POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 30, No. 3, March 2021, pp. 802–814
ISSN 1059-1478 | EISSN 1937-5956 | 21 | 3003 | 0802

DOI 10.1111/poms.13318

The Case for a Unified Science of Operations

Mark L. Spearman
Project Production Institute, 2229 San Felipe Street,Suite 1325, Houston, Texas 77019, USA, mspearman@spsinc.net

Wallace J. Hopp*
University of Michigan, Ann Arbor, Michigan 48118, USA, whopp@umich.edu

Throughout its history, the industrial engineering/operations management (IE/OM) field has relied heavily on axiomatic models and empirical studies of individual systems. But, unlike other engineering and management disciplines, it lacks a clear foundation in a descriptive science. Exceptional results like the famous "bullwhip effect" paper by Lee et al. (1997) hint at the powerful potential for a descriptive science of operations. But the very fact that such works are exceptional suggests that they are held to a high bar in the publishing process. This may be symptomatic of the cultural norms that have prevented our field from producing a rigorous scientific foundation. In this study, we make a case for why developing a unified science of operations is essential for IE/OM education, practice and research. We provide examples and a tentative framework to illustrate what such a science might look like and use this framework to generate a testable hypothesis about a powerful relationship between variability buffers in and operations system. We conclude with suggestions of measures we can take collectively to promote development of the science of operations.

Key words: operations science; variability; buffers; production; services
History: Received: September 2020; Accepted: October 2020 by Chris Tang, after 2 revisions.

1. Prologue

Industrial engineering is unique among engineering disciplines in lacking a rigorous descriptive science foundation. Students of all other engineering disciplines take engineering science courses, such as statics and dynamics for mechanical engineers and electricity and optics for electrical engineers. But in place of engineering science courses, industrial engineering students take mathematics courses on probability, statistics and optimization. Unlike engineering science courses that describe the behavior of real-world phenomena, mathematics courses provide methodologies for analyzing these behaviors.

An analogous situation exists in the business disciplines, most of which are based on social sciences. For example, accounting and finance are based on economics, organizational management is based on psychology and sociology, and (behavioral) marketing is based on psychology. However, operations management has no clear basis in either the social or physical sciences.

Why is this? We suspect the lack of a descriptive science basis for industrial engineering and operations management (IE/OM) stems from the historical fact that operations problems were approached right from the start in a purely prescriptive manner. Unlike fields such as mechanics, electricity and psychology, which were concerned with pure discovery as well as pragmatic application, the field of operations was narrowly concerned with application. As a result, IE/OM research consists almost entirely of analyses of specific practical problems. Such studies do not inspire the search for general descriptions and unified frameworks.

This is not to say, however, that IE/OM research lacks rigor. But in place of rigor in identifying universal behaviors (e.g., Ohm's Law in electrical science) the rigor in IE/OM has been devoted to developing general purpose methodologies and applying them to specific problems. In this light, the blurring of IE/OM into operations research in the 1960s and 70s makes perfect sense. Tools like linear programming, queueing theory, and decision analysis are highly adaptable to all manner of operations problems. But methodologies by themselves cannot tell us how operations systems behave.

Of course, IE/OM scholars have produced descriptive results about operations systems. Probably the most famous is Little's Law (Little 1961). This relationship is central to all treatments of flows because it relates flow rate, flow time and the number of entities in the system. Indeed, we described it as the "F = ma" of Factory Physics (Hopp and Spearman 2008, p. 240). However, unlike an empirical relationship drawn from observation of the real world, such as Ohm's Law, Little's Law follows directly from the assumptions of a queueing system in steady state. As

Little (1992) himself pointed out, Little's Law is actually "Little's Tautology".

Many classic results in the IE/OM canon are axiomatic relationships in the same vein as Little's Law. For example, despite the claim by Ford W. Harris (1913, p. 135) that "Experience has shown one manager a way to determine the economical size of lots", the EOQ formula follows directly (with a bit of calculus) from the assumptions of a fixed setup cost and linear holding cost. How well an axiomatic model fits reality depends on the extent to which the assumptions are satisfied, which is something that can only be determined empirically. Unfortunately, attempts to validate assumptions and to characterize the impact of deviations from them on the predictive power of axiomatic models are rare in the IE/OM field.

Of course, not all IE/OM research is based on axiomatic models. The majority of papers published in the *Journal of Operations Management* have long been empirical in nature, and the proportion of papers in other IE/OM journals that rely on empirical data has been growing in recent years. For example, a survey of papers published since 2002 in *Management Science* suggests that the fraction of operations papers with data has grown, although it still lags that of papers in other business disciplines such as finance, marketing, and information systems (Hopp and Simchi-Levi 2020). But most of these empirical papers analyze specific problems rather than seek out universal behaviors of operations systems. Consequently, few results from empirical IE/OM research have found their way into mainstream texts on production and operations management (e.g., Buffa 1987, Schmenner 1990, etc.), which have instead emphasized axiomatic models and general analysis techniques.

