Intervention as a Research Strategy

Rogelio Oliva
Mays Business School
Texas A&M University
College Station, TX 77943-4217
roliva@tamu.edu

Under review at the *Journal of Operations Management*.
Please contact the author for latest information for citation.

Intervention as a research strategy

Abstract

JOM has made Design Science central to its development strategy creating a department to incubate papers using this methodology. An inaugural editorial published in 2016 called for papers with a generic design supported by a *design proposition* that provide pragmatic guidelines to bridge the gap between the case specific and the universal. In this article, I propose ways in which the scope of interventions can be expanded beyond the editorial's proposal to foster theoretical developments and better align with *JOM's* mission. Specifically, I propose that rather than focus on the design propositions, we explore the role that interventions (the ultimate manifestation of design science) can play in testing and developing theory (the ultimate goal of an academic endeavor). Taking as a point of departure the principles of *action research* and explanatory framework of *process theories*, I propose frameworks for leveraging interventions as a mechanism for testing existing theory and develop theories about organizational and system transformation. I illustrate the application of these two frameworks in the context of a previously published research paper.

1 Introduction

The *Journal of Operations Management* (*JOM*) has made design science central to its development strategy creating a department to incubate papers using this methodology. Design, according to Simon (1981, p. 129), "is the principal mark that distinguishes the professions from the sciences," and its concern with artificial systems that do not yet exist contrasts with the natural sciences' interest in the description and explanation of existing systems. For design, the main question becomes "will it work?" as opposed to the natural sciences' "is it valid or true?"; design's view of knowledge is thus pragmatic (in the service of action) (Romme, 2003). Simon's ideas are informing the development of design methodologies that include formal efforts to devise design guidelines in organization and management studies (e.g., Romme, 2003; van Aken, 2004; van Aken and Romme, 2012), information systems research (see Rai, 2017 for an exploration of the diverse design science research strategies in information systems), and operations management (e.g., Holmström et al., 2009).

A common approach to design science entails a *generic design* supported by a *design proposition* that follows the pragmatic logic: "For this problem-in-**C**ontext it is useful to use this **I**ntervention, which will produce, through these **M**echanisms, this **O**utcome" (Denyer et al., 2008). The inaugural editorial of the Design Science department by van Aken et al. (2016) explored this approach, framing design propositions as pragmatic guidelines intended to bridge the gap between the case specific and the universal. The objective is a generic design that transcends the specific situation within which it was developed. Such a

generic design corresponds to a mid-range theory (Merton, 1968). The key criteria for judging a submissions following this approach are pragmatic validity (How strong is the evidence that the design will produce the desired result?) and practical relevance (Does the design make a valuable contribution towards addressing a significant problem?) (van Aken et al., 2016, pp. 1-2). Both criteria are predicated on an intervention, or series of interventions, to improve a problem in context. The editorial equates design science research with engineering for operations management (OM) (§1.2), but recognizes that human agency, prevalent in the socio-technical systems of interest to OM (e.g., Croson et al., 2013; Gino and Pisano, 2008; Morrison and Oliva, 2019), requires special handing beyond the technical treatment of most engineering disciplines.

The proposed intervention-based research initiative is a useful framework to enable empirically based theory development. First, theories, "ordered set[s] of assertions about a generic behavior or structure assumed to hold throughout a significant range of specific instances" (Sutherland, 1975, p. 9), are never finished products, but rather exist along a continuum (Mohr, 1982, p. 6; Runkel and Runkel, 1984, pp. 129-130; Weick, 1989, p. 516). That is, a theory is more complete as its generalizations become more ordered, the focus of its constructs on the generalizations become more generic, and the range of instances it explains expands (Weick, 1989, p. 517). Using theory to predict or control constitutes a more demanding test than using it to describe or explain, and reflects greater theory maturity (Edmondson and McManus, 2007). Second, using a theory as a basis for an intervention is a way of "being empirical" that transcends observing and recording regularities or testing theories through experimental manipulation (see Carlile and Christensen, 2005, for a description of the iterative process to develop descriptive and normative theories). Using theory to intervene provides an additional test of *practicality* and *usefulness* often ignored in OM, but critical in a managerial context with clear cost-benefit implications (Fisher, 2007; Gallien et al., 2015; Tang, 2015; Toffel, 2016). Finally, that any attempt to effect change in a human situation must contend with the possibility that autonomous human beings might modify their behavior to refute or conform to a pre-conceived theory "means that the would-be rational intervener in human affairs cannot separate theory and practice in the way that the natural scientist can" (Checkland, 1985, p. 757), necessitating continuous adjustments to the theory that informs the change as well as deployment of that theory. This gives rise to an autopoietic process whereby theory leads to practice, but practice is the source of further theoretical developments. OM, if is to have an impact in practice and give credence to Lewin's (1945) maxim that "there is nothing as practical as a good theory," must be open to the feedback that results from putting its theories to the practical test.

As framed in the editorial, however, the initiative suggests two significant issues. The first concerns the required standard of evidence for a design science research contribution. The editorial both requests a "saturated body of evidence of validity" (p. 6) to ensure that the proposed design is capable of producing

the desired outcomes (pragmatic validity), and solicits generic designs, that is, *generalizable* design propositions that will be relevant beyond a specific application domain (practical relevance). Whereas the practical relevance criterion aligns with the traditional concerns of generalizability and requires multiple tests beyond the specific application domain, the pragmatic validity expectation seems to require multiple tests within the specific domain; together, the standards of evidence seem onerous, at least in the usual way research for publication in academic journals is conducted. Indeed, one of the few publications in this department (Akkermans et al., 2019) describes an engagement with an organization that lasted more than fifteen years and required multiple iterations of interventions and changes to the design propositions.

The second concern that emerges from the guidelines offered by the editorial regards fit with *JOM*'s stated mission of publishing "original, empirical operations management research that demonstrates both academic *and* practical relevance." As discussed above, *design propositions* are, by definition, pragmatic. They certainly address the second requirement of *JOM's* mission. They are not, however, theory, that is, are not "explanatory generalizations … of behavior … conceptualized as being recurrent" (Mohr, 1982, p. 5). Theories require a different standard of evidence and development beyond "does it work?", the implicit standard of design science. Focusing on *design propositions* risks falling outside of *JOM's* originally intended remit and producing publications perhaps better suited to practitioner journals (e.g., *Journal of Applied Analytics*, *Journal of Applied Operational Research*).

I propose an alternative framing that leverages what I consider to be the principal merit of the *JOM* design science initiative, that is, submitting ideas/models/frameworks to the ultimate *empirical* assessment by testing their usefulness and applicability, while addressing the concerns I raise above regarding standards of evidence and potential theoretical contributions. This framing suggests shifting focus from *design propositions* (the goal of design science) to the role an *intervention* (the archetypical manifestation of design) can play in testing and developing *theory* (the goal of an academic endeavor and foundation for future practice). Specifically, I argue that interventions can be used to *test* existing, and provide rich and unique evidence for the *development* of new, theories.

The article is structured as follows. I present in the next two sections the theoretical underpinnings and logic that support the notion of intervention-based research. In §2, I describe the principles of *action research* and present a derivative framework that has been successfully used for theory testing and development. In §3, I elucidate the idea of *process theory* and how interventions are a unique source of evidence for developing this type of theory. I then illustrate the use of these frameworks by analyzing, in §4, a previously published OM intervention and showing how the intervention can be leveraged for insights and theoretical improvements. I conclude, in §5, by summarizing the argument and elaborating the benefits of this approach to *JOM* and the OM research and practice communities.

2 Interventions as a test of theory

Learning-by-doing (experiential learning) is well established in the pedagogical literature and a proven approach to professional learning and development (see Kolb, 2014, for a current statement of experiential learning theory). The idea of a purposeful intervention (i.e., doing), however, challenges one of natural science's core precepts: that the scientist must not influence the experimental outcome and should instead be a neutral and unbiased reporter of what happens. The extent to which OM, or, for that matter, most organizational theory, espouses the guiding principles of natural science explains why this method of learning is not widely adopted as a research strategy. This is also why controlled experiments remain the gold standard for theory development. Social scientists, however, have had to develop alternative research strategies for contexts in which it is neither ethical nor feasible to perform controlled experiments (see Krimerman, 1969, on the challenges of pursuing social sciences). Among many methods and research strategies developed to work around ethical and practical constraints in the social/human context (see Lewis-Beck et al., 2003, for comprehensive list), action research leverages purposeful intervention. In the remainder of this section I articulate the guiding principles of action research and expand how this research strategy can be used to gain insights from intervention.

