4-2020

Explanation and Rigor in Management Theorizing: Which Theory-Building Criteria Make for an Influential Contribution?

Jaume Franquesa
Ohio Northern University, j-franquesa@onu.edu

William Acar
Kent State University

Jino Mwaka
University of the Sacred Heart Gulu

Recommended Citation

Franquesa, J., Acar, W., and Mwaka, J.O. (2020). Explanation and Rigor in Management Theorizing: Which Theory-Building Criteria Make for an Influential Contribution? Journal of Management Research, 20(2), 75–98.

Journal of Management Research
Vol. 20, No. 2, April - June 2020, pp. 75-98

Explanation and Rigor in Management Theorizing: Which Theory-Building Criteria Make for an Influential Contribution?

Jaume Franquesa, William Acar and Jino O. Mwaka

Abstract

This paper advances and tests a comprehensive but parsimonious model of theory-building/evaluation criteria in management and organization science, and of the relationships between these criteria and a theory's eventual prominence within the discipline. The model is tested using survey data in which knowledgeable scholars are asked to provide a detailed assessment of the traits of one of a few well-known seminal theoretical articles that are used as vehicles. The results support the presence of three distinct but correlated dimensions of theory evaluation (*novelty*, *extendibility*, and *relevance to practice*) and further provide confirmatory evidence of an overarching, second-order construct, which we term the *explanatory meaningfulness* of a theoretical exposition. Moreover, we find this construct to be a positive and strong predictor of the subsequent perceived importance of a theoretical article among management scholars. By contrast, the *logical consistency* and the *falsifiability* of the theoretical exposition were not significantly associated with its perceived importance. Paradoxically, our findings suggest that the most influential theoretical articles in management are those that offer greater explanatory value (stemming from their originality and perceived usefulness for research and practice), regardless of other aspects associated with conventional prescriptions for rigorous theory-making.

Keywords: Management Theory, Theory Building, Theory Evaluation, Explanatory Meaningfulness, Scientific Rigor, Scholarly Impact

INTRODUCTION

Theories are central to the advancement of knowledge in any field of scientific inquiry. They guide the research agenda by organizing the complex empirical world and pointing to the important questions (Bacharach, 1989), suggesting appropriate experiments to answer such questions (Kaplan, 1964), and offering insightful predictions (Hambrick, 2007). Similarly, theories assist in organizing and making sense of what would otherwise be disconnected propositions and empirical findings, thus facilitating the effective accumulation of knowledge (Haveman, Mahoney, & Mannix, 2019; Turner, 1985). The precise definition and desirable qualities of theory, however, are matters of some dispute within the organization studies and management literature. There have been long-lasting and recurrent debates about what constitutes a theory (e.g., DiMaggio, 1995; Shapira, 2011; Sutton & Staw, 1995), and about the requisite elements of a "good" theory and thus the criteria for the evaluation of theoretical contributions (e.g., Ackoff, 1962; Bacharach, 1989; Corley & Gioia, 2011).

Jaume Franquesa (Corresponding Author)
James F. Dicke College of Business Administration
Ohio Northern University, Ada, OH 45810

William Acar
Graduate School of Management
Kent State University, Kent, OH 44242

Jino O. Mwaka
University of the Sacred Heart
Gulu, Uganda

In broad terms, the different schools of thought in the theory building/theory evaluation literature in management can be organized as (i) those that focus on the explanatory value or usefulness of account of the theoretical exposition (e.g., Astley & Zammuto, 1992; Corley & Gioia, 2011; Gioia & Pitre, 1990; McKinley, Mone, & Moon, 1999; Sutton & Staw, 1995); (ii) those that focus on the Popperian concept of falsifiability and/or other conditions for rigorous construction as well as exacting empirical tests of theories (e.g., Arend, Sarooghi, & Burkemper, 2015; Dubin, 1969; Miller & Tsang, 2010; Shapira, 2011); and (iii) those that place similar importance on both explanatory value and scientific rigor/validity requirements (e.g., Acar, Franquesa, & Mwaka, 2020; Ackoff, 1962; Bacharach, 1989; Eisenhardt, 1989b; Whetten, 1989). The criteria emphasized by these differing perspectives are only partially overlapping and, most importantly, often at odds with each other (i.e., there are trade-offs in the simultaneous pursuit of divergent sets of criteria).

Empirical studies may advance the debates by shedding light on the criteria associated with lasting and influential theories in management. Besides considerations like the reputation of the authors or the journal of publication (Cole & Cole, 1967; Judge et al., 2007), which characteristics of a theoretical contribution are most strongly associated with its eventual prominence? Are scientific rigor/accuracy aspects equally appreciated by management scholars as aspects related to richness/usefulness of account, or does one type of consideration prevail over the other? Beyond the personal viewpoints of various authors (e.g., Dubin, 1969), as well as the evolving and varying guidelines offered in editor notes of leading journals (e.g., Barney, 2018; Kilduff, 2006; Whetten, 1989), what empirically validated recommendations can we offer to management researchers for effective theory building?

To date, only a few studies have begun to explore the relationships between qualities of a theoretical contribution and its subsequent scholarly impact within the management field (Colquitt & Zapata-Phelan, 2007; Miner, 1984, 2003). Despite the unquestionable value of these contributions, they suffer from two important limitations: First, in all cases, they are circumscribed to only a few (and broadly defined) theory evaluation criteria under investigation at a time, so they offer rather partial and tentative evidence. Second, these studies have tended to measure the variables of interest as *ex post* outcomes—that is, some years after the publication of the theoretical work in question—as opposed to as characteristics of the theoretical treatise itself (e.g., in Miner, 1984, 2003). As such, they do not inform theorists and their evaluators of the desirable qualities during theory-making which will drive the ensuing prominence of a theoretical work.

This paper makes several contributions to this literature. First, we synthesize the extant conceptual literature to propose a comprehensive yet parsimonious model of *ex ante* theory-building/evaluation criteria that are likely to drive a theory's interest, adherence and, ultimately, salience within the management academic community. Second, we introduce a new, overarching theory-building/evaluation construct, which we term the *explanatory meaningfulness* of a theoretical exposition, and provide confirmatory evidence in support of this higher-order latent variable. Third, we advance and validate a survey instrument that operationalizes key theory evaluation constructs of interest in the prior literature. Finally, we explore the relationship between the *ex ante* theory evaluation traits of theoretical contributions and the eventual perceived scholarly importance of such contributions, thus pointing to recommendations for effective theory building by management scholars.

The rest of the paper is organized as follows: We begin by summarizing the embryonic empirical literature on the relationships between satisfaction of theory evaluation criteria and eventual prominence of theoretical contributions among management scholars. Next, we draw from the prior conceptual literature to develop our proposed model of key theory-evaluation dimensions that drive the subsequent importance of theoretical works. Ensuing sections describe, in turn, the methods used and the results of the empirical test of our model. We conclude with a discussion of our

findings, as well as of limitations of the study and implications for future research.

PRIOR EMPIRICAL RESEARCH AND FINDINGS

Three prior studies have explored relationships between theory evaluation correlates and the subsequent scholarly prominence of theoretical contributions in management: Miner (1984) analyzed 32 management theories and explored if (i) *scientific validity* (defined as the extent to which subsequent empirical tests had been found generally supportive; as measured by the author's own ratings) and/or (ii) *practical usefulness* (defined as the extent to which useful applications to management practice had resulted; also subjectively rated by the author) were related to a theory's scholarly *importance* (as measured by frequency of nomination by a panel of knowledgeable scholars). Surprisingly, neither scientific validity nor usefulness in practice appeared to be related to the scholarly consensus regarding a theory's importance.

In a re-test of the same relationships 20 years later, Miner (2003) broadened his study to a "reasonably complete listing" of 73 organizational behavior theories. A panel of knowledgeable scholars provided importance ratings for each of them, while the independent variables were measured in a similar fashion as in Miner (1984). This time, a significant correlation was found between the *scientific validity* of a theory and its *importance* rating, while *usefulness in practice* remained not significantly related to *importance*. In light of this change in results, Miner concluded that the field of organizational behavior had evolved into a "more mature science," albeit one that might "have become too academic" at the expense of disregarding "the matter of practical application" (2003, p. 262).

Subsequently, Colquitt and Zapata-Phelan (2007) investigated the "theoretical contribution" made by empirical articles in management by coding all papers published in the *Academy of Management Journal* between 1963 and 2007.[1] The theoretical contribution of an empirical article was defined and coded as its levels of (i) *theory building* (i.e., the extent to which it adds to existing theory by proposing a new mediator or moderator, a new relationship, or even a new construct) and (ii) *theory testing* (i.e., the extent to which predictions/hypotheses are grounded on something closest to a "true theory"). The authors' theory building and testing ratings were both found to be positively associated with an article's scholarly impact, as measured by citation rates.

Colquitt and Zapata-Phelan also found that the types of empirical articles that received the most citations were those with substantial *and* balanced theory building and testing ratings, where articles in the *mid-range of both* variables achieved as much prominence as those with *high scores in both*. From this, and based on the conceptual theory-evaluation literature (e.g., McKinley et al., 1999), they concluded that a *balance* between *novelty* on the one hand and *continuity* with an existing theoretical formulation on the other helps bring the most attention to an article among management scholars (2007, p. 1293), regardless of whether such extant formulation was a vague framework, a model, or a "true theory."