A notable exception of particular relevance in this issue honoring Hau Lee is the famous "bullwhip effect" paper (Lee, Padmanabhan and Whang 1997), which had been cited well over 5,000 times according to Google and has inspired many follow-on papers. By starting with descriptive data, as Newton and Kepler did when they based their models of planetary motion on detailed observations by Tycho Brahe, the authors were following in the tradition of the very best descriptive science. Like Newton and Kepler, they described a phenomenon that had been noted previously but was not well understood and proposed a parsimonious model to explain it. By doing this, they showed that the bullwhip phenomenon was a predictable consequence of rational behavior, that it had identifiable causes, and that specific policies could be used to mitigate the effect and its costly consequences.

This is a fine piece of work that deserves all the attention it has received and belongs in any future science of operations. Moreover, because it was so visible, it is plausible that this data-driven study helped inspire the rise of data-driven research in the years following its publication. However, the very excellence of this study is also a potential cautionary note. Would the bullwhip effect paper have been published had it only documented the behavior via careful data collection in the manner of Brahe? Would it have been published, had it only proposed a model that captured the behavior in a descriptive fashion in the manner of Kepler and Newton? We suspect not. Given the heavy emphasis on prescriptive results in the IE/OM field, the managerial conclusions about how to mitigate the bullwhip effect were probably necessary for acceptance of the paper.

But this is an exceedingly high bar to which other disciplines do not hold their authors. In most other fields, publishing well-documented research that describes fundamental behaviors in the physical or social worlds is standard procedure. It is such work that has produced the engineering science basis for engineering disciplines and the social science basis for management disciplines that IE/OM lacks.

The need for and lack of a scientific basis for IE/OM has been noted repeatedly in our history as a field. For example, Smiddy and Naum (1954, p. 2) proclaimed in the very first article published in *Management Science*:

"What has been, and still is, sought in all these recurring efforts to define and develop a true "Science of Managing" are, rather, those kinds of fundamental principles which are the essential scientific foundation of all generalization based on classified observations and which give meaning, accuracy and dependability to formulation of rules of action or policies, that, given a particular set of connections, can be used or applied as guides with confidence in their effectiveness."

Although Smiddy and Naum were confident that a scientific basis for all of management was within grasp in 1954, nearly a half century later, Bertrand and Fransoo, (2002, p. 244) bemoaned the lack of progress even within the narrow confines of operations management:

"Up to now OM research has not been very successful in developing explanatory or predictive scientific models of operational processes, that is, models that can be used to explain or predict the output or performance of the process as a function of process characteristics, process states and inputs to the process. This is a major roadblock for the development of the field, since the development of effective methods to improve performance assumes that scientific knowledge of the process is available."

Some individuals, including us, have made preliminary efforts to synthesize OM results into some kind of scientific structure. However, so far, these have either accumulated research results without a unifying framework (e.g., Schmenner and Swink 1998) or have provided limited frameworks for a specific industry (e.g., Hopp and Spearman 2008 for manufacturing, Hopp 2011 for supply chains, and Hopp and Lovejoy 2012 for hospitals). While these may represent precursors to a science of operations, they fall well short of a unified science that could serve as the descriptive foundation of IE/OM.

The reality is that a scientific field can only be brought into being by a community of scholars. Therefore, rather than simply make another plea for a science of operations, we first detail the need for such a science in IE/OM education, practice, and research. We then describe the scientific process with particular emphasis on the falsifiability standard and illustrate it by subjecting some well-known operations models to empirical tests. Next we sketch out a unified framework for operations science and use it to generate a testable hypothesis about the relationship between inventory buffers in operations systems. Finally, we conclude with suggestions of measures we can take collectively to promote development of the science of operations.

1.1. Wht Do We Need a Science of Operations?

The lack of a coherent science of how operations systems behave is an impediment for IE/OM education, practice, and research as we describe below.

When an electrical or mechanical engineering student graduates and goes into the workplace, he/she can lean on engineering science. For example, an electrical engineer knows from Ohm's Law that the size of the fuse needed to protect a device increases proportionally in the wattage of the device. A mechanical engineer may use a finite element tool to design components that has statics and dynamics principles built into it. But an IE/OM professional has only general-purpose analysis tools and axiomatic models of simplified situations to apply to complex real-world problems that often bear little resemblance to the stylized scenarios they saw in their classes.