2.1 Action research[1]

The tradition of intervention-based research can be traced to action research (AR), originally proposed by Lewin (1946) in response to "the limitations of studying complex real social events in a laboratory, the artificiality of splitting out single behavioral elements from an integrated system" (Foster, 1972, p. 530), and the conviction that, because homeostasis (dynamic equilibrium) reveals only a fraction of the interactions and dependencies in a complex social system, "one cannot understand a human system without trying to change it" (Schein, 1987, p. 29). Foster gives a formal definition of AR based on an earlier one by Rapoport (1970):

> A type of applied social research differing from other varieties in the immediacy of the researcher's involvement in the action process and the intention of the parties. ... It aims to contribute both to the practical concerns of people in an immediate problematic situation and to the goals of social science by joint collaboration within a mutually acceptable ethical framework (Foster, 1972, p. 534).

The nature of AR engagement fundamentally alters the traditional researcher's role. Schein (1987, 1993) argues that AR is the operating mode of clinical research, which he distinguishes from the non-obtrusive observation common in ethnographic research. Under the clinical research "agreement," the main concern of the researcher, much like that of a consultant, is to help the client achieve the agreed-upon goal, ceding control of the research area and duration of study to the client's appreciation of usefulness. This constraint has consequences for the processes of data acquisition and analysis and limits

[1] Some of the ideas presented in this section were originally developed in Oliva (1994).

the results that can be derived from a study (Schein, 1987). It also clearly entails some risk for the researcher to the extent that the focus on the success of the intervention might compromise desired research design protocols.

Action research, as defined by Lewin, involves iterative cycles of identifying a problem, planning, acting, and evaluating (Argyris et al., 1985). This generic intervention process—identify-plan-act-evaluate—implies a clearly stated methodology of intervention, a prediction of the outcome of the intervention, and an assessment of the effectiveness of the theory in use. That is, there cannot be action research without an explicit statement of assumptions and intentions, i.e., a normative theory (Schein, 1987), and without assessment of the theory's effectiveness. This later reflection is what separates AR from a typical consulting engagement. Note, however, the client might not be interested in the theoretical assessment once the intervention results have been achieved. Further, that the researcher becomes an active member of the group attempting to effect change implies that the intervention itself might be adjusting to the reactions of the agents in the situation, thereby rendering the intervention strategy dynamic. The adaptiveness of the intervention strategy as evidence emerges is what separates AR from evaluation research that attempts to assess the effectiveness of a past intervention (Pawson, 2002).

Checkland (1985) provides more details on the above process by proposing that an intervention in the AR context is guided by a deliberate strategy or methodology (M) that makes use of some linked ideas in a theoretical framework (T) to improve a situation (S).[2] What Checkland calls the intervention strategy or methodology is, in effect, a *design proposition* (in the design science sense) that attempts to provide specific guidelines for improving the situation by means of the ideas in the theoretical framework. The ideas in the theoretical framework refer to a set of theories considered relevant to the intervention that could include, among others, normative theories that inform the desired state after the intervention, theories about the intervention itself (e.g., change management), and theories that inform how to handle constraints to the implementation process (e.g., theory of incentives). See Table 1 for expanded definitions of these constructs.

[2] Checkland (1985) uses "conceptual framework (F)" instead of "theoretical framework (T)" and "area of application (A)" instead of "problem situation (S)"; I adapted the notation for easier mapping into the design science nomenclature used in the introduction.

T. Theoretical framework	Set of theories considered relevant to the intervention that could include, among others, normative theories that inform the desired state of the situation after the intervention, theories about the intervention itself, and theories that inform the implementation process.
M. Methodology for using T	Strategy that outlines how the theoretical framework T will be used or deployed to improve the problem situation S. Somewhat equivalent to a *design proposition* in that provides specific guidelines for improving the situation by means of the ideas in the theoretical framework/
S. Problem situation	Real-world situation that is thought it can be improved through the intervention.

Table 1. Elements of an intervention

A methodological intervention, i.e., an intervention in which the methodology is explicitly stated, documented, and reflected upon, will generate a number of mismatches between intentions and assumptions and what actually happens when the methodology is put into practice. Such mismatches generate opportunities to update and correct understanding of the problem situation and the models used to intervene in it (Argyris and Schön, 1978). In contrast to design science, which emphasizes the development of design propositions, Checkland argues that an intervention (i.e., the confrontation of the situation with the theoretical framework via the methodology) is capable of generating insights (learning) relevant to each context. Specifically, an intervention can yield lessons about: i) the problem situation S (which is what permits the situation to be improved); ii) the ideas in the theoretical frameworks T employed in the intervention; and iii) the approach M used to resolve the problem situation (i.e., the design that renders use of the framework in the problem situation feasible). Figure 1 depicts the three learning levels enabled by an intervention.

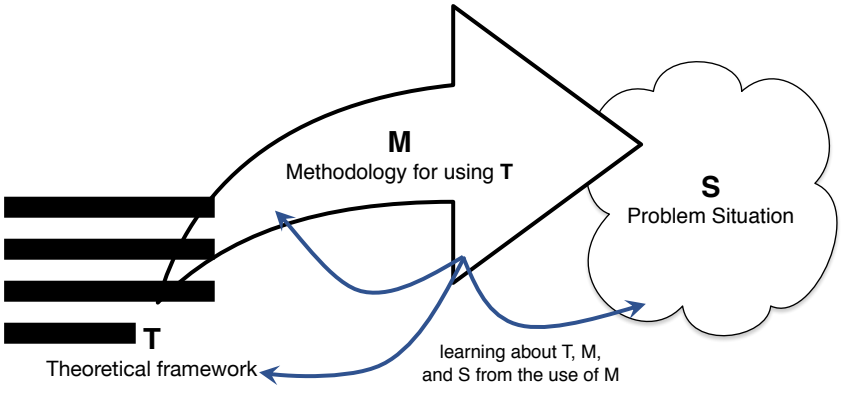

Figure 1. Learning from interventions
(adapted from Checkland, 1985)

The learning mechanisms portrayed in Figure 1 illustrate how its use in an intervention might reveal shortcomings in T or suggest needed improvements to the theory to explain the results obtained. I refer to this in the previous section as using an intervention to empirically test the usefulness and applicability of existing theory. As discussed above, lessons or insights emerge only when something unexpected happens, that is, when there is a disconnect between what a design suggests will happen and what actually

occurs when an intervention is executed. A successful intervention in which everything goes as expected is probably not useful in terms of learning as it merely confirms what is known and expected about S, T, and M. That there is no guarantee that a particular intervention will yield insights worth publishing represents yet another risk for the enterprising action-researcher.

The feedback loops proposed by Checkland also capture the iterative process described in design science whereby lessons learned from attempts at implementation inform subsequent adjustments to a design. Checkland employed this description of the AR process to portray the iterative research program deployed at Lancaster University to develop Soft Systems Methodology. The research program initiated in the late 1960s with the intent of using Bell Laboratories' systems engineering approach to solve problems in managerial contexts (human situations). Each year, a new cohort of graduate students was introduced to the latest version of the concepts T and methodological guidelines M and required to apply them in an unstructured real-world problem situation S. The master's thesis requirement was a project write-up, two-thirds of which was to be allocated to a description of what was done and why, the remaining third devoted to lessons or insights about T, M, and S gained from the experience. The lessons were used to update the theoretical frameworks and methodological guidelines taught to the next cohort of students. Active development of Soft Systems Methodology lasted 30 years and involved more than 300 projects (Checkland, 2010), my own master's thesis (Oliva, 1988) making a small contribution to the effort.