To further explore predictors of scholarly importance, our study builds upon the arguments and findings from the above investigations, as well as other conceptual theory evaluation work, albeit using a novel empirical approach: First, while the data used in the above studies were single-rater assessments of a few traits on a broad swath of theories (Miner, 1984, 2003) or empirical articles (Colquitt & Zapata-Phelan, 2007), our dataset is made of independent assessments by a large sample of knowledgeable scholars of a larger set of traits from a few seminal (and hence memorable) theoretical exemplars which serve as vehicles. Our data were collected from hundreds of members of pertinent divisions of the Academy of Management, who were asked to complete a detailed assessment of just one or two theoretical papers with which they professed to be conversant. Second, while measurement of explanatory variables in prior studies relied on single-item classifications derived from complex coding schemes, we advance and validate survey-based measurement scales for the

key constructs of interest. Our study focuses on uncovering the underlying relationships among key traits of a theoretical treatise, as well as between the latter and subsequent scholarly importance, as opposed to seeking to develop comparative ratings of different theoretical expositions. The next section develops our proposed conceptual model.

PREDICTORS OF THEORY IMPORTANCE

It has been argued that the ability of a theoretical contribution to explain important organizational phenomena is one of the determinants of its attractiveness, and ultimately salience, among management scholars (Astley & Zammuto, 1992; McKinley et al., 1999). Based on our synthesis of the theory building and evaluation literature in management (including commentary in the editorial statements of leading journals), we argue that there are three distinct and fundamental indicators of the perceived explanatory value of a theory to management scholars: (1) its *novelty* (e.g., Colquitt & Zapata-Phelan, 2007; Corley & Gioia, 2011; McKinley et al., 1999), (2) its *extendibility* or potential to generate further research (e.g., McKinley et al., 1999; Whetten, 1989), and (3) its *relevance to practice* (e.g., Corley & Gioia, 2011; Miner, 1984, 2003). Moreover, we contribute that, although distinct, the above traits interplay with one another in complex ways, so that the overall perceived value of a theory's insights will be best captured as a unifying dimension. Thus, we will propose below a new construct which we term the *explanatory meaningfulness* of a theory (or of a model, framework, or other theoretical contribution) and which captures its overall *a priori* informational value.

Novelty

Building on prior work, we define the *novelty* of a theory or theoretical exposition as the extent to which it challenges scholars' extant thinking on organizational phenomena by either modifying or extending current theories, or by offering an entirely new point of view (Barney, 2018; Bettis et al., 2014; Conlon, 2002; Davis, 1971; McKinley et al., 1999; Whetten, 1989). Thus, our construct encompasses two possible subdimensions of originality of a theoretical contribution, which Corley and Gioia (2011) respectively termed "incremental" (i.e., progressive advances in an established course of understanding) and "revelatory" (i.e., surprising new perspectives or new directions in understanding).

Novelty may create *defamiliarization*, which DiMaggio (1995, p. 392) defined as the process by which an academic discipline is led to see the world "with new eyes," and which he argued should be one of the aims of "good theory." Yet, it is important that novelty (even when it is of the revelatory kind) not be overdone and that there be some link or continuity with assumptions, rationales, constructs, and language already familiar to scholars so that the theory can be understood and seen as legitimate (Davis, 1971; DiMaggio, 1995; McKinley at al., 1999). Thus, a certain balance between defamiliarization and familiarity becomes an important consideration when introducing novel theoretical aspects.

It should also be noted that, per the above definition, our construct of novelty excludes theoretical invention that emanates from a lack of appreciation of constructs and rationales already available in the prior literature. This *faux* novelty contributes to a proliferation of redundant theoretical elements in our field, which muddies the pursuit of understanding and undermines the development of a coherent body of knowledge (e.g., Oxley, Rivkin, & Ryall, 2010).

Novelty, as defined here, could entail exploring previously unexamined relationships among existing constructs, proposing new causal mechanisms for already known relationships, or modifying relationships within extant theories by adding new mediators or moderators. Novelty could also go further by exploring entirely new research questions involving novel phenomena and/or introducing brand-new constructs and justifying relationships among them (Bartunek et al., 2006; Colquitt &

Zapatta-Phelan, 2007; Davis, 1971; McKinley at al., 1999).

We expect novelty to be a driver of a theory's attractiveness and thus of its eventual prominence among scholars. By separating its insights from the multitude of frameworks, constructs, and variables that compete for a scholar's attention, novelty helps the theory to get noticed (Colquitt & Zapatta-Phelan, 2007; DiMaggio, 1995; McKinley at al., 1999). Novelty also raises the interest among scholars by creating opportunities for follow-up scholarship, which promotes adoption (McKinley at al., 1999). Consistent with this, a survey of editorial board members of the *Academy of Management Journal* revealed that the most frequently cited reasons why an article was found interesting included, among others, whether it created new theory or whether it challenged established theory and created an "aha" moment (Bartunek et al., 2006). As a result of its effects on salience and interest, the novelty of a theory will contribute to the eventual impact of the ideas it presents (Whetten, 1989).

Extendibility

The *extendibility* of a theory refers to the degree to which it provides opportunities for, and stimulates, further theoretical development and/or broadened applicability (Acar et al., 2020). This includes opportunities to clarify or redefine its constructs, causal mechanisms, and/or boundary conditions, as well as opportunities to apply the theory to different phenomena or levels of analysis (Makadok, Burton, & Barney, 2018). A good theoretical contribution should offer an inspiring framework that triggers subsequent development and energizes the exploration of new questions in ways that foster the growth of knowledge (Bartunek et al., 2006; McKinley et al., 1999; Whetten, 1989). Thus, albeit different constructs, extendibility and novelty are mutually supportive: Novelty (especially of the *revelatory* kind) makes theories more extendible, and extendibility breeds further novelty.

Besides novelty, an important contributor to extendibility is a broad ideational *scope*, defined as the range of phenomena to which the theory could be applied (Bacharach, 1989; McKinley et al., 1999), as well as the possibility that it can be applied to different levels of analysis (Makadok et al., 2018). As scope increases generalizability, it adds to theoretical relevance. It also increases the variety of operationalizations and measures that may fit within the theory's frame of reference and, hence, the number of empirical studies that could be presented as tests of the theory (Astley & Zammuto, 1992; McKinley et al., 1999). Scope is fostered by the level of abstraction of a theory's formulation, as well as by the adequate broadness of construct definitions and of the boundary conditions under which the theory applies (Bacharach, 1989; McKinley et al., 1999; Suddaby, 2010). For example, Makadok et al. (2018, p. 1539) noted that when a theory's assumptions or boundary conditions are very general, opportunities exist to extend the theory to different sub-cases (i.e., by narrowing conditions), in order to derive more specific implications.[2] Similarly, Bacharach (1989, p. 507) remarked that constructs and variables with broader scope increase the "explanatory power" of the theory.

A related, albeit less desirable, contributor of extendibility is the *ambiguity* of construct definitions, causal mechanisms, or boundary conditions (Astley & Zammuto, 1992; McKinley et al., 1999; Weik, 1995). We define a construct's ambiguity as the extent to which the interpretation of its definition varies from scholar to scholar (Oxley et al., 2010). Likewise, ambiguity of the causal mechanism refers to vagueness with regard to the underlying logic that drives the theory's propositions. And ambiguity of boundary conditions refers to vagueness with regard to the contexts and levels of analysis where the theory applies. Such ambiguities may be unavoidable in the initial stages of a new theoretical stream, but they are also recognized as pervasive problems that plague management theory-making, in particular due to our over-reliance on informal (Makadok et al., 2018) or verbal (Oxley et al., 2010) modes of theorizing.[3] Although unhelpful to the accuracy of knowledge claims, ambiguity provides opportunities for refinements of the theory and,

thus, extendibility (McKinley et al., 1999; Suddaby, 2010). For example, Makadok et al. (2018, p. 1537) observed that ambiguity about a theory's boundary conditions offers an opportunity for subsequent research to debate and clarify its hidden assumptions. Similarly, other ambiguities may provide opportunities to clarify a theory's constructs or underlying rationale, or to expose internal inconsistencies in the latter (Makadok et al., 2018).

An additional contributor of extendibility is the *richness* of the theory, defined as the degree to which it attempts to encompass different theoretical perspectives and/or simultaneous organizational demands, even when inconsistent with one another, in order to fully account for multifaceted organizational realities (DiMaggio, 1995; Gioia & Pitre, 1990; Lado et al., 2006; Poole & Van de Ven, 1989).[4]

Extendibility increases the utilitarian value of a theory to scholars by facilitating and promoting a stream of theoretical and empirical research work, and publication, based on it. Indeed, we believe extendibility to be the key driver of the perceived research usefulness of a theory. Thus, we expect extendibility to increase the scholarly appeal of a theory, leading to greater use, and ultimately to heightened impact on the field in the form of a substantial cumulative literature (Astley & Zammuto, 1992; Bacharach, 1989; Makadok et al., 2018; McKinley et al., 1999).

Relevance to Practice

As a professional field, there is a long tradition that organization and management theory should have implications for, and effects on, practice (Corley & Gioia, 2011; Hambrick, 1994, 2007; Miner, 1984, 2003; Mintzberg, 2005; Sandberg & Tsoukas, 2011). Generating knowledge that is usable "in the real world" and addressing subjects that are timely and relevant to practitioners are seen as important factors that contribute to the quality of our theorizing (e.g., Kilduff, 2006; Ployhart & Bartunek, 2019; Whetten, 1989). As such, it is often a requirement for publication in management journals that a theoretical contribution should be "relevant to practice" (e.g., Rynes, 2005).