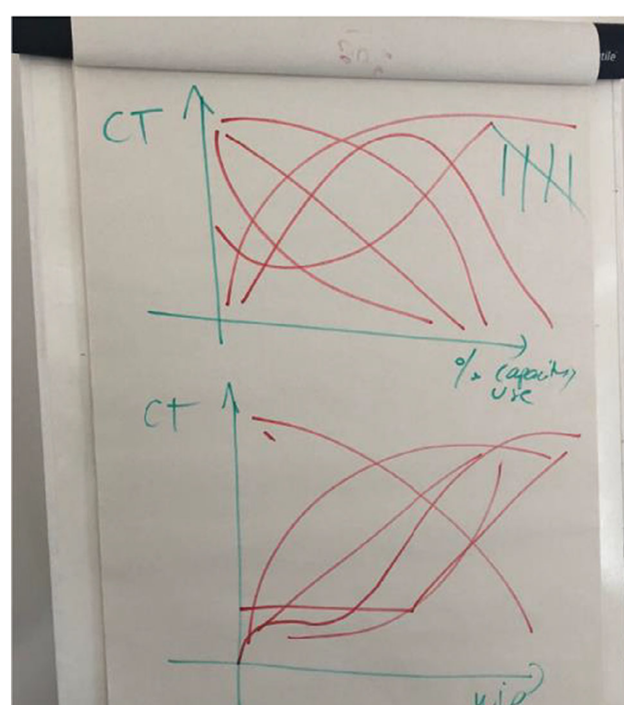
Hence, it is not at all surprising that we get calls from students in internships or first jobs asking things like: "Can I apply the EOQ model to this situation?", "Where do I get data for a Newsvendor model?" "Can I really compute waiting time using a queueing model?". Even more common are students who just dispense with the models and tools from their school days and adopt ad hoc methods, often implemented in spreadsheets, to guide decision making.

This is incredibly ineffective because the lack of a descriptive science of operations has also left operations professionals with poor intuition about basic relationships. For example, in our workshops, we have asked many experienced people in roles ranging from shop floor positions to executive management in automotive, semiconductor, healthcare, construction, energy, and other industries, to describe the relationship between utilization and cycle time and the relationship between cycle time and WIP.

Figure 1 shows the responses from one such session and is typical. Not only are most of the descriptions incorrect, but also the range of conceptual disagreement is simply staggering. Some think that cycle time goes up with utilization while others think it goes down. Some think that cycle time and WIP move in tandem, while others think they move in opposition. Unlike the sharp intuition of an electrical engineer making a fuse size decision, an IE/OM engineer making a capacity decision often does so with a very confused sense of the consequences.

This lack of intuition about basic relationships impacts more than individual decisions. Firms seeking to improve their competitiveness are forced to develop methods by trial and error or to copy methods of other firms without understanding how and why they work. In the former category, Taiichi Ohno, the principle architect of the Toyota Production System, used the trial-and-error approach over a period of more than 30 years (Ohno, 1988). But because most companies do not have such a single-minded genius or 30 years, they fall into the latter category, pursuing methods such as Lean and LeanSigma with very

Figure 1 Responses to Questions Regarding Basic Relationships [Color figure can be viewed at wileyonlinelibrary.com]



mixed success. Simplistic mottos such as "drain the water to see the rocks," "do only what creates value for the customer," and "deliver just in time," are poor substitutes for clear principles and intuition.

If the lack of commonly accepted, empirically validated principles is bad for the practice of IE/OR, it is even worse for research in these fields. The reason is that without explicit grounding in real-world behavior, research will be scattered, stagnant, or wrong.

Scattered research results because related problems are not seen as such. For example, the activity of a doctor treating a patient may be formally identical in some ways to the activity of a machine processing a part. But if this is not recognized, then research on one problem will not be leveraged in research on the other. The mechanics analogy would be a separate treatment of the dynamics of bullets and baseballs.

Stagnant research results because of the lack of attempts to refute existing models. Empirical anomalies are often the drivers of scientific progress. For example, the inability of Maxwell's equations to explain the photoelectric effect was an impetus for quantum mechanics. For many years, the IE/OR fields have focused on better solutions to established problems (e.g., the Traveling Salesman Problem), rather than reshaping models to better fit reality. Recent empirical studies of human behavior in operations settings (see Donohue et al., 2020) offer many opportunities to stimulate research progress. But since empirical testing and refutation of operations models is not an accepted practice in the IE/OM research community, we are unlikely to leverage these to their full potential.

Wrong research results from using the wrong building blocks. Since many fundamental IE/OR results, such as the EOQ and Newsvendor models, are axiomatic models without empirical foundations, many papers have built invalid representations into their models right from the start. For example, the seminal paper by Wagner and Whitin (1958) led to an avalanche of follow-on papers that continues to this day (see (Brahimi et al., 2017) for a recent survey providing 292 references). Many of these papers offer an easier algorithm to perform dynamic lot sizing while maintaining the "Wagner-Whitin" property described in the original text. While these papers do indeed minimize the costs of the proposed mathematical formulation, they have had little relevance to actual problems facing factories where lot sizes are determined more by capacity considerations and setup times than by setup costs. Moreover, the Wagner-Whitin property becomes much less valuable whenever there is variability in production and random yield losses. Other examples of overly simplistic building blocks leading to questionable representations of complex systems include one and two machine scheduling models, deterministic job-shop scheduling models, material requirements planning, Monte Carlo PERT networks, and supply chain models premised on inaccurate models of individual firm behavior.

2. What is Operations Science?

To define operations science, we must first define science. A typical starting point is something like, "The intellectual and practical activity encompassing the systematic study of the structure and behavior of the natural and social world through observation and experiment" (Google dictionary).