2.2 Insights from an intervention

The development of Soft Systems Methodology reveals how hundreds of interventions across different types of situations are capable of generating insights about theories, methodological guidelines, and even application domains. A single intervention, or even a series of iterative interventions on a particular site, however, present significant generalizability challenges. In what follows, I focus on the learning potential from individual interventions as they better match the traditional research strategy required to publish in academic journals.

In most cases, lessons from a single intervention about problem situation S will have low generalizability and be of interest only to managers in the focal situation. Lessons or insights related to methodological guidelines M are trickier as they bridge the application of general theories T to specifics of problem situation S. I posit that the generalizability requirement demanded by design science (or intervention-based research) can be satisfied only to the extent that methodological guidelines speak to better uses or applications of the theoretical framework T. Conversely, methodological insights derived to better cope with the constraints of problem situation S are, generally speaking, ad-hoc adjustments any capable engineer or consultant would make to achieve desired results. The insights that enable an

intervener to better adapt given methods to a problem situation have little generalizable value, being, by definition, adjustments to the specifics of situation S. Lastly, I submit that *JOM*, and academics in general, should be most interested in lessons related to theories T. Not only are reflections about theory more generalizable but contrasting evidence from interventions with documented statements from the theory renders a contribution unambiguous. For example, a reviewer might counter that insights from S and M, albeit new to a particular intervenor, are well known and do not constitute a substantive contribution. Claiming a theoretical insight, however, requires a well-documented statement of the theory, evidence of why the theory was insufficient to fully inform the intervention (or how it failed to anticipate some specific circumstance), and proposed amendments and improvements thereto. Although the onus remains on authors to defend their theoretical propositions, having the conversation relative to documented theories eliminates the subjectivity of assessments of whether or not a particular insight is new. From a *research* perspective, two fundamental questions should emerge from an intervention, what shortcomings of T were revealed by, and how should T be updated or expanded as a result of, the intervention.

This is not to say that insights about S and how to adjust M to fit better S are not useful or needed. Those insights, however, have a different audience than the target audience of a journal with theoretical aspirations, a point on which I elaborate when presenting examples of how to use this framework in §4 below.

3 Interventions as a source of theory

Moving on from the notion that interventions make explicit use of existing theory, I emphasize in this section that interventions also have a goal, specifically, that of changing problem situation S. Just as in the previous section we derived insights from "mismatches" between the expectations and reality of T, M, and S, I posit here that it is possible to use mismatches between the planned transformation path to desired outcome S* and actual transformation path to observed outcome S' to gather evidence about how the transition was achieved and what accounts for the outcome. Figure 2 depicts the transition achieved by the path to end-state S'.

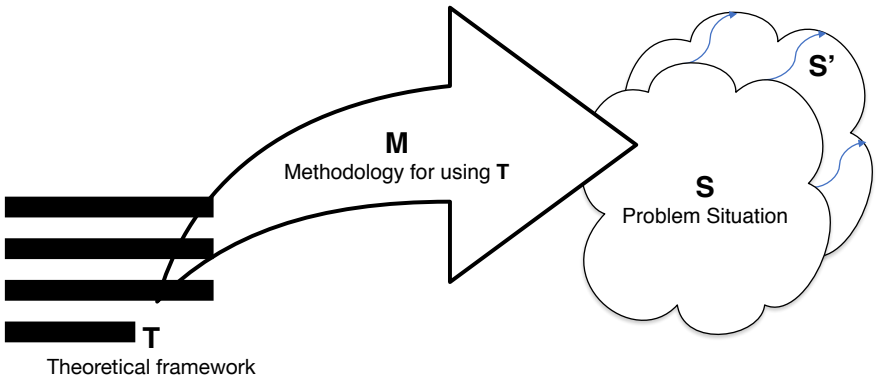

Figure 2. The outcome of an intervention

The emphasis here is not on the desired (S*) or achieved (S') endpoints, but rather on the transition path of the problem situation changing from S to S' instead of the expected path from S to S*. What we have in an intervention is a deliberate attempt to achieve a specific transformation. To the extent that a desired or targeted transformation is not achieved, despite deliberate planning and deployment of resources, the documented effort constitutes a rich and unique set of evidence with which to explore the questions: How did we get there? Why did we get there?

Although this might seem like a model for a case study with a sample of one, I believe an intervention can be leveraged for theoretical development if, instead of using the individual case as an occurrence to populate a variance theory, we use the evidence to develop (and implicitly test) process theories (Mohr, 1982; Monge, 1990; Poole et al., 2000), see Table 2. I explore below the main characteristics of process theories and how this type of theoretical framework can be used to leverage interventions for theory development.

S→S*	Expected transformation path as a result of the intervention from the initial problem situation S to the desired end-state situation S*.
S→S'	Actual transformation path as a result of the intervention from the initial problem situation S to the observed end-state situation S'.
Process theory	Explanations that identify how entities participate and are affected by events (see expanded definition and attributes in Mohr 1982 and Poole et al. 2000). Process theories provide causal explanations for how and why things happen and can be used to make sense of the observed transformation path as a result of the intervention.

Table 2. Explaining of the intervention outcome

3.1 Process theories

Mohr (1982) advanced the notion that there are two distinct approaches to explanation. These take the form of variance theories and process theories. The former are concerned with variables (quantifiable metrics) and efficient causes (in the Aristotelian sense), i.e., that which causes the change motion to start and stop. In variance theories, the basis of explanation is causality whereby the precursor is a necessary and sufficient condition for, but the time ordering among the independent variables immaterial to, the

outcome. Process theories, in contrast, identify how entities participate in and are affected by events and provide causal explanations of how and why things happen. The process theories articulated by Mohr reflect Simon's (1981) ideas about design science and earlier work that suggested them (e.g., Kaplan, 1964; Lave and March, 1975). Mohr takes these ideas further by articulating the logical and philosophical requirements for process theory development, specifically, that process theories focus on a final cause, that is, a purpose for the sake of which things are done, where precursor X is necessary for outcome Y, rather than the efficient causes required by variance theories. Moreover, time ordering among contributing events is critical to outcomes in process theories. The ideas and logic behind process theory have continued to evolve (see Chapter 2 of Poole et al., 2000 for an expanded set of assumptions) and were the focus of a special issue of *Academy of Management* (see Langley et al., 2013, for an overview and discussion of the current state of process theories). These efforts have been accompanied by the development of research strategies and designs for process studies (e.g., Dooley and van de Ven, 1999; Langley, 1999; Monge, 1990; Pentland, 1999; Pettigrew, 1990; van de Ven and Engleman, 2004; van de Ven and Poole, 1995) and corresponding methodological innovations in grounded theory development (e.g., Strauss and Corbin, 1990), econometric developments in time series (e.g., Greene, 1997) and panel data (e.g., Wooldridge, 2002), and extensive use of simulation to explore organizational dynamics (e.g., Sastry, 1997; Sterman et al., 1997).

Process theories have a clear standing in the natural sciences, but historically have played a minor role in the social sciences, where the literature exhibits a strong bias towards variance theories. Monge (1990) posits structural reasons for this bias. First, most empirical research has been conducted at a given time. Program and funding requirements and pressure to publish leave most researchers insufficient time to perform detailed longitudinal studies needed to develop process theories. Consider the theory of evolution, perhaps the archetypical process theory. It took five years on a round-the-world data-gathering trip funded by the Admiralty and 23 years of further data collection and analysis as a financially independent scientist before Darwin felt comfortable publishing *On the Origin of Species* in 1859. In the absence of such resources, researchers have traditionally performed cross-sectional studies that allow for the articulation of causal explanations in terms of variance across a sample. As Monge (1990, p. 407) bluntly states: "the field suffers from a form of methodological determinism." Monge further attributes the shortage of process theories in the literature to the fact that organizational and social scientists lack the conceptual tools to develop theories that involve dynamics. According to Weick (1987), organizational and social scientists tend to perform verbal and linguistic analyses akin to the kind of analysis historians might perform rather than use the formal mathematical apparatus (i.e., calculus) developed to describe process dynamics. Sastry (1997) is exemplary of the benefits of mathematically formalizing a process theory that had, to that point, been described only as an incomplete, and not always consistent, set of

verbal relationships among variables (Tushman and Romanelli, 1985). Formalization of a process theory also accommodates incorporation of "layers of explanation ranging from immediate to distal" (Poole et al., 2000, p. 46). For example, Oliva and Sterman (2001) integrated a number of hypotheses about service delivery independently developed as variance theories into a formal simulation model that, calibrated to a specific site, revealed the mechanisms (a process theory) of the observed erosion of service quality, and expanded the investigation to identify the range of parameters that would sustain that erosion across varied service settings (Oliva, 2001).