Drawing on Astley and Zammuto (1992), we define the *a priori relevance to practice* of a theoretical contribution as its potential to influence managerial (or other practitioners') action through either (i) feasible tools and techniques that may derive from the theory and will be directly applicable in practice or, more generally, (ii) concepts, ideas, and language that may increase practitioners' ability to frame, analyze, and solve business problems, as well as to legitimate chosen courses of action. Relevance to practice is also enhanced by (iii) the timeliness and importance to management practice of the phenomena that the theory explains (Corley & Gioia, 2011; Mintzberg, 2005; Ployhart & Bartunek, 2019).

Thus, our construct is broader than Miner's "usefulness in practice" (1984, 2003), which focused only on practical tools or "applications" generated form the theory. Consistent with prior observations that organization science theories tend to be rather abstract and rarely reducible to specific managerial tools (e.g., Beyer & Trice, 1982), Miner (2003, p. 267) reported that most of the 73 theories in his study rated on the low end of his usefulness-in-practice scale. Indeed, the bulk of the contributions to practice by management theories may rather occur through the provision of conceptual language that enhances the problem-framing and problem-solving capabilities of managers and consultants (Astley & Zammuto, 1992).[5]

Based on our review of the theory building/evaluation literature, we expect that the relevance to practice of a theoretical contribution will increase its appeal among scholars, thus leading to its greater adoption and eventual impact on the field (Bartunek et al., 2006; Miner, 1984; 2003). Consistent with this expectation, for example, Corley and Gioia (2011, p. 19) reported that an observable pattern distinguishing the papers that received the *Academy of Management Review* Best Article Award in any given year from the papers that ended up being the most cited over time (out of the same AMR volumes) is that the most cited papers could also "be characterized as higher in utility for practice."

Overall Explanatory Value: The Concept of Explanatory Meaningfulness of a Theory/ Model

Although distinct traits, we expect that the *novelty*, *extendibility*, and *relevance to practice* of a theoretical contribution are connected elements that mutually reinforce each other. Given this, we propose that they will be best conceptualized as different manifestations of a single, tridimensional higher-order construct. For example, *ceteris paribus*, greater novelty enhances the opportunities for subsequent scholarship on new and unexplored questions and, thus, extendibility. Simultaneously, as noted above, extendibility breeds further novelty. Similarly, practical relevance enhances extendibility, and vice-versa: greater practical relevance facilitates direct observation of the relevant phenomena as well as of manifestations of the theoretical knowledge "in use," which may provide new ideas/insights to modify/extend the theory. At the same time, greater extendibility magnifies the practical relevance of the theory as it broadens its applicability to a greater number of phenomena, contexts, or problems of consequence to practitioners.

In short, we expect complex interdependence among the dimensions associated with the explanatory significance and usefulness of a theoretical contribution, so that their discrete impact on the attractiveness of the contribution to scholars will be difficult to untangle. Consequently, we argue that their impact will be best captured through a comprehensive, unified dimension. We thus propose the notion of *explanatory meaningfulness* of a theory as the overarching explanatory value perceived by management scholars based on a theory's combination of novelty, extendibility, and relevance to practice. Moreover, we propose that this overarching construct can be best operationalized as a second-order factor with novelty, extendibility, and relevance as its three reflective lower-order indicators. Formally:

Hypothesis 1: *Perceived novelty, extendibility, and relevance to practice will be distinct yet interrelated constructs whose covariance will be effectively captured by a reflectively measured second-order factor.*

Also, consistent with the arguments above for each of its component dimensions, we expect this overarching second-order *meaningfulness* construct to be positively related to the eventual prominence of a theory among management scholars. Namely:

Hypothesis 2: *This tridimensional second-order construct, which we will refer to as the* a priori *explanatory meaningfulness of a theoretical treatise, will be positively related to the treatise's eventual perceived importance among management scholars.*

Scientific Rigor

Besides the meaningfulness of its substance and insights, the scientific rigor/validity of a theory/ model is a conventional assessment criterion emphasized by prior theory-building authors (Arend et al., 2015; Bacharach, 1989; Dubin, 1969; Eisenhardt, 1989b; Shapira, 2011) as well as by editors of leading publications (e.g., Bartunek et al., 2006; Whetten, 1989), and which has been related to the prominence of theoretical works in management (Colquitt and Zapata-Phelan, 2007; Miner, 2003). Drawing from philosophy of science precepts, two rigor/validity aspects in particular have been emphasized in the prior literature as necessary requirements for sound theoretical construction: (1) *logical consistency* (Ackoff, 1962; Campbell, 1953; Oxley et al., 2010; Shapira, 2011) and (2) *empirical falsifiability* (Bacharach, 1989; Popper, 1959; Shapira, 2011).

We define *logical consistency* as the extent to which the theoretical contribution is a well-constructed deductive system resulting in *coherent* explanation and concluded insights. Namely, starting with clear construct definitions as well as explicit statement of assumptions or axioms and, based on these, solid logical derivations follow which lead to specific predictions (Ackoff, 1962; Campbell, 1953; Dubin, 1969; Haveman et al., 2019; Shapira, 2011). Logical consistency, therefore, implies that there are no tautologies or near-tautologies in the arguments (Arend et al., 2015; Bacharach, 1989; Priem &

Butler, 2001) and that the logic is otherwise reliable and coherent (Ackoff, 1962; Campbell, 1953; Dubin, 1969; Shapira, 2011; Whetten, 1989). Since logical consistency has been argued to be an important aspect of "good" theory, we expect this trait to increase the attractiveness of a theory to management scholars and, thus, their allegiance to it, leading to eventual impact. Hence:

Hypothesis 3: *The logical consistency of a theoretical treatise will be positively related to its eventual perceived importance among management scholars.*

We define the *empirical falsifiability* of a theoretical contribution as the extent to which predictions formulated from it can be refuted by empirical evidence (Arend et al., 2015; Bacharach, 1989; Miller & Tsang, 2010; Shapira, 2011).[6] Empirical falsifiability is related to the ability to operationalize and measure constructs accurately and reliably; to the precision of predictions regarding the nature of the relationships among them as well as the specifications of boundary conditions under which such relationships should hold; and to the ability to obtain pertinent data, including enough variation in the object of analysis (Astley & Zammuto, 1992; Bacharach, 1989). Falsifiability therefore facilitates empirical tests of the theory, as well as the derivation of conclusions from such tests, thus contributing to cumulative learning.

It is important to note that, in contrast to the hard sciences, falsifiability of organization and management theories is obstructed by (i) the complex, open, and changing nature of organizational phenomena; (ii) the inherent vagueness of our theories, which rarely can be specified in mathematical terms; and (iii) the pervasive lack of clarity regarding assumptions and boundary conditions (Acar et al., 2020; Bacharach, 1989; Miller & Tsang, 2010). Therefore, strict falsifiability leading to definitive confirmation/refutation is an ill-fitting notion and thus an unreasonable demand on management theories, as well as on theories in the social sciences in general (Acar et al., 2020). Rather, falsificationism in the social sciences needs to be understood as a more modest pursuit, where verification and falsification claims should be regarded as tentative or possibly "fallible" (Miller & Tsang, 2010). Still, seeking to plausibly affirm or reject theories on the basis of rigorous empirical tests is essential for the production of knowledge in organization science (Miller & Tsang, 2010; Oxley et al., 2010).

Given the emphasis in the theory-making/theory-evaluation literature upon practicable testability leading to a sense of empirical validity of management theories (Bacharach, 1989; Eisenhardt, 1989b; Miller & Tsang, 2010; Oxley et al., 2010; Shapira 201), we expect empirical falsifiability to be

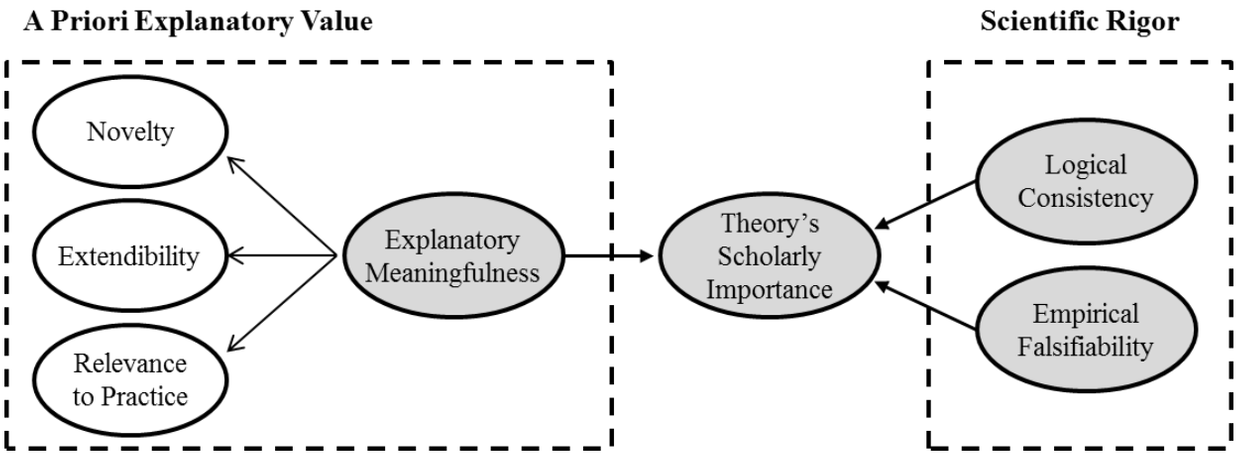

Figure 1: Hypothesized Model of Theory Importance

one of the determinants of a theory's attractiveness to scholars, contributing to its use (McKinley et al., 1999) and, thus, to its eventual prominence in the field (Miner, 2003). Therefore:

Hypothesis 4: *The empirical falsifiability of a theoretical treatise will be positively related to its eventual perceived importance among management scholars.*

Our hypothesized model is depicted in Figure 1. We now turn to the description of our empirical test of this model.