Operations science involves some elements of the natural world, such as physically moving materials from one place to another. But, since organizations, and all the operations within them, are human constructs, it is very much part of the social world.

This definition immediately makes it clear that simply collecting data is not science. The "systematic study of the structure" of science implies some form of tentative paradigm or for interpreting data and organizing our understanding of it. Two excellent examples of paradigms are the Standard Model of particle physics and the genome of biology. Kuhn (1962) famously described a scientific field as progressing through a series of punctuated equilibria separated by paradigmatic revolutions in between which scientists engage in the "normal science" of testing consequences and filling in details.

We should note however that scientific paradigms are not without controversy and there is a great deal of literature (see, e.g., Fuller 2004) debating the ideas of Kuhn (1962) and Popper (1963) regarding scientific paradigms. We will duck this controversy by interpreting Kuhn's theories as an attempt to describe how science is *actually done* and Popper's theories as an effort to define the way science *should be done*. From his sociological perspective, Kuhn described "normal science" as "puzzle solving" needed to fill in the gaps in the current scientific paradigm. But over time results would accumulate that could not be explained by that paradigm and would force a "paradigm shift" or revolution. From his philosophy of science perspective, Popper focused on the proper method of experimentation for validating a paradigm. He used the logical "fallacy of affirming the consequent" to point out that no amount of experimentation can ever definitively determine a cause of an effect and so science is more productive when it seeks refutations to its conjectures instead of confirmations.

The fallacy with which Popper is concerned goes as follows: Theory A predicts observation O, we observe O, and conclude that theory A is true. But there could be many other possible causes for observation O and it is impossible to know which is correct. Although this could make science seem hopeless, Isaac Asimov found it to be exhilarating and explained, "The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!' but 'That's funny ...'". That is, it is not the big confirmations that push science forward but rather the small refutations that force a revision of the paradigm.

In Popper's terms, normal science consists of activities do not result in refutations and paradigm shift is the result of significant refutations. Of course, not all refutations result in wholesale rejection of an entire paradigm. Instead, most refutations lead to refinements and modifications of a paradigm. For example, when Uranus was discovered, its orbit did not appear to obey Newton's Law of Gravity. Instead of developing a new theory, a search was made for what could have caused the perturbation and Neptune was discovered. However, when it was found that Mercury's orbit failed to obey Newton and after a search for a planet inside the orbit of Mercury turned up negative, Newton's Law entered a "crisis" stage (Kuhn's term). And when Einstein's theory of General Relativity predicted the observed perturbations, it put the final nail in Newton's coffin.

The fundamental role of refutations in driving scientific progress led Popper to conclude that the necessary and sufficient criterion of science is *falsifiability*. That is, an empirical science must yield testable hypotheses. For example, Einstein's general theory of relativity also predicted that starlight is deflected by a small but measurable amount by the gravity of the sun, a phenomenon that, unlike Mercury's orbit, had *not* been observed. A single valid observation contradicting this prediction would have rendered general relativity false. However, when careful measurements of star positions during a solar eclipse in 1919 confirmed the predicted deflection, Einstein and his theory were lionized by the public. Nevertheless, relativity was not proven correct. It was merely not disproven. Such is the nature of empirical science.

In contrast, Freud's theory that human personality is made up of the id, ego and superego is inherently untestable. What experiment could disprove it? Popper considered Freudian psychology to be pseudoscience because it was based on observations and was presented as "scientific" but failed to meet the falsifiability standard.

Note, however, that not all nonscience is pseudoscience. Mathematics, philosophy, mythology, religion, and metaphysics are examples of systems of thought that make claims not subject to empirical refutation. But these self-contained systems are internally consistent. That is, their claims follow from the assumptions that define them. They do not purport to be sciences and therefore cannot be called pseudosciences.

Furthermore, nonscience can be useful. Mathematics is not a science but is used to great advantage within most sciences. Philosophy and religion have been powerful forces in shaping human society and can help individuals guide their daily lives. Furthermore, one can state scientific hypotheses about nonscientific, or even pseudoscientific, practices. For example, a hypothesis that individuals who meditate daily are less prone to committing acts of violence than individuals who do not is certainly testable. It would require some logic and mathematics to evaluate the causality of this claim, but a valid negative result would render it false. Hence it meets the falsifiability standard for science.

With this falsifiability standard in mind, the practice of science must involve making empirical observations, classifying these observations, postulating relationships in the form of a testable hypothesis, experimentally testing the hypothesis, and iteratively modifying the hypothesis until there are no more refutations. At this point, we tentatively accept the hypothesis and call it a theory or a law depending on our confidence in the results. During the process, we may develop any number of models to describe the relationships and to facilitate meaningful experiments. Such models are the mark of a mature field that has advanced beyond the stages of observation and classification. Models also eventually come to describe the paradigm of the field and, by noting what they fail to explain, can guide research for expanding our understanding.