3.2 Process insights from an intervention

I believe intervention-based research has the potential to address the foregoing challenges to development of process theories in the social sciences. Interventions are an effective way to short-circuit the delays normally associated with longitudinal analysis. As explained above, they entail a deliberate effort to achieve a change, and continuous adjustment of resources and strategies to achieve the desired end. The behavior occasioned by the change is observable as the intervention unfolds. It is possible to derive from a properly monitored and documented intervention a *dynamic hypothesis* that explicitly articulates how the structure of the organizational context, changes introduced, and behavioral response of stakeholders in the situation are jointly responsible for the observed behavior (Morrison and Oliva, 2019; Sterman et al., 2015).[3] The process of articulating a formal description of how structure (i.e., the physical and behavioral elements of the system) is responsible for observed behavior creates a refutable causal model with "multiple points of testing" (Bell and Bell, 1980; Bell and Senge, 1980). See Oliva (2003) for testing strategies for hypotheses that link structure to behavior.

This is in line with Schein's (1987) argument that it is through change that we understand and learn from complex social systems. Through a detailed description of the change initiative and careful observations of the response of the social system it is possible to derive a process theory that explains the evolution of a system that describes how entities participate and are affected by events. Moreover, application of this process perspective is not limited, as it is the case for action research, to interventions where the researcher is an active participant of the change team. For example, Oliva and Watson (2009, 2011) were able, following a major successful intervention during which they were not engaged with the focal organization, to conduct a post-hoc analysis of the intervention strategy and system response

[3] Note that the explanation for why the system transitioned to S' is not simply the deployed methodology M. Had only M been responsible for the observed transition we would have expected (assuming the methodology was developed with some competence) to reach S*. That S* was not achieved suggests that other structural explanations beyond M need to be explored, such as unanticipated side effects, policy resistance, and so forth (Sterman, 2000).

through semi-structured interviews, secondary data, and direct observation of operations post-intervention. By reconstructing the timeline of the intervention strategy, documenting the main actors' responses and design team's adjustments to the intervention strategy, and triangulating stated responses with operational data, they were able to reconstruct the rationale for the responses and adjustments and develop a dynamic explanation (a process theory) for successful organizational alignment despite purposefully misaligned incentives across departments.

Regarding the second obstacle to the dissemination of process theories, I submit that OM scholars by and large, by virtue of having a *process* focus and caring about process performance and possessing the training and mathematical background to formalize process descriptions and dynamics, are better suited than organizational and social scientists to leverage process theories. That we have not, however, placed sufficient emphasis on process theory is evidenced by a Google Scholar search of the term "process theory" (absent "dual," to exclude references to psychology's "dual process theory") in the top three operations management journals (*JOM*, *M&SOM*, and *POM*), which yielded 16 articles out of 5,582 entries. I hypothesize that we have not articulated process theories because we are not trained to look for them (Monge's methodological determinism). I believe the proposed initiative to leverage intervention-based research could afford the data needed to pursue development of such theories.

4 Example

I illustrate here how the questions derived from the frameworks above—How should T be updated as a result of an intervention? What explains the change seen through the transition from S to S'?—as well as the frameworks themselves can be used to articulate theoretical developments from interventions that satisfy the methodological and theoretical expectations of *JOM*. I do so by applying the frameworks to a study previously published in an OM journal. I adopted this strategy for two reasons. First, using a published intervention enables me to abbreviate its presentation here by referencing peer-reviewed documentation. Second, it is my intention to show how post-hoc application of the proposed frameworks can yield insights far removed from the initial intent of the intervention. I was a member of the team that published the original research, but am solely responsible for any errors and omissions in these post-hoc analyses.

Note that the study selected for this exercise cannot be considered a pure intervention, as at the time of publication such a focus was not a viable publication strategy in leading OM journals. The study was designed and written so as to adhere to the publication requirements for the chosen research design, a randomized field experiment, and the original publication does not include all the evidence to support my post-hoc analyses. Where appropriate, I provide some of the missing evidence. Despite these tradeoffs, I

hope to show how the study could have evolved as viable intervention-based research had the experimental design not been in place.

4.1 Intervention design and results

The intervention in this example was undertaken at the request, and with the assistance, of a consumer product manufacturer (supplier) and retail service provider (RSP) interested in assessing whether it was possible to develop an external auditing service to complement retailer operations in order to reduce the number of shelf-out-of-stocks (shelf-OOS), that is, instances in which a product is available in the store, but not accessible to the customer. Full descriptions of the experimental setup, experimental design, and results are provided in Chuang et al. (2016).

The main idea behind the intervention was to develop and test an alert system that utilized principles of Statistical Process Control (SPC) to trigger shelf audits by the RSP in the event that unlikely sales patterns were observed (Fisher and Raman, 2010). Specifically, we leveraged retailer point-of-sales (POS) data to trigger audits whenever a sequence of zero-sales days unlikely to be observed if the item was available to consumers was detected. The intervention was set up as a randomized field experiment with 60 stores, evenly divided between test and control groups, and was to run for twelve weeks. We monitored the sales patterns for four of the supplier's stock keeping units (SKUs). Although the retail stores operated seven days a week, the RSP were capable of executing audits only during weekdays. The review process was implemented as follows.

> Each morning the supplier retrieved the retailer's POS data and inventory position for the item-store pairs in the treatment group (4 SKUs in 30 stores). The data were then transferred to the retail service firm where we assessed sales, inventory positions, and the probability of an OOS. By noon, we triggered store visits to be executed by the auditors the following day. The auditors downloaded these requests at the end of the business day, as they prepared for the inventory audits to be conducted the following day.
> When executing the physical audits, auditors were supposed to verify and correct inventory information and address store and shelf-OOS for all the SKUs in the sample, and not just those items responsible for the triggered audit. After each visit, the auditors submitted an audit report that reflected the conditions found relative to the item status and inventory record. (Chuang et al., 2016, p. 939).

One challenge of field, as opposed to laboratory, experiments is that it is often impractical (or impossible) to control for exogenous sources of variance. Our field experiment was no exception. The intervention was designed to assess the effect of audits on sales, but the net effect of those audits had the potential to modify the demand patterns used as a baseline to develop the predictive models for assessing the likelihood of zero sales. Moreover, we were unsure of the RSP's ability to execute the audits in a timely manner and with acceptable quality. We thus expected changes not only in the performance of the stores, but also in the quality of the predictive models used to trigger, and the RSP's ability to execute, the audits. In the original publication of the research, we refer to this dynamic set up as three simultaneous

plan-do-check-act cycles that compromised the experimental protocol (see Figure 1 in Chuang et al., 2016).

Adjustments to the original design had to be made to accommodate existing capabilities and contextual idiosyncrasies, the first even before the experiment began. While developing the demand models to estimate the baseline probability of a zero-sale day, we found that approximately 95% of the SKUs' demand patterns in the experiment fit a negative binomial distribution. After a week of searching for alternative models for the other five percent of SKUs, a decision was made by the RSP that fitting 95% of the SKUs was "good enough" as full-scale deployment of the service would not accommodate too many or too complex demand models being evaluated. We thus adopted the negative binomial distribution for all item-store demand in the treatment group. Second, after one week of operations, we changed the lower level of all triggers, regardless of the predictive model's suggestions, to wait until the consecutive number of zero-sales days was at least equal to three. We did this because we found that the retailer was often capable of correcting the shelf-OOS within a day or two and therefore many of the triggered auditors' trips were unnecessary. This adjustment resulted in requiring more evidence that the retailer was not aware of the shelf-OOS before we trigger an audit. Third, as we discovered that many of the audit trips were unnecessary, i.e., the auditor did not find a shelf-OOS condition when visiting, we adjusted the sensitivity of the triggering criteria, not only to require further evidence of possible OOS, but also to incorporate a cost component (expected sales loss) to justify the cost of the audit. Lastly, during the first days of execution, we also found that RSP representatives were not executing all triggered audits in a timely fashion and not all information required to assess audit effectiveness was being captured. We made changes to both the instructions for the RSP representatives and forms used to collect data. As a result of the foregoing changes, we did not have a stable auditing protocol, and thus useful experimental data, until the third week of the experiment.

