metrics, nor does it seem poised to do so anytime in the foreseeable future. Other than perhaps the most obvious of bromides - women being more feminine than men being the perhaps the archetypal example – management is pervaded by not only dissensus regarding mutually-contradictory predictions but also by an intractable 'meta-dissensus' regarding how the dissensus of its predictions can be resolved in the first place. For example, the debate about whether organizations optimally tend to evolve towards the same organizational form as New Institutional Theory would hold, or whether they will evolve towards a kaleidoscope of varieties as (Neo)Contingency Theory would predict, seems to be no closer to resolution than it was when the question was first broached decades ago. Pity the poor practitioner who peruses the management literature in search of answers only to discover that the literature can essentially support whatever prediction he wants. {In stark contract, the management gurus and consultants likely benefit from the dissensus of the literature precisely because it can support whatever prediction they want!}

The fundamental problem is that the field of management has yet to generate any true paradigms. No management theories have successfully ascended to the apex of the Christensen-Carlisle pyramid. For a variety of reasons do theories fail to ascend the pyramid – whether through not generating falsifiable/testable predictions at all, to generating only obvious predictions, to generating predictions that have not been truly validated and replicated. Perhaps most importantly of all, especially given the mutually exclusive predictions that various management theories generate, the <u>ejection</u> of failed theories is the true key to progress of the field.

However, management academia has no history of ejecting any once-popular theories; the management theoretical landscape is never pruned but rather grows monotonically like kudzu. Certain proposals to reform management academia –the elevation of qualitative research, the greater use of practitioners, or the leveraging of dedicated/insulated cadres - seem unlikely to instigate a culture of ejection. While such proposals might well succeed in developing new theories that are reliable and non-obvious, they would still likely fail to generate a <u>consensus</u> among the community at large by convincing it to eject past theories. Rather, such proposals would be akin to the proverbial tree falling in the woods where nobody is around to hear it, does it truly generate any sound? Analogously, if certain researchers successfully develop truly rigorous theory yet the community fails to incorporate it as part of its consensus body of knowledge, does that theory truly generate impact? We should recall that Copernicus's historical impact stems not from the misconception that he was the first person to ever propose a heliocentric model of the universe, for such models had been proposed centuries beforehand. Rather, his true impact stemmed from the fact that he convinced the community to eject extant geocentric models in favor of his.

One possible way to then generate greater ejection is to enforce the tenets of the scientific method by rigorously re-examining extant theories for their falsifiability, their non-obviousness, and their predictive performance. Journals should directly pit theories against each other through the system of adversarial collaboration where proponents of each theory would be required to stipulate clear predictions with falsification conditions. Replication and retraction should not be the rare oddities of the literature that they are today but rather should be commonplace. One might even entertain the notion of public scientific wagers amongst different schools of management thought of the same vein as the celebrated Simon-Ehrlich wager or the Thorne-Hawking-Preskill wager, where the losers of such wagers would have to publicly concede that they lost.

However, such reforms will assuredly require at least the consent if not the outright intervention of the academic journals; such consent is unlikely to be forthcoming anytime soon. Hence, as an interim deliverable to the rest of the world, management academia could declare that it has simply failed to generate any consensus regarding most of the important managerially relevant questions of modern times. The current state of literature simply cannot support any consensus answers regarding the 'best' style of leadership, the 'best' manner to foster innovation/entrepreneurship, the 'best' competitive strategy to implement, or the 'best' way to manage corporate culture – and such lack of knowledge is itself important knowledge. Through the public airing of such dissensus, the management academic community could provide tremendous practical value to the world by directly challenging the bombastic, unwarranted claims of the legions of gurus and consultants who do purport to know the answers to those questions.

Management academia could also provide tremendous pedagogical value to the world by teaching students and practitioners to be more rigorous in their methodological approach. Insofar as the methodological tools of management researchers are useful in elucidating truth and causality, those tools should be taught to students. {On the other hand, if those tools are not useful, then management researchers ought to ask themselves the discomfiting question of why they persist in using them.}   Rather than merely inculcate greater skepticism amongst students regarding the claims made by In Search of Excellence or Good to Great, students should be taught why they should be skeptical, and, more importantly, how one might go about rigorously testing those claims. Such skepticism would not only impart greater wisdom amongst practitioners by demonstrating a healthy appreciation for what we don't know, but might also spark greater appreciation for rigorous research, which might then ultimately spur true research progress. One can dream.

Abbreviated Bibliography

Barney, J. 1991. Firm Resources and Sustained Competitive Advantage. Journal of Management. 17(1).

Barney, J. 2001. Is the Resource Based View a Useful Perspective for Strategic Management Research? Yes. Academy of Management Review. 26(1).

Carlile, Paul & Clayton M. Christensen. 2004. The Cycles of Theory Building in Management Research.

Davis, Gerald F. 2010. Do Theories of Organizations Progress? Organizational Research Methods. 13:690.