From this perspective, IE/OR looks like a field that has developed categories and many uncoordinated models, but which has few basic empirical relationships and certainly has no "standard model" or framework. Moreover, while much of IE/OR research is nonscience in the form of mathematics, much of IE/OR practice is nonscience in the form of pseudoscience. For example, the Lean Institute defines the Five Principles of Lean as:

1. Identify value.
2. Map the value stream.
3. Create flow.
4. Establish pull.
5. Seek perfection.

Obviously, there are no testable hypotheses inherent in the way these are stated. They sound like religious commandments. Indeed, Lean is sometimes presented as a philosophical or almost religious framework of beliefs. At the same time, Lean is sometimes presented as a fact-based, scientific approach to operations management. Hence, it seems fair to call it pseudoscience. As we noted above, it is entirely possible for nonscientific systems of beliefs to have positive impacts on human behavior. Whether a management system like lean is effective in achieving the goals of a business is a question that requires scientific analysis to answer.

It is unrealistic to expect that complex socioeconomic systems like businesses should be as amenable to scientific inquiry as the physical phenomena studied in physics. But this does not mean that a useful science is impossible. Economics, which attempts to characterize the most complex socioeconomic systems on earth, relies heavily on models guided by empirical studies. If they can do it, we can too.

3. Models and Experiments

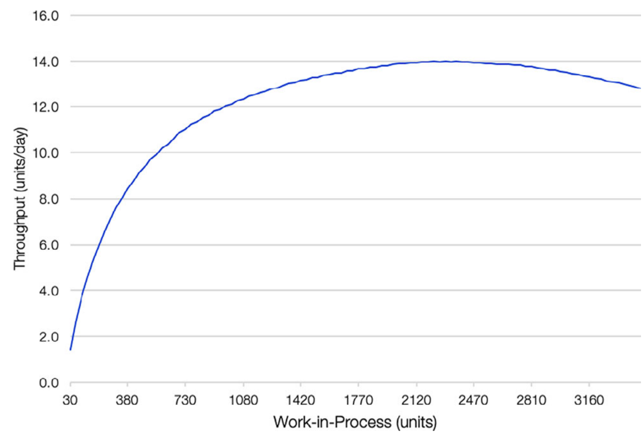
To illustrate how the above "conjecture and refutation" process could work to advance the empirical science of operations, we offer some examples of OM problems that have data and provide hypotheses that could possibly be refuted.

First, we consider a classic axiomatic model of a closed queueing network made up of a series of single resource stages. We assume each stage has random process times that are iid and independent of the amount of WIP in the system. This queueing system has been well studied and the relationships between various performance metrics have been characterized. In particular, Adan and Van der Wal (1989) proved that throughput increase monotonically in the WIP level.

As an empirical test of the validity of this model, we found a system that matches the characteristics of this system closely. This was a large capital construction project that managed a multi-stage workflow by releasing standardized work packages into the system in a manner that held the total amount of work in the system constant. As an experiment, the manager of this system decided to reduce the cap on the work level. To our surprise (but not his) throughput increased in violation of the Adnan and Van der Wal result. Of course, since throughput must go to zero as WIP approaches zero, we know that if we reduce WIP too much throughput will eventually decrease. Hence, based on the empirical observations of this system, the throughput curve is not monotonically increasing but is instead something like the curve shown in Figure 2.

Obviously, the closed queuing network model failed to capture the behavior of this real-world system. Since it is an axiomatic model, the problem is not the analysis by Adan and Van der Wal. Instead, there must be something wrong with the assumptions. The assumption that seems most suspect is that the process times are independent of the amount of WIP in the system. As the manager suspected, it could be the case when WIP gets too high, it takes more time to move it, sort it and find the next unit to work on. This WIP-dependent work consumes some of the available capacity and hence slows down the process.

Figure 2 **Throughput Degradation with Increases in WIP [Color figure can be viewed at wileyonlinelibrary.com]**



This level of refutation is not sufficient to reject the closed queuing network altogether as a representation of workflows like the one in the capital construction project. Instead, it is analogous to the manner in which friction alters the physics of a projectile (i.e., in a vacuum a projectile follows a parabolic arc, but with quadratic friction it follows an asymmetric arc). To understand and optimize a workflow with WIP-dependent process times, we would need to collect data to describe this effect and find a modified queueing model that incorporates it.

For our second empirical test, we consider the ubiquitous Poisson model of demand processes. This model is very widely used but rarely tested (a notable exception being Inman (1999)), so we lack understanding of where it represents reality and where it does not. We make our empirical test by using demand data from suppliers to a manufacturer of electrical surge protection equipment. These data include both the date of the demand and the quantity. We aggregate the data into different sized time buckets (i.e., 1 day, 2 days, etc.). If demands are iid but not necessarily Poisson, there will be a linear relationship between the variance of the demands and the bucket size. This can be modeled as a renewal process with the variance of demand in a time bucket equal to the variance of the number of renewals in an equivalent time interval (Cox 1962). A renewal process is a counting process where the times between renewals are independent and identically distributed random variables. An equilibrium renewal process is one that has been running sufficiently long that the initial conditions can be ignored. Cox derives the variance of the number of renewals in an interval of length t of an equilibrium process as,