In the original publication, we reported the experimental results along five dimensions.

- Audit frequency: After a rush of audits when the review process was first deployed, we found that by the eighth week of the experiment, the system had reached a new steady state where very few audits were being triggered per week.
- Trigger accuracy: Although designed to have low sensitivity and high specificity, we found trigger accuracy to improve over the duration of the experiment as inventory record inaccuracies (IRI) were corrected as a result of the audits.
- Operational impact: A final audit of test and control groups at the end of the twelve-week period revealed the control group to be 3.4 times more likely to exhibit shelf-OOS, 1.4 times more likely to exhibit IRI, and 6.6 times more likely to exhibit shelf-OOS and IRI.

- Store performance: From a difference-in-difference estimation, we calculated the test group to experience a 21% sales lift during the last four weeks of the experiment.
- Economic viability: Taking into account monthly net benefits (sales lift – audit cost), we calculated one month of steady-state operation to be sufficient to cover the cost of the eight-week transition to the new steady state.

We noted in the original paper the potential for generalizability, as the reported results had been achieved with: i) a retailer that possessed excellent operational capabilities; ii) a product family with extremely low demand elasticity to supply (the product has strong customer loyalty and long shelf-life); and iii) an external (i.e., costly) monitoring audit team. Settings with more traditional values for these parameters would be likely to experience higher sales lift, greater economic benefit, or both.

As a proof of concept for the new service, the intervention was successful and the results encouraging (note that this is the design science criterion: it worked!). The article's initial submission nevertheless encountered some resistance from the editorial and review teams, the latter's main concerns being that the audit trigger policy was not optimal, and that the experimental protocol was compromised in the execution. Indeed, we had adjusted the protocol to ensure that the service provider could execute the triggers in an appropriate time frame and at reasonable cost. Optimality of the trigger policies was not even a concern to us; aware that our policy was, at best, directional, we were interested in testing whether the signal was sufficient to trigger adequate responses. Our challenge during the review process was to demonstrate that our results were robust despite the necessary adjustments to the research protocol, that optimality was beyond the scope of the experiment, and that we had produced transferable insights. The original publication aimed to satisfy those requirements. Absent from both the original publication and review team's line of inquiry were detailed justifications of or questions related to the changes needed to secure implementation. Implementation and execution of the intervention were taken for granted, the only concern being whether any of the changes compromised the generalizability of the findings. While a justifiable focus in the context of reporting the results of a randomized experiment, this omission removes from the reflection process the set of ideas with which design science is concerned and the potential theoretical developments that could emerge from reflection on the intervention process. I attempt to address these gaps below, using the frameworks proposed in §2 and §3.

4.2 Insights from the intervention

The framework proposed in Figure 1 identifies the changes and adaptations made to the original design in order to implement the triggering system and execute the audits. Specifically, in looking for surprises about the problem situation S we can point to the adaptations we had to make to the triggering criteria to

a) consider the retailer's ability to self-detect and -correct shelf-OOS, and

b) accommodate for the RSP representatives' response capacity.

Note that neither of these possibilities—self-correcting processes and limited response capacity—is considered in the guidelines for designing and deploying SPC (e.g., Montgomery, 1991).

Similarly, on the methodological M side, we were confronted with a variety of compromises from the ideal SPC implementation

i. Because of IRI, we were working with an inaccurate control signal.
ii. For some SKUs, we were working with inappropriate predictive demand models.
iii. Our control mechanism had a 36-hour response lag.
iv. We purposefully modified our triggers to adjust for a high-cost of executing the audits (low sensitivity, high specificity).
v. Actions occasioned by the SPC trigger were not oriented towards identifying the root causes of the problem, but instead limited to eliminating the shelf-OOS condition without further investigation of how to prevent future occurrences.

Although addressing, or at least accounting for, these challenges was necessary to make the intervention viable, our manner of accommodating them, via ad-hoc adjustments to algorithms and human resource practices, leaves little to be said about generalizable *design propositions*. The experiment worked *despite* those workarounds, themselves unremarkable adjustments made to deal with the reality of the resources available for the intervention and cultural feasibility of the deployed policies. My concern is that if we focus exclusively on *design propositions,* we will limit the caliber of insights and confidence on the claims will instead need to come from multiple implementations.

Whereas insights associated with S and M were relatively easy to identify as modifications requisite to the intervention's success, insights about T were not immediately obvious. I explain the difficulty of articulating insights about the theory in terms of what Argyris calls double-loop learning (Argyris, 1976). Whereas in single-loop learning, behavior is adjusted based on the mismatch between expectations and reality (e.g., insights about S and M), in double-loop learning the challenge is to question the values and assumptions that are the source of those expectations (i.e., the theory, T). Argyris (1985) and Argyris and Schön (1978) list the defense routines and cultural assumptions we use to maintain the congruency of our mental models. The defense mechanisms are in line with the Quine-Duhem thesis, which explains that core theories are often protected by attributing responsibility for surprising results to auxiliary assumptions (Cook and Campbell, 1979). As such, the process of questioning the theoretical framework requires deeper reflection, so it often happens in a context outside the immediacy of making the necessary changes so that the intervention works. In our case, that reflection process was aided by the effort to write and justify the experimental results. This reflection, however, is often a luxury that the practitioner responsible for the intervention cannot afford, thus limiting the potential learning from interventions. As a

result of the shortcomings of the theoretical framework in the foregoing intervention we developed three major lines of questioning (future research).

1. Why use zero-sales (z) to infer the probability of OOS? If we are seeing z, OOS may already have occurred rendering required adjustments late. Can we predict shelf-OOS? If not, is it possible to predict IRI and use that to infer the probability of shelf-OOS?
2. Although SPC triggers are not cost sensitive, ours were nevertheless compromised by cost considerations. In its original conception, SPC does not have to be cost sensitive; it assumes that a triggered intervention will address root causes, thereby eliminating the error-inducing process and yielding a perpetuity of savings from the eliminated problem. Our audits, however, did not remove the root-causes of, but merely corrected, the OOS condition. Clearly, cost tradeoffs need to be considered when triggering audits.
3. Although our triggers responded to the behavior of individual SKUs, triggered audits were performed on all SKUs in the sample. In practice managers typically think and talk about auditing policies for groups of SKUs. Can we develop audit policies for groups rather than individual SKUs?

Note how addressing the insights about T yields more generalizable and transferable concepts than addressing the insights about S and M. The foregoing questions are germane to any intervention to design audit policies based on sales data. Conversely, insights about M have more to do with how the theoretical frameworks were adjusted or compromised to accommodate specific constraints of the problem situation. That said, as discussed in §2.2, the distinction between lessons in M and T is often not clear and insights often overlap. Indeed, one can argue that insights iv and v above revealed theoretical shortcomings that led to the research outlined in the second line of questioning.

While this section illustrates how the framework informs reflection on the intervention to generate insights, this, in itself, does not guarantee that enough insights will be generated to warrant publication. Although identifying the potential contribution of addressing the foregoing research questions does not constitute a sufficient contribution, these questions nevertheless have become an active research agenda presently being pursued with one of the co-authors of the original study, and two papers addressing these questions are currently under review. I expand on the requirements for publication in the discussion section below.