METHODS

Participants, Procedures, and Study Sample

Participants in our study were academic members of the Business Policy and Strategy (BPS), Organization and Management Theory (OMT), and Entrepreneurship (ENT) divisions of the Academy of Management who were affiliated with institutions within the United States.[7] Potential respondents were contacted by email and invited to participate in an online survey that entailed a detailed assessment of up to two theoretical exemplars, chosen from a short list of eight well-known works in strategy and organization theory. To promote response, the articles offered for assessment were seminal pieces within four of the most prominent theoretical streams within strategic management and organization theory. At the same time, to provide variance in the variables of interest, they represented different theoretical approaches and/or different levels of influence within their respective literatures. The works offered as assessment choices were Jensen and Meckling (1976) and Eisenhardt (1989a) for agency theory, DiMaggio and Powell (1983) and Greenwood and Hinings (1996) for institutional theory, Barney (1991) and Amit and Schoemaker (1993) for the resource-based view, and Williamson (1975) and Ouchi (1980) for transaction-cost economics.

At the start of the survey, participants were asked to select the theoretical work that they were most knowledgeable about to assess.[8] After completing the assessment instrument for the chosen theoretical exemplar, respondents could opt to either end the survey or select a second article form the list and complete a second assessment. To ensure that respondents only rated the theoretical exemplars they were most familiar with, the survey closed after the second assessment round.

Of 2,918 scholars invited to participate, 641 entered and completed the survey, for a 22% response rate. Among those that saw the survey to its end, 187 respondents reported insufficient familiarity to provide a detailed assessment of any of the listed theoretical exemplars, 153 respondents completed the assessment instrument for only one exemplar, and 301 assessed two exemplars, producing a total of 755 theory assessment observations. Of these, 137 assessments included an "I don't know" response to one (or more) evaluation item(s) and were dropped from the present study[9] in order to limit the sample to entirely knowledgeable raters. To further screen for expert assessments, we also dropped an additional 106 observations where the respondent had rated his/her familiarity with the subject article below 5 in a 7-point Likert scale. These screens resulted in a study sample of 512 observations.

Among the participants comprising the final sample, 38% were members of the BPS division only, 19% were OMT members only, 9% were ENT members only, 32% were members of two of these divisions, and 1% reported membership in all three. Over 97% held doctorates and had had this degree for an average of 13 years. After screening for this, the average reported familiarity with the theories evaluated was 6.04 out of 7, and 94% of sample respondents reported using the assessed article at least once as a reference in their own work.

Measurement

Scholarly importance. Following Miner (2003), our dependent variable was measured using a single-item indicator of the management scholars' perceived importance of the theoretical exemplar. Respondents were asked to "rate the importance of the theory to the field of management," and their responses were recorded on a 7-point scale ranging

from "not at all important" to "very important." The validity of this measure was explored by contrasting it with citation data from the EBSCO *Business Source Complete* database: The correlation of average perceived importance with total number of citations was .75, and that with average annual citations was .73, which supports the convergent validity of our measure.

Theory evaluation constructs. Since *novelty*, *extendibility*, *relevance to practice*, *logical consistency*, and *empirical falsifiability* had not been operationalized via survey instruments before, we developed a theory evaluation scale to measure the independent variables in our proposed model. Design of this instrument was based on our review of the theory building/evaluation literature in management and was guided by the conceptual definitions offered earlier in this paper. It also underwent two separate phases of refinement before being used to test the hypotheses in this study.

In the first phase, different versions of our proposed theory evaluation scale underwent sequential review by two separate panels of judges: A first draft of the subscales for each theory evaluation construct was subjected to scrutiny by three professors from sociology and management departments who were knowledgeable about the relevant literature, and who reviewed for content validity and clarity. Items were dropped, added, or amended based on the feedback received, resulting in a preliminary 31-item overall theory evaluation scale. The latter, along with other items capturing variables to be used as screens or controls in the study, was then reviewed by a larger panel of faculty and doctoral students, who focused on (i) the clarity and interest of the questions, (ii) appropriateness in terms of the time and intellectual demands on respondents, and (iii) sequencing of the different parts of the survey instrument. Besides adjustments to other survey items, the theory evaluation scale underwent further editing, and one of the preliminary items was altogether dropped.

The second phase of scrutiny consisted of validation and refinement of the factor structure of the theory evaluation scale through exploratory factor analysis (EFA). Scores on the 30-item battery from the study sample were factor analyzed using principal components extraction and oblique (*promax*) rotation. Five factors were initially retained based on the Kaiser-Guttman rule (Kaiser, 1960) and the scree test (Cattell, 1966). Following rotation, five items were eliminated which either had cross-loadings or failed to load onto any factor. A subsequent factor analysis on the remaining 25-item instrument, using the same methods and retention criteria, resulted in five factors that accounted for 64.6% of item variance. The extracted five-factor structure was consistent with our proposed constructs, and there were no items with cross-loadings above .35 on the rotated factors. Loadings, item communalities, percent variance explained by each factor, and Cronbach's α reliability estimates of the final subscales are presented in Table 1.

The resulting measures from the validated 25-item theory evaluation instrument are as follows: The *novelty* of the insights proposed by a theoretical exemplar was measured using a six-item scale (α = .89). Sample items are: "the theory offered different explanations for previously established relationships" and "the theory created an 'aha' moment when proposed." The *extendibility* of the constructs and rationale presented in the theoretical exemplar was operationalized via a four-item scale (α = .73). Examples of items are: "the theory as presented could be extended to a broad range of phenomena" and "the theory's rationale could be extended to shed light on new questions." *Relevance to practice* was measured using an eight-item scale (α = .91). Items include: "understanding the theory helps practitioners formulate action plans" and "the theory offers a good tool for business consultants." The *logical consistency* of the theoretical exposition was captured with a three-item scale (α = .63), including "there were inherent contradictions in the theory" and "the propositions derived from the theory were tautological," both reverse-scored. Finally, the e*mpirical falsifiability* of predictions derived from the theory was measured using a four-item scale (α = .85) that includes items such as "predictions from the theory can be contrasted with empirical evidence" and "the theory's constructs

Table 1: Exploratory Factor Analysis of (Refined) Theory Evaluation Scale[a]

			Factor pattern (standardized regression coefficients)					
Item	**Mean**	**s.d.**	**Relevance to practice**	**Novelty**	**Falsifiability**	**Extendibility**	**Logical consistency**	h^2
P1	5.26	1.68	**0.88**	0.01	0.04	0.03	-0.12	0.79
P2	5.34	1.65	**0.84**	0.08	0.02	-0.02	-0.06	0.74
P3	5.26	1.58	**0.74**	-0.12	0.02	-0.13	0.29	0.61
P4	5.52	1.47	**0.76**	0.11	0.09	0.00	0.02	0.73
P5	5.22	1.66	**0.90**	-0.10	-0.05	0.01	-0.07	0.71
P6	6.31	1.01	**0.42**	0.14	-0.04	0.31	0.12	0.53
P7	5.37	1.58	**0.67**	-0.04	-0.08	0.02	0.21	0.51
P8	5.46	1.59	**0.66**	0.13	0.15	0.03	-0.07	0.63
N2	5.74	1.42	-0.05	**0.82**	0.10	-0.16	-0.03	0.61
N3	5.92	1.29	-0.01	**0.82**	0.10	-0.07	-0.05	0.66
N4	5.97	1.31	-0.03	**0.84**	0.03	-0.09	-0.06	0.61
N5	6.18	1.21	0.02	**0.79**	-0.04	0.10	0.02	0.70
N6	6.28	1.07	0.10	**0.82**	-0.09	0.06	-0.01	0.73
N7	5.71	1.46	0.01	**0.70**	-0.04	0.10	0.12	0.62
F2	5.41	1.57	0.00	0.08	**0.71**	0.08	0.21	0.77
F3	5.52	1.49	-0.01	0.07	**0.69**	0.09	0.23	0.76
F4	4.62	1.79	0.03	-0.04	**0.89**	0.03	-0.04	0.77
F5	5.46	1.54	0.09	0.01	**0.74**	-0.02	-0.10	0.55
E1	5.93	1.31	0.03	0.29	0.03	**0.61**	-0.04	0.62
E2	5.02	1.83	0.11	-0.17	0.13	**0.70**	-0.19	0.50
E3	5.81	1.43	-0.10	-0.15	0.09	**0.86**	0.01	0.65
E5	6.21	1.12	-0.02	0.13	-0.07	**0.68**	0.04	0.53
LC1	5.54	1.72	0.02	0.11	-0.18	0.34	**0.54**	0.56
LC2	4.65	1.75	0.17	-0.10	0.01	-0.13	**0.76**	0.57
LC3	4.72	1.89	-0.12	0.06	0.20	-0.06	**0.76**	0.69
Total communality								16.15
% Variance explained			19.10	17.01	10.76	10.21	7.50	64.58
Alpha coefficient			0.91	0.89	0.85	0.73	0.63	

[a] n = 512. Promax-rotated factor solution reported. Factor loadings (i.e., coefficients ≥.35) are indicated in boldface. Interfactor correlations are in the range from 0.26 to 0.43. h^2 = item communalities. Percent variance explained statistics are reported from the pre-rotation varimax solution, since, once factors are allowed to covary with the additional procrustean transformation, variance explained overlaps and communality cannot be uniquely partitioned.

are easily translated into measurable variables." All measures used a Likert response scale, ranging from 1 = "strongly disagree" to 7 = "strongly agree." The complete text of the theory evaluation instrument is reproduced in the Appendix.