$$\mathrm{var}(N_t) = \frac{\sigma^2 t}{\tau^3} + \left(\frac{1}{6} + \frac{\sigma^4}{2\tau^4} - \frac{\tau_3}{3\tau^3}\right),$$

where N_t is the number of renewals (demands) and τ is the mean, σ^2 is the variance and τ_3 is the third central moment of the time between renewals. Substituting in the appropriate values for a Poisson demand process yields the much simpler well-known expression,

$$\mathrm{var}(N_t) = \frac{t}{\tau} = \lambda t$$

Thus, if the demand is Poisson, we should see a linear relationship with zero intercept between the size of the time bucket and the total demand. Any other pattern would represent a refutation of the Poisson model.

Figure 3 shows the relationship for the demand for part 3068. There is a clear linear correlation with a r^2 factor of 0.996. The probability of an r^2 factor greater than 0.823 for 12 data elements is only 0.001 so the probability of having one that is 0.996 and uncorrelated is extremely low. But the intercept is far from being zero. Of course, it is possible that part 3068 is unique. So, we performed the same test for a large number of other parts sourced by the manufacturer and found similar results (Figure 3).

Given these results we can reject the hypothesis that the demand is Poisson. Moreover, it is easy to show that a compound Poisson process (one with Poisson demand instances but with iid demand sizes) will also have a linear relationship with zero intercept. Thus, we can reject the often-made assumptions that the demand is either an ordinary or a compound Poisson process. This outcome highlights the importance of making empirical tests of modeling assumptions, both to ensure the validity of the model for its proposed purpose and to identify opportunities for improving or extending our modeling capabilities. The fact that very few IE/OM papers make such empirical tests is an obstacle to progress in our field.

4. Toward an Operations Science Paradigm

As we noted earlier, empirical testing of results is necessary for science, but it is not sufficient. To become a legitimate field of scientific study an area also requires a widely accepted paradigm, which consists of the theorems, principles, and models that describe the generalizable features of the field. Because IE/OM has progressed largely as a series of efforts to understand and improve individual systems, our paradigm, if we can be said to have one, is highly fragmented and ill-defined.

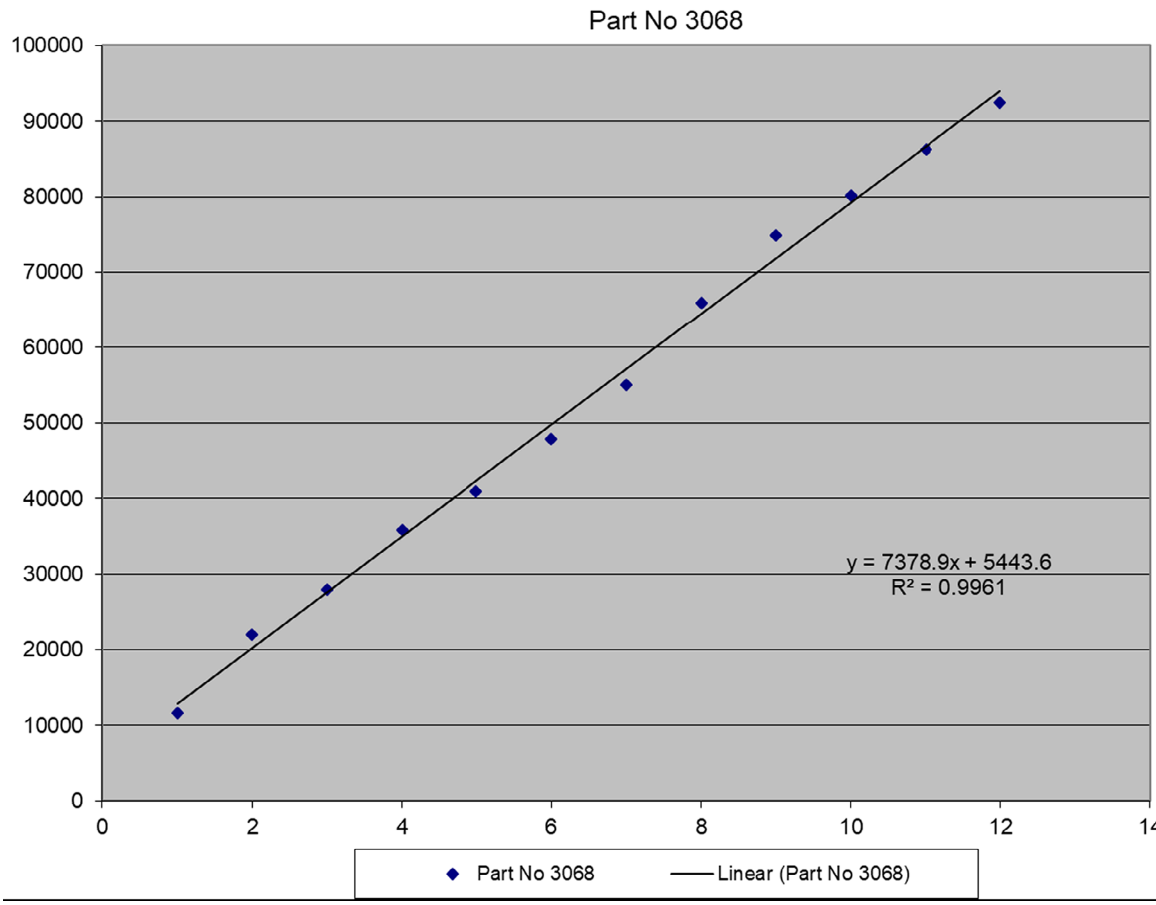
Given the rudimentary state of descriptive research in IE/OM, it is unreasonable to expect a paradigm on a par with the Standard Model of particle physics, which is rigorous, precise and accurate in its predictions. A more realistic example might be the paradigm of neoclassical microeconomics, which consists of many different types of axiomatic and empirical models of economic activity. But in neoclassical microeconomics, these models were united by their focus on supply and demand as the key drivers of economic output and prices. They also relied consistently on rational choice theory to model the behavior of the people in economic systems by assuming that they act rationally on the basis of full information to maximize their utility. In recent years this assumption has been subject to widespread refutations from behavioral economists who have shown that people frequently behave in irrational but predictable ways (see, e.g., the following books by Nobel Prize winning behavioral economists Kahneman (2011) and Thaler (2015)). Nevertheless, the enormous body of work describing economic behavior under the rational choice assumption prepared the field for the behavioral refutations and possibly a paradigm shift that will make microeconomics more accurate in making predictions and informing decision making.