4.3 Results of the intervention

I described above the main insights that emerged from confronting the concepts of statistical process control with the realities of retail inventory audits. The process launched a series of explorations to further expand the use of SPC-triggering capabilities in a context in which addressing root causes of the detected problem was not an objective of the intervention. In this subsection, I illustrate the use of the second

framework proposed above "What explains the change observed as a consequent of the transition from S to S'?" to develop a process theory that explains the surprising outcome. In what follows, I use the system dynamics (SD) notation of stocks, flows, and feedback loops (see Sterman, 2001, for an overview) to articulate the elements of the process theory (i.e., how entities participate in and are affected by events) and explain the observed behavior (see Sterman et al., (2015) for a discussion of the use of SD concepts to articulate process theories in the OM context).

As described in §4.1, we found the number of triggers to decline over the course of the intervention to a relatively low level of audit activity after the eighth week of the experiment. More specifically, we found the number of triggers exhibited an exponential decay consistent with a process improvement half-life of two weeks (Keating et al., 1999; Schneiderman, 1988), see Figure 3.

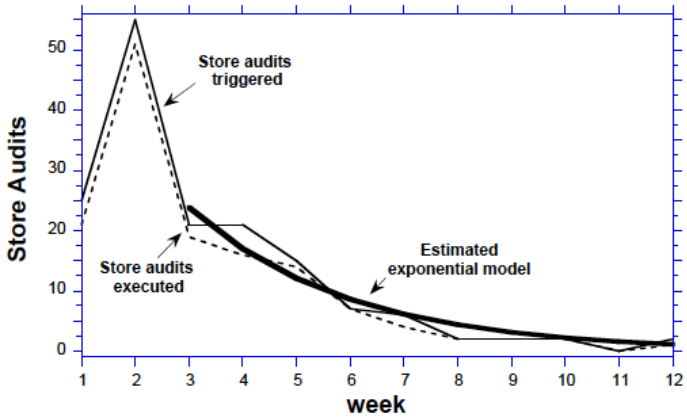

Figure 3. Weekly Store Audits
(source: Chuang et al., 2016)

Exponential decay of problems as a result of a process improvement intervention is a well-established empirical regularity (Keating et al., 1999; Schneiderman, 1988; Sterman et al., 1997). The explanation of this behavior is predicated on the notion that underlying process capabilities are being improved and fewer errors are thus being introduced into the system. In the context of retail inventory audits, a traditional Total Quality Management (TQM) intervention would have used the zero-sales signal to identify process improvements that would improve the retailer's operational capabilities, for example, reducing (perhaps through better shelving discipline) error introductions to shelves (loop B1 in Figure 4) and boosting (perhaps through more aggressive internal shelf checks and sorting) the shelf error detection and correction rates (loop B2 in Figure 4). Improved operational capabilities lead to a lower inflow and higher outflow to the stock of *errors in shelf*, which results in a decreasing number of *errors in shelf* and reduction of the audit triggers.

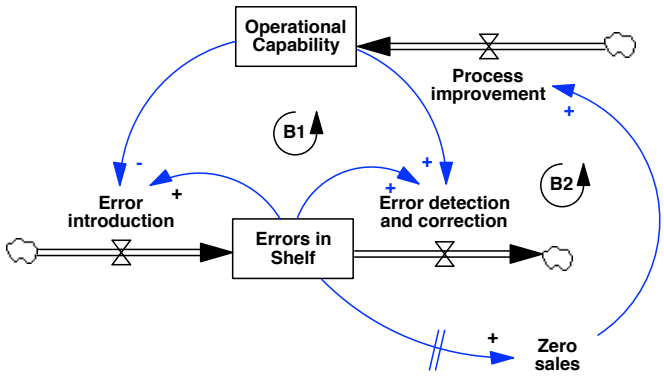

Figure 4. A TQM intervention

The intervention described in §4.1, however, is not represented by the diagram in Figure 4 as the audits were not used to identify the root causes (i.e., potential process improvements) of the identified errors, but rather were used exclusively to correct shelf errors (e.g., missing labels, stockouts, IRI, etc.). As such, the audits did not result in structural changes to the retailer's capabilities, but merely provided temporary relief by flushing out existing errors from the *errors in shelf* stock (see Figure 5). As long as the extraordinary audits (i.e., audits beyond the normal error detection and correction rate) are maintained, the process of continuously flushing the *errors in shelf* is sufficient to lower the steady state level as the dynamic equilibrium is now determined by the three flows in and out of *errors in shelf*.

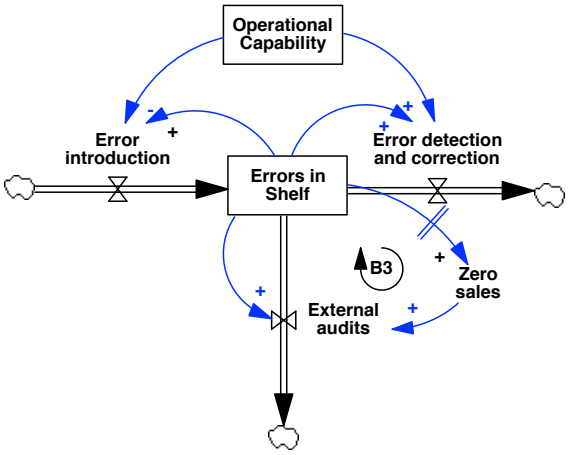

Figure 5. The effects of shelf-corrections on errors in shelf

That extraordinary audits are triggered by suspicion of OOS explains why the process of reaching the new steady-state follows exponential decay: the evidence that triggers audits becomes scarcer as the level of errors in the stock declines.[4] We can thus articulate the necessary conditions and causal mechanism required to specify a process theory (Mohr, 1982) as follows.

[4] A possible alternative explanation for the decay in audits is that the retailer's employees became aware of the special status of the SKUs in the experiment and were more careful in the re-stocking process. We rule out this

In normal operating conditions, retailers experience a steady state level of errors in shelf determined by the retailer operating practices and the interactions with the customers. In this context, an extraordinary audit and correction effort triggered by zero sales will lower the steady state of the *errors in shelf*. As evidence of errors becomes scarcer with the corrections, the audits will be triggered less frequently, resulting in a constant fractional (i.e., exponential) rate of decline of errors in shelf.

Once the theory that explains the dynamics of *errors in shelf* is in place, it is possible to expand the theory to explore under what conditions such external intervention would be sustainable long term. The sustainability argument is economic—extraordinary audits should be executed as long as the value generated exceeds their cost—and we can assess this by comparing the cost of executing the audits to the potential elimination of losses from avoided lost sales. The aggregate cost of the intervention is the sum of the number of audits multiplied by the cost to execute each, the benefit, is the reduction of losses from lost sales due to errors in shelf. The cost of losses is determined by demand elasticity (how many of the errors turned into lost sales as opposed to sales of an alternative product) and the retailer's margin. Figure 6 illustrates the structural determination of these costs and losses. The benefits of eliminating lost sales will outweigh the cost of the audit program if cost per audit is low or avoidance of losses high due either to a sufficiently high steady state of errors in shelf or high demand elasticity or margins.