Control variables. To minimize potential omitted variable problems when testing for hypothesized relationships to perceived theory importance, we controlled for theory age, as well as for possible rater biases due to years of experience, closeness to the theory, and respondent research and consulting

intensity. *Theory age* was measured as the logarithm of the number of years since publication of the evaluated article. *Rater experience* was measured as *log* (1 + number of years since respondent earned his/her doctoral degree). *Rater familiarity* was measured as the logarithm of the reported number of instances of use of the rated article as a reference in the rater's own publications or presentations. *Research intensity* and *consulting intensity* were the self-reported percentages of the rater's professional life spent in research and consulting activities, respectively.

Analyses

We used covariance-based structural equation modeling (SEM) with LISREL software. Following recommended procedures (e.g., Anderson & Gerbin, 1988), we employed a two-step analytical strategy whereby, prior to fitting the structural model, we examined and determined the best measurement model through confirmatory factor analysis (CFA). In our case, this step also served to test our first hypothesis. All latent variables were modeled using the discrete item indicators from our validated instrument (as opposed to using item averages), and models were fit using the maximum likelihood estimator, which has been found to be rather robust to distributional variations (e.g., Chou & Bentler, 1995).

We compared five models to identify the measurement structure that best fit the data. Model 1 was a first-order unitary factor model, with all items loading onto a single latent variable. Model 2 was a first-order model with the five orthogonal factors, and with the loading structure as derived from the EFA. The comparison between these two models would serve to confirm the multi-dimensionality of our theory evaluation instrument. Model 3 was the correlated (oblique) five-factor model. The comparison between Model 2 and Model 3 therefore would serve to confirm that there is significant covariance between the first-order factors which, as hypothesized, might then best be captured through a second-order factor. Model 4 was the hypothesized measurement model with one second-order factor (explanatory meaningfulness) and five first-order factors, where only three of them (novelty, extendibility, and relevance to practice) loaded on the second-order factor. Finally, Model 5 was an alternative structure with one general second-order factor with all five first-order factors loading onto it.[10] The comparison among the last three models served as the test of Hypothesis 1.

To gauge and compare model fit, five recommended measures were used: the chi-square/*df* ratio; the root mean square error of approximation (RMSEA); the non-normed fit index (NNFI), also called the Tucker-Lewis index (TLI); the comparative fit index (CFI); and the consistent Akaike information criterion (CAIC) (e.g., Bentler, 1990; Kline, 1998). This is a comprehensive set that includes absolute, incremental, and parsimony fit indices. Chi-square/*df* is a measure of absolute fit, indicating how closely the estimated covariances in a fitted model match the covariances in the original data. Normally, values of 3.0 or less in this statistic are interpreted as indicating good fit to the data (Kline, 1998), although others have proposed a cut-off of 5.0 or less as indicative of acceptable fit (Wheaton et al., 1977). RMSEA is another absolute fit index which is adjusted for sample size and thus recommended to complement the chi-square fit test in larger samples. A value smaller than .08 in this index is considered to indicate good fit (MacCallum et al., 1996). RMSEA also has the advantage of a known distribution, allowing us to also report the 90% confidence interval around point estimates. By contrast, NNFI and CFI are incremental fit indices, indicating improved overall fit relative to a null model where all variables are assumed to be uncorrelated. For both indices, values equal to or exceeding .95 are indicative of good fit (Hu & Bentler, 1999). Finally, CAIC is a parsimony fit index (i.e., it penalizes for model complexity) that also adjusts for sample size. Because it is not bounded, there is no suggested cut-off for this statistic. Rather, the model with the lower value in this index is seen as superior (Akaike, 1974).

Beyond fit index statistics, we also used chi-square difference tests between nested models to assess which of these provided a marginally better fit.

Models 2 and 3 are nested models, and Model 5 is also nested within Model 4.

Once the best measurement model had been identified, SEM was performed to estimate the fit of the overall hypothesized model to the data. Beyond overall fit, path coefficients for proposed relationships were used to test the remaining hypotheses.

RESULTS

Measurement Model

Table 2 shows the fit indices for the alternative measurement models. Model 1 exhibited a rather poor fit to the data (χ^2/df = 11.41; RMSEA = .18; NNFI = .83; CFI = .84). Model 2, with the five-factor structure as revealed by the EFA, produced substantially better fit indices, although not yet demonstrative of good fit (χ^2/df = 6.02; RMSEA = .11; NNFI = .92; CFI = .92; CAIC = 2,417 versus 5,361 for Model 1). Therefore, the comparison of fit indices between these two models confirms the multidimensionality of our theory evaluation scale.

Model 3, with the five correlated factors, demonstrated good fit to the data (χ^2/df = 3.66; RMSEA = .075; NNFI = .96; CFI = .96; CAIC dropped to 1,460). Moreover, the chi-square difference test between Models 2 and 3, $\chi^2_{diff.}(10)$ = 686, p < .001, strongly suggests that the latter is preferred. Additionally, all parameter estimates in Model 3 were significant at p < .001, and the average estimated factor correlation was quite high (0.53), although well below 0.9 in all cases[11] thus providing evidence of the discriminant validity of our constructs (e.g., Venkatraman, 1990). These results lend strong support to our theory evaluation constructs being distinct, yet covariant, dimensions. The high correlations among the first-order constructs boded well for second-order factor models.

Model 4, our hypothesized model with a three-dimensional second-order meaningfulness factor, demonstrated marginally better fit than Model 3 (χ^2/df = 3.65; RMSEA = .075; NNFI = .96; CFI = .96; CAIC improving to 1,445). Thus, Model 4 is preferred, largely on the grounds of parsimony (i.e., it provides as good a fit while involving fewer parameter estimates). Moreover, all parameter estimates in Model 4 were strongly significant (p < .001 in all cases), including the loadings of novelty, extendibility, and practical relevance onto the second-order construct, which were .77, .77, and .64 respectively.

Finally, Model 5, where the variances among all five first-order factors were explained by one general second-order factor, produced a slightly worse fit than either Model 3 or Model 4 (χ^2/df = 3.68; RMSEA = .075; NNFI = .96; CFI = .96; CAIC = 1,459). Also, a nested comparison between Model 4 and Model 5 shows that the former (with only novelty, extendibility, and practical relevance

Table 2: Fit Indices of Alternative Measurement Models[a]

Model	χ^2	*df*	χ^2/df	RMSEA (90% CI)	NNFI	CFI	CAIC
Model 1	3,138.8	275	11.41	.18	.83	.84	5,361
Model 2	1,654.7	275	6.02	.11	.92	.92	2,417
Model 3	968.7	265	3.66	.075 (.070 – .080)	.96	.96	1,460
Model 4	980.7	269	3.65	.075 (.070 – .079)	.96	.96	1,445
Model 5	994.3	270	3.68	.075 (.071 – .080)	.96	.96	1,459

[a] n = 512. Model 1 has one first-order unitary factor. Model 2 has five orthogonal first-order factors. Model 3 has five correlated (oblique) first-order factors. Model 4 has five first-order factors and one second-order factor (Explanatory Meaningfulness) with only three of the first-order factors (Novelty, Extendibility, Relevance to Practice) as its underlying dimensions. Model 5 has one general second-order factor with all five first-order factors as its underlying dimensions. 90% CI = 90% confidence interval around RMSEA point estimate.

loading on the second-order factor) is strongly preferred: $\chi^2_{diff.}(1) = 13.6$, $p < 0.001$. In other words, the hypothesized measurement structure provided better fit than alternative second-order models.

Overall, these results support the notion that novelty, extendibility, and practical relevance are distinct dimensions, but also collectively reflective of a second-order construct, which we have termed the *explanatory meaningfulness* of a theory or model. Thus, Hypothesis 1 is supported.

Structural Model

After finding support for our measurement model, we proceeded to test the complete hypothesized structural model of theory importance, as depicted in Figure 1. To provide a rigorous test of our hypotheses, we also included the control variables discussed above. The independent variables (explanatory meaningfulness, logical consistency, and empirical falsifiability) were allowed to covary, as indicated by our test of the measurement model, and as is the default in LISREL. The hypothesized model exhibited good fit to the data ($\chi^2/df = 2.11$; RMSEA = .06; NNFI = .95; CFI = .96), and it explained 86% of the variability in perceived importance. The standardized second-order loadings and path coefficient estimates in the fitted model are shown in Figure 2.

Hypothesis 2, which postulates that the explanatory meaningfulness of theoretical exemplars will be positively related to the eventual perceived importance by management scholars, was supported by a positive and strongly significant path coefficient between these variables (β = 1.42, $p < .001$). By contrast, Hypothesis 3, which predicts that the logical consistency of theoretical exemplars will also be positively related to eventual scholarly importance, was not supported, as the path coefficient was actually negative although not statistically significant (β = – .57, p = .12). Our results also fail to support Hypothesis 4, as the estimated coefficient from empirical falsifiability to scholarly importance was not statistically significant either (β = – .15, p = .35).