Like economics, IE/OM focuses on supply and demand. Most textbook definitions of operations describe the discipline as focusing on delivery of goods and services to customers. But unlike economics, which abstracts away the details of the processes used to create the supplies that meet demands, IE/OM is intensely interested in these specifics. To capture this central focus, we define the core activity of all operations systems as follows:

Definition (Operation.): *An operation is an act that utilizes resources to transform one or more attributes of an entity or set of entities into some good or service that is required to satisfy some external demand.*

Note that by defining resources, transformation and demand broadly enough, this definition can encompass a huge range of human activities. Manufacturing, service, health care, finance, information technology, consulting government, non-profit, military, and religious organizations all make use of some kind of resources to meet some kind of demand. For example, a manufacturer uses labor, machines, and information to transform components, consumable materials, and energy into physical products. A retailer uses facilities, labor and information to transform the location of goods from manufacturers and distributors to customers. A primary care physician uses tools, supplies, information, and creativity to

Figure 3 Demand Data for Part 3068 [Color figure can be viewed at wileyonlinelibrary.com]



transform data and observations into diagnostic and treatment advice, and so on. At this level of abstraction, virtually every organization makes use of operations to carry out their mission. Certainly, the vast majority of research papers in IE/OM journals deal with operations that fit this general definition.

Importantly, operations are a means to an end. Unlike projectiles in motion or electrons in a current, operations have a purpose. But the operations function in an organization takes this purpose as an input rather than as a decision. This means that the values and preferences of both the customers and the organization are taken as given. On the demand side, this means that what customers want is simply the underlying driver of demand. Members of a religious organization may be seeking spiritual guidance, while consumers of Hollywood blockbusters may be seeking cartoonish violence. Regardless, the role of operations is to deliver on the expectations of customers. Similarly, on the supply side, a manufacturing company may be concerned about profits, a non-profit may be concerned about the welfare of stakeholders, and a public agency (e.g., NASA) may be concerned about a mission-based goal. Operations scholars and practitioners may have personal opinions about the values of customers or organizations, but these are outside the scope of the field.

However, operations specialists are concerned with metrics for measuring how well an operation serves its purpose, once that purpose has been defined. On the demand side, metrics of concern to customers are often grouped into cost, quality, time, and variety. But these are broad headings that can encompass many dimensions. For example, customers for a given product or service might be concerned about sustainability, which we might lump into the quality category. But we could also think of carbon emissions or pollution as a cost to the customer, albeit one that is felt by the public as a whole rather than by the individual customer alone.

On the supply side, metrics of concern to managers of operations systems include internal measures that assure satisfaction of customer needs (e.g., quality, responsiveness) as well as metrics that relate to financial performance of the firm (e.g., throughput, efficiency, inventory). Because different kinds of organizations make different types of strategic value propositions to their customers, performance metrics are one place where operations will differ by sector or industry.