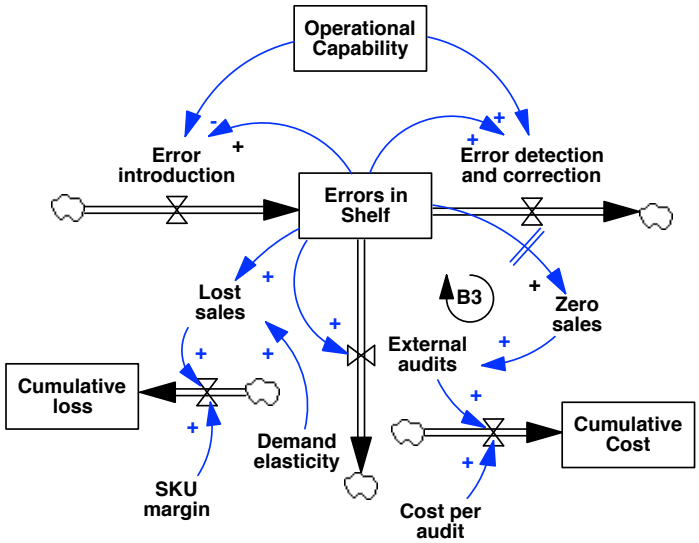

Figure 6. Cost and savings tradeoffs of zero-sales triggered audits

Note how the parameters in Figure 6 (operational capability, demand elasticity, SKU margin, and cost per audit) provide the context for potential generalization of the findings discussed above. Specifically, the retailer's inherent operational capability determines the steady state of the *errors in shelf* stock and

---

explanation, as the retailer's employees could not have identified the special status of these SKUs in their normal operating context (see the explanation on pg. 939 and in footnote 1 of Chuang et al., 2016).

whether there is sufficient room for potential savings, demand elasticity affects potential savings from the elimination of lost sales, and cost per audit partially defines the viability of sustaining the audit program. Note that these three parameters map one-to-one onto the generalizable claims made in the originally published paper (points i, ii, and iii in §4.1). Indeed, the framework could be used by the RSP (or for that matter the retailer itself) to perform a quick feasibility evaluation by assessing the value of these parameters to determine whether deployment of extraordinary audits would be economically sustainable.

Although the research team was not able to articulate the SPC insights described in §4.2 immediately after the intervention, the explanation of the observed dynamics of the transition emerged soon after the data was first analyzed, albeit not as fully developed as presented here. It is interesting to note that despite our enthusiasm for the process explanation, we were asked by the review team to exclude it from the article because they did not serve the purpose of reporting the results of a randomized field experiment.

5 Discussion and conclusion

In this paper, I argue that the empirical motivation of design science is a suitable (required?!) principle of OM theory development and practice. I propose, however, that rather than focus on the design propositions called for in the inaugural editorial in *JOM* (van Aken et al., 2016), we instead explore the use of interventions to make improvements (the ultimate manifestation of design science) that serve to develop and test theories (the ultimate goal of an academic endeavor). Taking as a point of departure the principles of *action research* and explanatory framework of *process theories*, I propose frameworks for leveraging intervention as a mechanism for testing existing theory (how should T be updated as a result of the focal intervention?), and develop theories about organizational and system transformation (what explains the change observed in the transition from S to S'?). I illustrate the application of these two frameworks in the context of a previously published research paper (Chuang et al., 2016).

Having been designed as a randomized field experiment, it is not surprising that the insights from post-hoc analyses were not part of the original publication in the intervention I used to illustrate the application of the frameworks; the research designs and controls and description of the results met the original publication requirements. I believe, however, that these analyses reveal the potential for a well-structured intervention to generate insights useful for further theory development. Specifically, I show how some of the conceptual shortcomings identified in the intervention suggested a new research direction for theoretical improvements in the use of statistical signals to trigger audit corrections, and how generating a structural endogenous explanation for the observed dynamics in the organizational transformation yielded a generalizable process theory for improved performance even in the absence of structural process change.

The question that remains to be answered is whether an intervention will provide sufficient evidence for either of the proposed frameworks to generate useful generalizable theory. As stated above, I do not believe that every intervention will be useful for the type of development proposed here. Indeed, the more successful an intervention is in delivering intended results, the less useful it will be in revealing shortcomings of existing theories or generating insights about the transformation process. To be useful as a source of theoretical improvement, an intervention must yield somewhat surprising (either unexpectedly good or unexpectedly bad) results.[5] Furthermore, an intervention should not be expected to yield insights under both of the proposed frameworks. For instance, the intervention used to illustrate the use of the frameworks (§4) yielded more interesting and compelling insights under the process theory than under the TMS framework. Although directionally significant, the insights gained about T require further work and development. That is, the intervention was useful to signal, but not, of itself, resolve the theoretical shortcomings. Lastly, theory development and testing remain difficult processes that must adhere to well-established methodological protocols and standards of evidence. Neither an intervention nor the data it generates constitute an excuse to shortcut these protocols and standards of evidence. For instance, insights from action research will need to be documented in ways that support recoverability (Checkland and Holwell, 1998), and I suspect most interventions that claim to yield insights about T will need to at least attempt a second iteration in order to demonstrate that the improved theoretical framework is capable of better addressing the issue at hand, essentially an experiment on itself. Interventions that claim to provide evidence for process theories, despite the advantages of multiple testing points cited in §3, face the same difficulties and challenges as single-case research that attempts to justify grounded theory development (Strauss and Corbin, 1990; Yin, 2003). That we have these protocols and standards to guide theory development process is another reason why I argue that the developmental focus should be on theoretical developments over design propositions.

My proposal calls for using interventions as a source of data for theory development and testing. I argue that, by testing the usefulness and applicability of theories in the real world, interventions represent a third avenue for pursuing empirical research, supplementing the use of observations and measurements to develop descriptive theories and experiments to test causal explanations. Despite well-established foundations (e.g., Argyris et al., 1985; Checkland, 1985; Simon, 1981) and our community's unequivocal success in improving real-world situations (INFORMS, 2019), the reflective process of using practice to inform theory has, to date, been largely ignored by the OM community. In a gratifying strange-loop sense (Hofstadter, 2007), the test of whether the proposed frameworks will be useful in supporting intervention-

[5] In the example intervention in §4, the surprising result was that it was possible to maintain a reduced level of *errors in shelf* with an external intervention, notwithstanding that we were not addressing the retailer's operational capabilities.

based research aimed at further development of OM theory is itself an intervention. I look forward to an emergence of intervention-based research in *JOM* that leads to improvements to the theoretical frameworks proposed here and improved process theories of interventions that increase our ability to derive research insights.

References

Akkermans, H., van Oppen, W., Wynstra, F., Voss, C., 2019. Contracting outsourced services with collaborative key performance indicators. Journal of Operations Management 65, 22-47.

Argyris, C., 1976. Increasing Leadership Effectiveness. Wiley, New York.

Argyris, C., 1985. Strategy, Change, and Defensive Routines. Pitman, Boston.

Argyris, C., Putnam, R., Smith, D.M., 1985. Action Science: Concepts, Methods and Skills for Research and Intervention. Jossey-Bass, San Francisco.

Argyris, C., Schön, D.A., 1978. Organizational Learning: A Theory of Action Perspective. Addison-Wesley, Reading, MA.

Bell, J.A., Bell, J.F., 1980. System dynamics and scientific method, in: Randers, J. (Ed), Elements of the System Dynamics Method. Productivity Press, Cambridge, MA, pp. 3-22.

Bell, J.A., Senge, P.M., 1980. Methods for enhancing refutability in system dynamics modeling. TIMS Studies in the Management Sciences 14, 61-73.

Carlile, P.R., Christensen, C.M., 2005. The cycles of theory building in management research, Working Paper #2005-03. Boston University, School of Management, Boston.

Checkland, P.B., 1985. From optimizing to learning: A development of systems thinking for the 1990s. Journal of the Operational Research Society 36, 757-767.

Checkland, P.B., 2010. Researching real-life: Reflections on 30 years of action research. Systems Research and Behavioral Science 27, 129-133.

Checkland, P.B., Holwell, S., 1998. Action research: Its nature and validity. Systemic Practice and Action Research 11, 9-21.

Chuang, H.H.C., Oliva, R., Liu, S., 2016. On-shelf availability, retail performance, and external audits: A field experiment. Production and Operations Management 25, 935-951.

Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design & Analysis Issues for Field Settings. Houghton Miffin, Boston.

Croson, R., Schultz, K., Siemsen, E., Yeo, M., 2013. Behavioral operations: The state of the field. Journal of Operations Management 31, 1-5.

Denyer, D., Tranfield, D., van Aken, J.E., 2008. Developing design propositions through research synthesis. Organization Studies 29, 393-413.

Dooley, K.J., van de Ven, A.H., 1999. Explaining complex organizational dynamics. Organization Science 10, 358-372.

Edmondson, A.C., McManus, S.E., 2007. Methodological fit in management field research. Academy of Management Review 32, 1155-1179.

Fisher, M., 2007. Strengthening the empirical base of operations management. Manufacturing & Service Operations Management 9, 368-382.