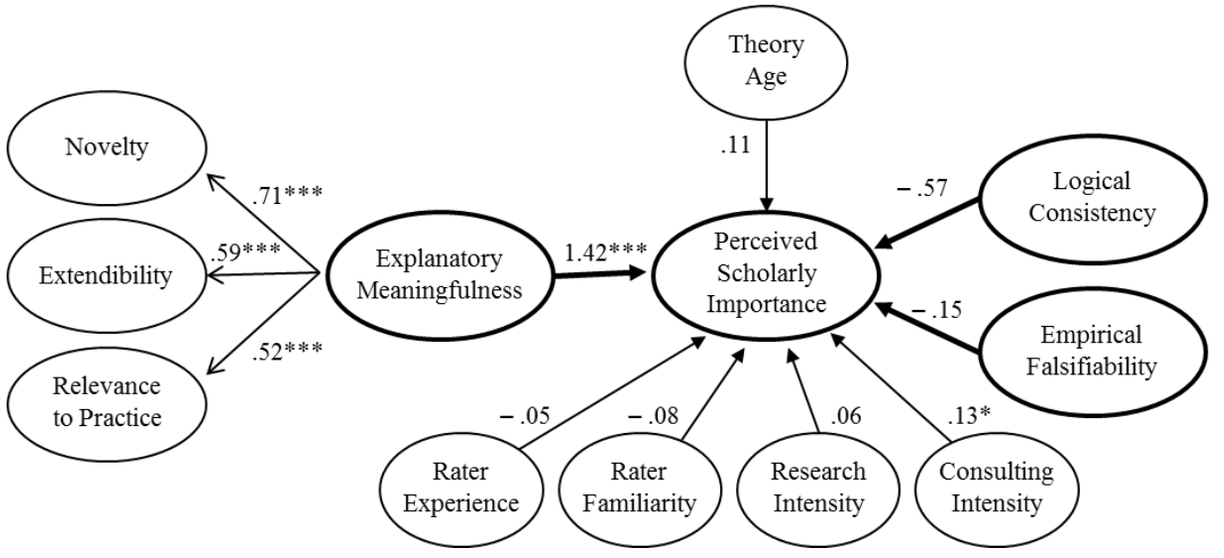

Figure 2: Structural Equation Modeling Results[a]

[a] n = 512. Standardized estimates reported. Hypothesized relationships displayed in bold.
* p ≤ .1
** p ≤ .01
*** p ≤ .001, two-tailed.

DISCUSSION

This paper synthesizes the prior literature on theory building and evaluation in management to propose a model of the key aspects of a theoretical treatise that drive its eventual scholarly importance. In our model, the *novelty*, *a priori extendibility*, and *practical relevance* of a theory's elements capture its explanatory value and lead to its adoption, allegiance, and eventual salience among scholars. Moreover, we propose that these aspects are interconnected and mutually reinforcing, so that they are best understood as subdimensions of an overarching construct, which we have termed the *explanatory meaningfulness* of the theoretical work. Therefore, we hypothesize that novelty, extendibility, and practical relevance are the reflective indicators of a second-order latent variable, and that this higher-order variable will be positively related to the subsequent perceived importance of theoretical exemplars among management scholars.

Alongside explanatory value, the scientific rigor of the exposition is also seen in most of the theory making/evaluation literature in management as either an essential requirement or, at least, a desirable trait of productive theory. Therefore, we expect that rigor will also impact the prominence of theoretical works. In particular, we hypothesize that the *logical consistency* and *empirical falsifiability* of theoretical exemplars will be positively related to their eventual perceived importance among management scholars.

Our model was tested with data from members of the OMT, BPS, and ENT divisions of the Academy of Management, who provided a detailed assessment of one or two seminal theoretical exemplars with which they purported to be very familiar. Respondents were asked to rate evaluatory traits of the particular theoretical treatise in its original published form, as well as to assess its current importance as a contribution to the field of management. EFA findings support the convergent and discriminant validity of our measurement scales across a structure composed of five distinctive but correlated dimensions, which are consistent with the proposed theory evaluation constructs of novelty, extendibility, practical relevance, logical consistency, and empirical falsifiability.

In turn, CFA findings support the existence of the hypothesized second-order theory evaluation factor, with novelty, extendibility, and practical relevance as its underlying dimensions, and thus are consistent with our proposed explanatory meaningfulness construct. Finally, SEM findings show this explanatory meaningfulness factor to be a statistically significant, and very important, predictor of the perceived current importance of a theoretical article to the field of management. By contrast, after controlling for explanatory meaningfulness, the logical consistency and empirical falsifiability properties of theoretical works are not significantly related to their perceived current importance.

Research Implications

Our findings have implications for the theory making/evaluation literature in management. In particular, they inform the ongoing debate between those that emphasize the explanatory role of theories (Astley & Zammuto, 1992; Corley & Gioia, 2011; McKinley et al., 1999; Sutton & Staw, 1995) and the importance of novel, rich, and relevant accounts of multifaceted organizational realities (Davis, 1971; Eisenhardt, 1989b; Poole & Van de Ven, 1989) and those that insist on the importance of rigorous construction and empirical refutability of theoretical frameworks (Arend et al., 2015; Miller & Tsang, 2010; Oxley et al., 2010; Shapira, 2011).

An important undercurrent in this debate stems from the existence of trade-offs between explanatory and scientific rigor aims. For example, there is a trade-off between extendibility through broadly defined boundary conditions and constructs, on the one hand, and the precision of predictions and thus falsifiability of a theory, on the other hand (Astley & Zammuto, 1992; Bacharach, 1989; Makadok et al., 2018). Also, construct vagueness in particular is deplored by those who emphasize theoretical rigor and point to its unfortunate consequences,

including muddiness of thought and infinite regress problems (e.g., Ackoff, 1962; Hallberg & Felin, 2020; Kaplan, 1964; Turner, 1989), as well as obfuscated empirical testing (Miller & Tsang, 2010; Oxley et al., 2010). Similarly, there is a trade-off between the construction of encompassing theories through the use of paradox and the pursuit of logical consistency (Poole & Van de Ven, 1989; Smith & Lewis, 2011).

There has been a paucity of empirical work contrasting the above perspectives and shedding light on some of the extant debates; hence the relevance of the present work. Results from this study suggest that the perceived explanatory significance and usefulness of a theory drives its eventual prominence within the research community in organization theory and strategic management, irrespective of its degree of scientific rigor. Management scholars appear to prize contributions that provide meaningful explanation (through greater novelty, extendibility, and practical relevance), while not much weight is being put on the greater logical consistency and falsifiability of such treatises.

As such, our findings support Davis's (1971, p. 309) classic assertion for the social sciences that "a theorist is considered great, not because his theories are true, but because they are interesting. . . . In fact, the truth of a theory has very little to do with its impact," or McKinley et al.'s (1999, p. 636) proposition that "empirical validity tends to recede into the background as a determinant of where organization theorists place their scholarly allegiances." Our findings also support those who advocate for theory-making that embraces inconsistencies, tensions, or contradictions rather than simplifying them away (e.g., Lado et al., 2006; Poole & Van de Ven, 1989; Smith & Lewis, 2011). What is sought are models that offer new language, perspective, rationale, constructs, and/or relationships, which can be used to improve scholars' ability to make sense of complex business realities and derive insights on a broad range of important organizational phenomena, as well as help practitioners identify, characterize, and interpret such phenomena in the real world.

Does this mean that scientific rigor plays no relevant role in the minds of most management scholars and can be disregarded by those seeking to make lasting and influential theoretical contributions? We do not believe the findings from the present study warrant such a radical conclusion. Unless a theoretical exposition is minimally bounded by (even implicit) assumptions, is based on reasonably well-defined constructs and a coherent rationale, and provides a frame from which to deduct fairly explicit and testable predictions, it is unlikely to meet the requirement for publication in management journals in the first place (e.g., Bartunek et al., 2006). Likewise, a measure of theoretical rigor will be necessary for subsequent tests of the theory to lead to reasonably unequivocal knowledge claims, thus generating a sense of constructive knowledge accumulation from it. This, in turn, will underpin persistent concord on the theory's usefulness and, thus, its longevity. In short, a minimum level of rigor is an obvious necessary condition for a management theory to be seen as a valid and beneficial vehicle for knowledge production, and hence for it to be adopted and used.

Rather than rigor playing no role in the subsequent adoption and prominence of management theories, we interpret the results from the present study to suggest that, beyond acceptable levels, greater logical consistency and empirical falsifiability do not contribute to greater appreciation of the theoretical work by management scholars. In other words, we speculate that scientific rigor matters for scholarly impact only up to a threshold level. Once this threshold of acceptability is met, additional formality and/or precision of argument do not breed additional impact within the management academic community, at present.

The threshold level of rigor where effects on impact are exhausted may also evolve with the state of development of a particular research literature. In the initial stages of scientific inquiry into a new domain, vague conceptual frameworks may be needed before models and, eventually, theories can be built (Oxley et al., 2010; Priem & Butler,

2001; Shapira, 2011). To the extent that this is the case, our focus on seminal works in the present study may have underplayed rigor effects: i.e., our exemplars may have required low thresholds of rigor at the dawn of their conceptual streams to be considered important theoretical contributions. In other words, despite significant variation in rigor in our data, the positive effect of rigor on impact (below the acceptability threshold) may have been censored due to a lack of variation in the stage of theoretical life cycle among the exemplars used.