Christensen, Clayton M. and David M. Sundahl. 2001. The Process of Theory Building. Draft.

Collins, Jim. 2001. Good to Great. HarperCollins Publishers.

Collins, Jim. 2009. How the Mighty Fall: And Why Some Companies Never Give In. JimCollins Publishing.

Collins, Jim and Jerry Porras. 1994. Built to Last: Successful Habits of Highly Visionary Companies. HarperBusiness.

Collins, Jim and Morton Hansen. 2001. Great by Choice: Uncertainty, Chaos, and Luck – Why Some Thrive Despite Them All. HarperBusiness.

Davis, Gerald F. 2006. Mechanisms and the Theory of Organizations. Journal of Management Inquiry. 15:114.

Davis, Gerald F. and Christopher Marquis. 2005. Prospects for Organization Theory in the Early Twenty-First Century: Institutional Fields and Mechanisms. 16:4.

David, Gerald F. and Mayer N. Zald. 2008. Sociological Classics and the Canon of the Study of Organizations. Paul Adler.

Davis, James A. 1994. What's Wrong With Sociology? Sociological Forum. 9:2.

Davis, Murray S. 1971. That's Interesting: Towards a Phenomenology of Sociology and a Sociology of Phenomenology. Philosophy of the Social Sciences. 1:4.

DeLong, Brad & Kevin Lang. 1992. Are All Economic Hypotheses False? Journal of Political Economy. 100:6.

DiMaggio Paul J. and Walter W. Powell. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. American Sociological Review. 48:2.

Hambrick, Donald C. 1994. 1993 Presidential Address: What if the Academy Actually Mattered? Academy of Management Review. 19:1.

Hambrick, Donald C. 2004. The Disintegration of Strategic Management: It's Time to Consolidate Our Gains. Strategic Organization. 2:1.

Hambrick, Donald C. 2007. The Field of Management's Devotion to Theory: Too Much of a Good Thing? Academy of Management Journal. 50:6.

Hamel, Gary and C.K. Prahalad. 1996. Competing for the Future. Harvard Business Review Press.

Harlow, Lisa L., Stanley A. Mulaik, and James H. Steiger (Eds.) 1997. What if There Were No Significance Tests? Lawrence Erlbaum Associates.

Khurana, Rakesh. 2010. From Higher Aims to Hired Hands: The Social Transformation of American Business Schools and the Unfulfilled Promise of Management as a Profession. Princeton University Press.

Kuhn, Thomas. 2012. The Structure of Scientific Revolutions (50th Anniversary Edition). University of Chicago Press.

Lakatos, Imre. For and Against Method: Including Lakatos's Lectures on Scientific Method and the Lakatos-Feyeraband Correspondance. University of Chicago Press.

Lakatos, Imre, John Worrall, and Greg Currie. 1980. Mathematics, Science, and Epistemology: Volume 2, Philosophical Papers. Cambridge University Press

Lakatos, Imre, John Worrall and Greg Currie. 1980. The Methodology of Scientific Research Programmes: Volume 1 Philosophical Papers. Cambridge University Press
Leamer, Edward. 1978. Specification Searches. John Wiley & Sons.

McGrath, Rita Gunther. 2007. No Longer a Stepchild: How the Management Field Came Into Its Own. Academy of Management Journal. 50:6.

McKinley, William. 1993. The Uniqueness Value and its Consequences for Organization Studies. Journal of Management Inquiry. 2:284.

McKinley, William. 2003. Postmodernism and Management: Pros, Cons, and the Alternative Research in the Sociology of Organizations. 21.

McKinley, William 2007. Managing Knowledge in Organizational Studies Through Instrumentalism. Organization. 14:1.

McKinley, William. 2010. Organizational Theory Development: Displacement of Ends? Organization Studies. 31:47.

McKinley, William and Mark A. Mone. 1998. The Re-Construction of Organizational Studies: Wrestling with Incommensurability. Organization. 5:169.

McKinley, William, Mark A. Mone, and Gweyan Moon. 1999. Determinants and Development of Schools in Organization Theory. Academy of Management Review. 24:4

Miller, Danny. 2007. Paradigm Prison, or in Praise of Atheoretic Research. 5:2.

Miller, Danny. 2009. The Social Ecology of Fads: A Commentary on Starbuck's "The Constant Causes of Never-ending Faddishness in the Behavioral and Social Sciences". Scandinavian Journal of Management. 25.

Miner, John B. 1973. Management Process: Theory, Research, and Practice. Collier McMillan Inc.

Miner, John B. 1980. Theories of Organizational Behavior. Dryden Press.

Miner, John B. 1984. The Validity and Usefulness of Theories in an Emerging Organizational Science. Academy of Management Review. 9:2.

Miner, John B. 2002. Organizational Behavior: Foundations, Theories, and Analyses. Oxford University Press.

Miner, John B. 2003. The Rated Importance, Scientific Validity, and Practical Usefulness of Organizational Behavior Theories: A Quantitative Review. Academy of Management Learning & Education. 2:3.