Fisher, M.L., Raman, A., 2010. The New Science of Retailing: How Analytics are Transforming the Supply Chain and Improving Performance. Harvard Business Press, Boston.

Foster, M., 1972. An introduction to the theory and practice of action research in work organizations. Human Relations 25, 529-556.

Gallien, J., Graves, S.C., Scheller-Wolf, A., 2015. OM Forum—Practice-based research in operations management: What it is, why do it, related challenges, and how to overcome them. Manufacturing & Service Operations Management 18, 5-14.

Gino, F., Pisano, G., 2008. Towards a theory of behavioral operations. Manufacturing & Service Operations Management 10, 676-691.

Greene, W.H., 1997. Econometric Analysis, 3rd ed. Prentice-Hall, Upper Saddle River, NJ.

Hofstadter, D., 2007. I am a Strange Loop. Basic Books, New York.

Holmström, J., Ketokivi, M., Hameri, A.P., 2009. Bridging practice and theory: A design science approach. Decision Sciences 40, 65-87.

INFORMS, 2019. O.R. & Analytics Impact Everything. https://www.informs.org/Impact, Last accessed: May 9, 2019.

Kaplan, A., 1964. The Conduct of Inquiry. Chandler, San Francisco.

Keating, E.K., Oliva, R., Repenning, N., Rockart, S.F., Sterman, J.D., 1999. Overcoming the improvement paradox. European Management Journal 17, 120-134.

Kolb, D.A., 2014. Experiential Learning: Experience as the Source of Learning and Development, 2nd ed. Pearson FT Press, Upper Saddle River, NJ.

Krimerman, L.I. (Ed), 1969. The Nature & Scope of Social Science: A Critical Anthology. Meredith Corporation, New York.

Langley, A., 1999. Strategies for theorizing from process data. Academy of Management Review 24, 691-710.

Langley, A., Smallman, C., Troukas, H., van de Ven, A.H., 2013. Process studies of change in organization and management: Unveiling temporality, activity and flow. Academy of Management Journal 56, 1-13.

Lave, C.A., March, J.G., 1975. An Introduction to Models in the Social Sciences. Harper & Row, New York.

Lewin, K., 1946. Action research and minority problems. Journal of Social Issues 2, 34-46.

Lewis-Beck, M., Bryman, A.E., Liao, T.F., 2003. The Sage Encyclopedia of Social Science Research Methods. Sage, Thousand Oaks, CA.

Merton, R.K., 1968. Social Theory and Social Structure, 1968 Enlarged ed. Free Press, New York.

Mohr, L.B., 1982. Explaining Organizational Behavior. Jossey-Bass, San Francisco.

Monge, P.P., 1990. Theoretical and analytical issues in studying organizational processes. Organization Science 1, 406-430.

Montgomery, D.C., 1991. Introduction to Statistical Quality Control, 2nd ed. Wiley, New York.

Morrison, J.B., Oliva, R., 2019. Integration of behavioral and operational elements through System Dynamics, in: Donohue, K., Katok, E., Leider, S. (Eds), The Handbook of Behavioral Operations. Wiley, New York, pp. 287-321.

Oliva, R., 1988. A systems study of 'Departmental Purpose Analysis' in British Telecom (MA Thesis), Department of Systems and Information Management. Lancaster University, Lancaster, UK.

Oliva, R., 1994. A framework for reflecting in methodology, System Dynamics Group Working Paper D-4458-2. MIT Sloan School of Management, Cambridge, MA.

Oliva, R., 2001. Tradeoffs in responses to work pressure in the service industry. California Management Review 43, 26-43.

Oliva, R., 2003. Model calibration as a testing strategy for system dynamics models. European Journal of Operational Research 151, 552-568.

Oliva, R., Sterman, J.D., 2001. Cutting corners and working overtime: Quality erosion in the service industry. Management Science 47, 894-914.

Oliva, R., Watson, N.H., 2009. Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. Production and Operations Management 18, 138-151.

Oliva, R., Watson, N.H., 2011. Cross-functional alignment in supply chain planning: A case study of sales and operations planning. Journal of Operations Management 29, 434-448.

Pawson, R., 2002. Evidence-based policy: In search of a method. Evaluation 8, 157-181.

Pentland, B.P., 1999. Building process theory with narrative: From description to explanation. Academy of Management Review 24, 711-724.

Pettigrew, A.M., 1990. Longitudinal field research on change: Theory and practice. Organization Science 1, 267-292.

Poole, M.S., van de Ven, A.H., Dooley, K., Holmes, M.E., 2000. Organizational Change and Innovation Processes: Theory and Methods for Research. Oxford University Press, Oxford.

Rai, A., 2017. Editor's comments: Diversity of design science research. MIS Quarterly 41, iii-xviii.

Rapoport, R.N., 1970. Three dilemmas in action research. Human Relations 23, 499-513.

Romme, A.G.L., 2003. Making a difference: Organization as design. Organization Science 14, 558-573.

Runkel, P.J., Runkel, M., 1984. A guide to usage for writers and students in the social sciences. Rowman & Allanheld, Totowa, NJ.

Sastry, M.A., 1997. Problems and paradoxes in a model of punctuated organizational change. Administrative Science Quarterly 42, 237-275.

Schein, E.H., 1987. The Clinical Perspective in Fieldwork. Sage, Newbury Park, CA.

Schein, E.H., 1993. Legitimating clinical research in the study of organizational culture. Journal of Counseling & Development 71, 703-708.

Schneiderman, A.M., 1988. Setting quality goals. Quality Progress 21, 51-57.

Simon, H.A., 1981. The Sciences of the Artificial, 2nd ed. MIT Press, Cambridge, MA.

Sterman, J.D., 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. Irwin McGraw-Hill, Boston.

Sterman, J.D., Oliva, R., Linderman, K., Bendoly, E., 2015. System dynamics perspectives and modeling opportunities for research in operations management. Journal of Operations Management 39-40, 1-5.

Sterman, J.D., Repenning, N.P., Kofman, F., 1997. Unanticipated side effects of successful quality programs: Exploring a paradox of organizational improvement. Management Science 43, 503-521.

Strauss, A., Corbin, J., 1990. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Sage, Newbury Park, CA.

Sutherland, J.W., 1975. Systems; Analysis, Administration and Architecture. van Nostrand, New York.

Tang, C.S., 2015. OM forum—Making OM research more relevant: "why?" and "how?". Manufacturing & Service Operations Management 18, 178-183.

Toffel, M.W., 2016. Enhancing the practical relevance of research. Production and Operations Management 25, 1493-1505.

Tushman, M.L., Romanelli, E., 1985. Organizational evolution: A metamorphosis model of convergence and reorientation. Research in Organizational Behavior 7, 171-222.

van Aken, J.E., 2004. Management research based on the paradigm of the design sciences: The quest for field-tested and grounded technological rules. Journal of Management Studies 41, 219-246.

van Aken, J.E., Chandrasekaran, A., Halman, J., 2016. Conducting and publishing design science research: Inaugural essay of the design science department of the Journal of Operations Management. Journal of Operations Management 47-48, 1-8.

van Aken, J.E., Romme, A.G.L., 2012. A design science approach to evidence-based management, in: Rousseau, D. (Ed), The Oxford Handbook of Evidence-Based Management. Oxford University Press, Oxford, pp. 43-57.

van de Ven, A.H., Engleman, R.M., 2004. Event- and outcome-driven explanations of entrepreneurship. Journal of Business Venturing 19, 323-358.

van de Ven, A.H., Poole, M.S., 1995. Explaining development and change in organizations. Academy of Management Review 20, 510-540.

Weick, K.E., 1987. Theorizing about organizational communication, in: Jablin, F.M., Putnam, L.L., Roberts, K.H., Porter, L.W. (Eds), Handbook of Organizational Communication. Sage, Newbury Park, CA, pp. 122-135.

Weick, K.E., 1989. Theory construction as disciplined imagination. Academy of Management Review 14, 516-531.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

Yin, R.K., 2003. Applications of Case Study Research, 2nd ed. Sage, Thousand Oaks, CA.