Interestingly, our results, as well as the above notion of truncated rigor effects on impact, lead to an alternative interpretation of past findings in the incipient empirical literature on the subject. In particular, Colquitt and Zapata-Phelan (2007) found that (i) articles that introduced new mediators or moderators of existing relationships, and which relied on loose conceptual arguments or (at best) graphical models or diagrams (i.e., medium level of rigor), as well as (ii) articles that either introduced new relationships among existing constructs or reconceptualized constructs, and which relied on either models/diagrams or "true theory" (i.e., higher level of rigor), were the most impactful. From this, they interpreted that the articles that achieve the greatest impact in management are those that offer "a balance between novelty and continuity" with existing theoretical formulations (p. 1293). The present study suggests a plausible more parsimonious explanation for Colquitt and Zapata-Phelan's findings: Simply, that the most cited articles are those that utilize the minimum necessary theoretical rigor while maximizing the richness and flexibility of their newly introduced elements. As noted above, there are trade-offs between the precision of prediction that comes with greater theoretical rigor and the explanatory meaningfulness of the account. Therefore, the most impactful articles may be those that meet the thresholds of logical consistency and falsifiability required by the stage of development of their literature, but no more, in exchange for greater broadness and richness of their exposition.

For those who advocate for greater rigor in management theory-making (e.g., Acar et al., 2020; Oxley et al., 2010), our results provide a measure of the challenge ahead: The current incentives and norms in the field appear to run counter to their aspirations for more consistently reliable knowledge claims, as well as more productive accumulation of a coherent body of understanding of organizational phenomena. At present, logical consistency and precision of prediction appear to receive (at best) limited appreciation from management scholars.

Our research also has direct implications for theorists, as well as for editors and reviewers engaged in theory evaluation in management. The recommendation for theorists is straightforward: our results underscore the importance of advancing constructs, rationales, and hypothesized relationships that are novel, extendible, and practically relevant, as this will increase the meaningfulness of the resulting theory to management scholars, which in turn will drive it to prominence over time. Still, theory writers should avoid novelty that emanates from a lack of awareness of prior work and thus results in "reinventing the wheel" without really making a substantial contribution. Writers should also avoid extendibility that emanates from unjustified vagueness of constructs, rationale, and boundary conditions, including from unnecessarily obscure and convoluted writing style (e.g., Tourish, 2020).

Given the trade-offs between precision and meaning, as well as the limited impact effects of rigor, management theorists are also encouraged not to shy away from the embryonic exploration of important new phenomena and rationales, or to compromise the potential broadness of their ideational scope and/or the multifaceted richness of their account, in exchange for excessive exactitude of their derivation. After meeting reasonable requirements for clarity, coherence, and precision, in correspondence with the stage of conceptual development of the domain as well as the current norms in the field, increasing the formal elegance and meticulousness of the exposition is unlikely to increase the appreciation of their work.

This implies that the skill set for impactful theory writing in management today is very different from, say, in economics or finance. Impactful theory building may also be particularly challenging for budding management theorists, as they must be willing to leave behind the "safety" of an airtight formulation to, for example, accept the vagueness associated with theorizing about incipient but relevant organizational phenomena and managerial concerns.

As to editors and reviewers, we hope that our synthetic model of theory evaluation dimensions (novelty, extendibility, relevance, internal consistency, and falsifiability) may help consolidate and organize the core set of assessment criteria for theoretical treatises, alongside other essential considerations like familiarity and interconnection with the prior literature or writing quality. With regard to the weight given to different criteria, findings from our study suggest that richer and more encompassing theories that necessarily give away some rigor and coherence of argument are likely to achieve greater salience. Thus, our findings may be consistent with recent editorial calls to relax conventional deductive theory-making requirements so as to broaden the gamut of accepted approaches to theorizing (e.g., Haveman et al., 2019). Still, we believe that the apparent lack of impact effects from greater rigor in the present study should give pause to those associated with management journals. Journals play an important role in setting the incentives and quality norms for research in the discipline. Arguably, a greater appreciation for clean and tidy logical deduction and for precise prediction in management theories would contribute to more productive knowledge creation and accumulation going forward. Thus, further scientific maturity of the management field may require a proactive recalibration of the importance afforded to rigor in theory evaluation policies and norms.

Limitations and Future Directions

As in any empirical study, our research design involved trade-offs whereby advantages were obtained at the expense of some limitations. First, we used seminal works as vehicles for data collection, which, as discussed above, might have contributed to the censoring of scientific rigor effects. Subsequent investigations could further validate our model by relying on rater assessments of different sets of less foundational conceptual works. Second, our study focused on a few research streams in organization theory and strategy, which might have introduced biases in our data. Although (in further analyses not reported here) the relationship between explanatory meaningfulness and importance was found to be stable across each of our four theoretical stream subsamples, it would be important to investigate if findings generalize across other theoretical streams and management disciplines. In particular, our focus on macro-level theories might be complemented with similar studies using micro-theories and respondents from associated disciplines (i.e., in human resource management and organizational behavior). Likewise, future research could explore how drivers of importance may differ across disciplines with disparate theorizing traditions (e.g., organization science versus operations research and management science). Third, as this is the first attempt at testing our parsimonious model, our study was confined to American researchers in order to facilitate follow-up during data collection, as well as to control for likely confounding issues arising from differing cultural contexts and research traditions. Nevertheless, it would be important to extend tests of the model to other cultural settings. It is thus hoped that this study will generate a stream of systematic tailored investigations, as well as cross-national contrasts.

The discovery in this study of the higher-order construct of *explanatory meaningfulness* also opens up promising avenues for further research. We provide preliminary evidence of its nomological validity in the form of its relationship to perceived scholarly importance, but other aspects of its extended nomological network could be further explored. For example, is explanatory meaningfulness also related to the application and influence of a theory among practitioners? Likewise, the relationship between various aspects of scientific

rigor in theorizing and the scholarly prominence of theoretical treatises in management deserves further inquiry. Is there no relationship, a truncated relationship (as speculated here), or even a quadratic relationship whereby higher levels of rigor actually detract from subsequent impact?

CONCLUSION

This article makes several contributions to the sparse empirical literature on theory evaluation dimensions, and on how the latter relate to the eventual prominence of theoretical works in the organization and management field. Most importantly, we propose and test a comprehensive but parsimonious model of attributes that drive the scholarly importance of a theoretical treatise. We also introduce the *explanatory meaningfulness* construct (a tripartite second-order theory evaluation dimension reflected in a theory's novelty, extendibility, and practical relevance) and provide empirical evidence of its manifestation, as well as of its being a strong and positive predictor of long-term scholarly importance.

Our findings point to clear recommendations for theory building and evaluation in management. Specifically, given current norms in the field, our study suggests that theory writers might want to err on the side of lesser formality and precision in their formulation in exchange for greater originality, relevance, and multifaceted richness (i.e., meaning) of their account. The model and results presented here pave the way for subsequent conceptual and empirical research on drivers of scholarly impact of theoretical contributions in the field of management. A fuller understanding of the etiology of management theories' impact should be an important and, hence, welcome pursuit.

REFERENCES

Acar, W., Franquesa, J., & Mwaka, J. O. (2020). Conceiving perfectible theories in management through adaptive framing. *International Journal of Strategic Decision Sciences*, *11*(1), 1–20.

Ackoff, R. L. (1962). *Scientific method*. New York: Wiley.

Akaike, H. (1974). A new look at the statistical model identification. *IEE Transactions on Automatic Control*, *19*(6), 716–23.

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

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two step approach. *Psychological Bulletin*, *103*, 411–423.

Arend, R. J., Sarooghi, H., & Burkemper, A. (2015). Effectuation as ineffectual? Applying the 3E theory-assessment framework to a proposed new theory of entrepreneurship. *Academy of Management Review*, *40*(4), 630–651.

Astley, W. G., & Zammuto, R. F. (1992). Organization science, managers, and language games. *Organization Science*, *3*, 443–460.

Bacharach, S. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, *14*(4), 496–515.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120.

Barney, J. B. (2018). Editor's comments: Theory contributions and the AMR review process. *Academy of Management Review*, *43*(1), 1–4.

Bartunek, J. M., Rynes, S. L., & Ireland, D. (2006). What makes management research interesting, and why does it matter? *Academy of Management Journal*, *49*(1), 9–15.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bettis, R. A., Gambardella A., Helfat, C., & Mitchell, W. (2014). Editorial: Theory in strategic management. *Strategic Management Journal*, *36*, 1411–1413.

Bernaards, C. A., & Sijtsma, K. S. (2000). Influence of imputation and EM methods of factor analysis when item nonresponse in questionnaire data is nonignorable. *Multivariate Behavioral Research, 35*, 321–364.

Beyer, J. M., & Trice, H. M. (1982). The utilization process: A conceptual framework and synthesis of empirical findings. *Administrative Science Quarterly*, *27*, 591–622.

Campbell, N. R. (1953). *What is science?* New York: Dover.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.

Chou, C. P., & Bentler, P. M. (1995). Estimation and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37–55). Thousand Oaks, CA: Sage.

Cole, S., & Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review*, *32*, 377–390.

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. *Academy of Management Journal*, *50*(6), 1281–1303.

Conlon, E. (2002). Editor's comments. *Academy of Management Review*, *27*(4), 489–492.

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: What constitutes a theoretical contribution? *Academy of Management Review*, *36*(1), 12–32.

Davis, M. S. (1971). That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences*, *1*, 309–344.