Peters, Thomas J. and Robert H. Waterman Jr. 1982. In Search of Excellence: Lessons from America's Best Run Companies. Warner Books. New York.

Pfeffer, Jeffrey. 1981. Four Laws of Organizational Research. Perspectives on Organizational Design and Behavior (Andrew H. Van de Ven and William F. Joyce Editors)

Pfeffer, Jeffrey 1982. Organizations and Organization Theory. Pitman.

Pfeffer, Jeffrey. 1993. Barriers to the Advancement of Organizational Science: Paradigm Development as a Dependent Variable. Academy of Management Review. 18:4.

Pfeffer, Jeffrey. 1995. Mortality, Reproducibility, and the Persistence of Styles of Theory. Organization Science. 6:6.

Pfeffer, Jeffrey. 1997. New Directions for Organization Theory: Problems and Prospects. Amazon Digital Services.

Pfeffer, Jeffrey. 2005. Why do Bad Management Theories Persist? Academy of Management Learning & Education. 4:1.

Pfeffer, Jeffrey. 2007. A Modest Proposal: How We Might Change the Process and Product of Managerial Research. Academy of Management Journal. 50:6.

Pfeffer, Jeffrey. 2008. What Ever Happened to Pragmatism? Journal of Management Inquiry. 17:57.

Pfeffer, Jeffrey 2009. Renaissance and Renewal in Management Studies: Relevance Regained. European Management Review. 6.

Pfeffer, Jeffrey and Christina T. Fong. 2002. The End of Business Schools? Less Success Than Meets the Eye. Academy of Management Learning & Education. 1:1.

Pfeffer, Jeffrey and Christina T. Fong. 2005. Building Organization Theory from First Principles: The Self-Enhancement Motive and Understanding Power and Influence. Organization Science. 16:4.

Pfeffer, Jeffrey and Gerald R. Salancik. 2003. The External Control of Organizations. Stanford University Press. Stanford, CA.

Pfeffer, Jeffrey and Robert Sutton. 2006. Evidence-Based Management. Harvard Business Review OnPoint.

Pfeffer, Jeffrey and Robert Sutton. 2006. Hard Facts, Dangerous Truths and Total Nonsense: Profiting from Evidence-Based Management. Harvard Business School Publishing.

Popper, Karl. 2002. The Logic of Scientific Discovery. Routledge Classics.

Powell, Thomas. 2001. Competitive Advantage: Logical and Philosophical Considerations. Strategic Management Journal. 22.

Powell, Thomas. 2002. The Philosophy of Strategy. Strategic Management Journal. 23.

Prahalad, C.K. and Gary Hamel. 1990. The Core Competence of the Corporation. Harvard Business Review.

Priem, Richard L. and John E. Butler 2001. Is the Resource Based "View" a Useful Perspective for Strategic Management Research? Academy of Management Review. 26:1.

Priem, Richard L. and John E. Butler. 2001. Tautology in the Resource-Based View and the Implications of Externally Determined Resource Value: Further Comments. Academy of Management Review. 26:1.

Rosenthal, Robert. 1979. The "File Drawer Problem" and Tolerance for Null Results. Psychological Bulletin. 86:3.

Rosenzweig, Phil. 2007. The Halo Effect: …and the Eight Other Business Delusions That Deceive Managers. Simon and Schuster.

Sala-I-Martin, Xavier X. 1997. I Just Ran Two Million Regressions. American Economic Review. 87:2.

Samuelson, Paul. 1969. The Way of an Economist. in International Economic Relations: Proceedings of the Third Congress of the International Economic Association (P.A. Samuelson Ed.).

Scott, W. Richard and Gerald F. Davis. 2006. Organizations and Organizing: Rational, Natural, and Open Perspectives. Pearson.

Starbuck, William H. 2003. Turning Lemons into Lemonade: Where is the Value in Peer Reviews? Journal of Management Inquiry. 12:4.

Starbuck, William H. 2005. How Much Better are the Most-Prestigious Journals? The Statistics of Academic Publication. Organization Science. 16:2.

Starbuck, William H. 2007. Living in Mythical Spaces. Organization Studies. 28:21.

Starbuck, William H. 2009. The Constant Cases of Never-Ending Faddishness in the Behavioral and Social Sciences. Scandinavian Journal of Management. 25.

Stern, Robert N. and Stephen R. Barley. Organizations and Social Systems: Organization Theory's Neglected Mandate. Administrative Science Quarterly.

Stigler, George. J. 1978. The Literature of Economics: The Case of the Kinked Oligopoly Demand Curve. Economic Inquiry. 16.

Tsang, Eric W.K. 2006. Behavioral Assumptions and Theory Development: The Case of Transaction Cost Economics. Strategic Management Journal. 27.

Webster, Jane and William H. Starbuck. 1988. Theory Building in Industrial and Organizational Psychology. International Review of Industrial and Organizational Psychology (C.L. Cooper and I. Robertson Eds.)