DiMaggio, P. (1995). Comments on "What theory is not." *Administrative Science Quarterly*, *40*, 391–397.

DiMaggio, P., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160.

Dubin, R. (1969). *Theory building*. New York: Free Press.

Eisenhardt, K. M. (1989a). Agency theory: An assessment and review. *Academy of Management Review*, *14*(1), 57–74.

Eisenhardt, K. M. (1989b). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550.

Gioia, D. A., & Pitre, E. (1990). Multiparadigm perspectives on theory building. *Academy of Management Review*, *15*(4), 584–602.

Greenwood, R., & Hinnings, C. R. (1996). Understanding radical organizational change: Bringing together the old and the new institutionalism. *Academy of Management Review*, *21*(4), 1022–1054.

Hallberg, N. L., & Felin, T. (2020). Untangling infinite regress and the origins of capability. *Journal of Management Inquiry*, *29*(1), 17–32.

Hambrick, D. C. (1994). 1993 Presidential address: What if the Academy actually mattered? *Academy of Management Review*, *19*(1), 11–16.

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, *50*(6), 1346–1352.

Haveman, H. A., Mahoney, J. T., & Mannix, E. (2019). Editors' comments: The role of theory in management research. *Academy of Management Review*, *44*(2), 241–243.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

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

Judge, T. A., Cable, D. M., Colbert, A. E., & Rynes, S. L. (2007). What causes a management article to be cited: Article, author, or journal? *Academy of Management Journal*, *50*(3), 491–506.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20,* 141–151.

Kaplan, A. (1964). *Conduct of inquiry: Methodology for behavioral science.* Scranton, PA: Chandler Publishing.

Kilduff, M. (2006). Editor's comments: Publishing theory. *Academy of Management Review*, *31*(2), 252–255.

Kline, R. B. (1998). *Principles and practices of structural equation modeling.* New York: Guilford Press.

Lado, A. A., Boyd, N. G., Wright, P., & Kroll, M. (2006). Paradox and theorizing within the resource-based view. *Academy of Management Review*, *31*(1), 115–131.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–49.

Makadok, R., Burton, R., & Barney, J. (2018). A practical guide for making theory contributions in strategic management. *Strategic Management Journal*, *39*, 1530–1545.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher-order factor models and their invariance across groups. *Psychological Bulletin*, *97*(3), 562–582.

McKinley, W., Mone, M. A., & Moon, G. (1999). Determinants and development of schools in organization theory. *Academy of Management Review*, *24*(4), 634–648.

Miller, K. D., & Tsang, E. W. K. (2010). Testing management theories: Critical realist philosophy and research methods. *Strategic Management Journal*, *32*, 139–158.

Miner, J. B. (1984). The validity and usefulness of theories in an emerging organizational science. *Academy of Management Review*, *9*(2), 296–306.

Miner, J. B. (2003). The rated importance, scientific validity, and practical usefulness of organizational behavior theories: A quantitative review. *Academy of Management Learning and Education*, *2*(3), 250–268.

Mintzberg, H. (2005). Developing theory about the development of theory. In K. G. Smith & M. A. Hitt (Eds.), *Great minds in management: The process of theory development* (pp. 355–372). Oxford: Oxford University Press.

Ouchi, W. G. (1980). Markets, bureaucracies, and clans. *Administrative Science Quarterly*, *25*(1), 129–141.

Oxley, J. E., Rivkin, J. W., & Ryall, M. D. (2010). The Strategy Research Initiative: Recognizing and encouraging high-quality research in strategy. *Strategic Organization*, *8*(4), 377–386.

Ployhart, R. E., & Bartunek, J. M. (2019). Editors' comments: The is nothing so theoretical as good practice – a call for phenomenal theory. *Academy of Management Review, 44*(3), 493–497.

Poole, M. S., & Van de Ven, A. H. (1989). Using paradox to build management and organization theories. *Academy of Management Review*, *14*(4), 562–578.

Popper, K. R. (1959). *The logic of scientific discovery.* New York: Basic Books.

Priem, R. L., & Butler, J. E. (2001). Tautology in the resource-based view and the implications of externally determined resource value: Further comments. *Academy of Management Review*, *26*(1), 57–66.

Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*, 537–560.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*, 473–489.

Rynes, S. L. (2005). From the editors: Taking stock and looking ahead. *Academy of Management Journal*, *48*(1), 9–15.

Sandberg, J., & Tsoukas, H. (2011). Grasping the logic of practice: Theorizing through practical rationality. *Academy of Management Review*, *36*(2), 338–360.

Shapira, Z. (2011). "I've got a theory paper –Do you?": Conceptual, empirical, and theoretical contributions to knowledge in the organizational sciences. *Organization Science*, *22*(5), 1312–1321.

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review*, *36*(2), 381–403.

Suddaby, R. (2010). Editor's comments: Construct clarity in theories of management and organization. *Academy of Management Review*, *35*(3), 346–357.

Sutton, R., & Staw, B. (1995). What theory is not. *Administrative Science Quarterly*, *40*(3), 371–384.

Tourish, D. (2020). The triumph of nonsense in management studies. *Academy of Management Learning & Education*, *19*(1), 99–109.

Turner, J. H. (1985). In defense of positivism. *Sociological Theory*, *3*(2), 24–30.

Turner, J. H. (1989). A positivist's reflections on the problems and prospects of general theory in sociology. *International Review of Sociology*, *3*(2), 79–106.

Venkatraman, N. (1990). Performance implications of strategic coalignment: A methodological perspective. *Journal of Management Studies*, *27*(1), 19–41.

Weick, K. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. (1977). Assessing reliability and stability in panel models. *Sociological Methodology*, *8*(1), 84–136.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, *14*(4), 490–495.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications.* New York: Free Press.

APPENDIX

Survey Instrument (30 items) and Resulting Theory Evaluation Scale (25 items)

After choosing a theoretical exemplar that they reported being most familiar with, respondents were given the following instructions:

Please answer the following set of questions with respect to [Author(s) (year): "Title." *Publication Journal*, volume(issue): page#–page#].

Please indicate the extent to which you agree or disagree with the following statements:

Item #	Dropped	Item text
N1	x	The theory explored novel phenomena.
N2		The theory introduced original constructs.
N3		The theory proposed new relationships between constructs.
N4		The theory offered different explanation for previously established relationships
N5		The theory's rationale and core propositions were intriguing
N6		The theory contributed to a different way of looking at organizations
N7		The theory created an "aha" moment when proposed
E1		The theory's rationale could be extended to shed light on new questions
E2		The theory could be used to understand relationships at both the "micro" and "macro" levels of analysis
E3		The theory as presented could be extended to a broad range of phenomena
E4	x	The theory as presented did not realize its full potential
E5		The theory provided opportunities for further development
P1		Understanding of the theory helps practitioners formulate action plans
P2		A manager who understand the theory will make better decisions than one who doesn't
P3		The theory is so abstract that it has limited impact on real life decisions [reverse scored]
P4		The theory generates usable knowledge in the real world
P5		The theory offers a good tool for business consultants
P6		The theory addresses issues relevant to organizations
P7		Predictions from the theory are not interesting to practitioners [reverse scored]
P8		The theory leads management practice to address crucial questions
F1	x	The theory makes unambiguous predictions
F2		The theory can be tested against facts
F3		Predictions from the theory can be contrasted with empirical evidence
F4		The theory's constructs are easily translated into measurable variables
F5		It is easy to identify the independent and dependent variables of the theory
LC1		The theory did not improve human understanding of the world [reverse scored]
LC2		There were inherent contradictions in the theory [reverse scored]
LC3		The propositions derived from the theory were tautological [reverse scored]
LC4	x	The theory's propositions were properly deduced from its assumptions
LC5	x	The theory provides a strong justification of its predictions

Responses were recorded electronically using the following scale:

Strongly Disagree	Moderately Disagree	Slightly disagree	Neither Agree nor Disagree	Slightly agree	Moderately agree	Strongly agree
1	2	3	4	5	6	7

Endnotes

1. The co-authors divided this task so that each article was coded by a single rater.
2. We should note here the existence of a trade-off between extendibility through broader scope and the precision of a theory's predictions (e.g., Makadok et al., 2018). We will come back to a discussion of this trade-off later on.
3. In other words, some of the ambiguity in management theorizing is unjustified.
4. The trade-offs here are with parsimony and, importantly, with logical consistency, which will also be discussed below.
5. The rise of *Knowledge Management* as a distinct sub-discipline in the field of management provides a case in point.
6. Note that empirical falsifiability is only one aspect of the broader concept of Popperian falsifiability (Popper, 1959). Indeed, Popper's concept of falsifiability of a theory subsumes both internal consistency and empirical testability as defined here.
7. Some participants were members of two, or all three, of these divisions. Limiting the study to U.S.-based scholars was done to facilitate telephone follow-up during data collection. As discussed below, future studies with other samples might be warranted.
8. Respondents who reported lack of sufficient familiarity with any of the listed references were not allowed to proceed, and the survey would end for them.
9. As opposed to using regression imputation (Rubin, 1996), personal mean imputation (Bernaards & Sijtsma, 2000), mean substitution (Roth, 1994), or any other sampling distribution inference technique.
10. Note that this measurement model is also equivalent to alternative models with any four factors loading on the second-order factor, plus a covariance between the exogenous factor and the second-order factor. Any such models will produce the exact same fit as when all five factors load on the higher-order factor (e.g., Marsh & Hocevar, 1985).
11. The maximum inter-factor correlation was .67.