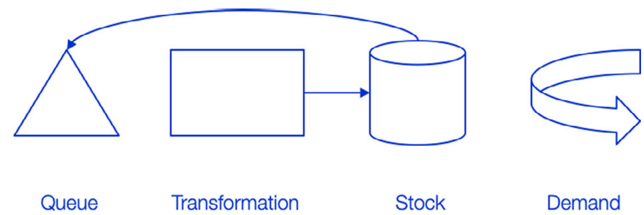
Nevertheless, we can think of this broad description of a purpose-driven operation as a general framework for describing specific operations settings. As such, it plays an analogous role in descriptive research that general purpose analytic tools, such as linear programming and regression analysis, play in prescriptive research. Obviously, the details of what constitutes resources, transformation, and demand, as well as the overarching purpose for the operations, will differ greatly across organizations. The benefit of starting from a general framework, rather than from the particulars of a specific system is that it encourages us to determine where differences matter and where they do not. It probably makes good sense to treat supply chains, retail systems and health care organizations as separate subfields within IE/OR. But it also makes sense to see commonalities across the operations within these in order to leverage insights from one industry in another.

5. The Basic Elements of Operations Science

To move from definitions and classification to modeling, we note that because we know from empirical observation that there is always randomness and variability in demand and transformation that it is virtually never the case that demand and transformation occur at precisely the same time. If transformation occurs before demand, we create *inventory* and if transformation occurs after demand, there will be some amount of *wait time* between demand occurrence and satisfaction. With this in mind, we offer a description of the "primitive element" of an operation in Figure 4.

The basic behavior described in this figure is as follows. If a demand arrives (from the right) to find a suitable item in stock it is satisfied immediately. If not, the demand enters the queue on the left and waits. The transformation activity uses resources to satisfy demands in the queue. Both the stock and queue could be capacitated. Demands could balk if not filled immediately or leave the queue if not filled quickly enough. The goods or services produced by the transformation could have types that must be matched to demands. Other variants in the details are possible as well.

Figure 4 Primitive Element of an Operation [Color figure can be viewed at wileyonlinelibrary.com]

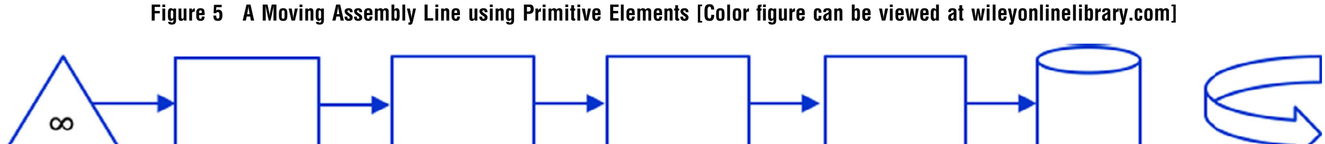


In conventional IE/OM terminology, this primitive element is a simple capacitated inventory system. But it includes all of the components of an operation—demand, transformation, resource(s), inventory, and wait time—in the above definition. Moreover, by changing the operating parameters, a wide range of situations can be represented. Setting the stock level to zero models most simple service operations as a queueing system (e.g., a triage station in a hospital emergency department). More complex service systems can be represented by combining primitive elements into a network of queues. A linear combination of elements with finite stock levels models a Kanban line, while a linear combination of elements with no intermediate stocks represents a push production line. Combining a linear set of elements, eliminating all intermediate queues and stocks, with demand arriving according to a takt time results in a moving assembly line as illustrated in Figure 5 (the infinity in the queue indicates that there will always be work to start subject to the takt time).

To be consistent with the IE/OM literature, we will call a linear sequence of elements a *flow*. An element that requires multiple entities that are themselves outputs of flows represents an *assembly* operation. An element that transforms multiple types of entities that move on to different elements is a *splitting* operation. By combining elements with these various options, we can represent almost any production or service system, at least in concept.

The above model deliberately echoes many of the mathematical models that are common in IE/OM. Hence a reasonable question is, what is potentially new about Operations Science? The main answer is that we have set up this framework as the foundation for a true descriptive science that will enhance the predictive and decision aiding power of IE/OM models far beyond that of the current axiomatic and ad

Figure 5 A Moving Assembly Line using Primitive Elements [Color figure can be viewed at wileyonlinelibrary.com]



hoc empirical models. We will take up this stream of development in the next section.

But there is another benefit to having a common framework for representing all operations. This can be useful because it gives us a way to make sure we have a complete model. Whenever we begin a study to improve an operation, we should ask:

- What and where does the demand come from?
- What is the transformation?
- What resources are used?
- Which resources are consumed, and which are capacity resources?
- How much variability is there in demand and transformation and what is its source (e.g., lack of information or inherent randomness or both)?
- How is this variability mitigated?

To illustrate this type of generic thinking, we consider an example of energy generation. Demand for electrical energy comes from individual users and is known to be highly variable across both short and long time frames. From Factory Physics (Hopp and Spearman 2008, Chapter 9) we know that there are three potential ways to mitigate this variability: inventory, time, and capacity. Of these, most systems in developed countries make almost exclusive use of capacity to accommodate variability. In less developed countries with inferior infrastructures, time in the form of blackouts is often employed as a default buffer, but this is an undesirable alternative. Accommodating "peak load" demand requires a large capacity buffer. Because capacity is very expensive, power companies also try to mitigate variability in demand by offering discounts to entities that have flexibility in their power usage. For example, some utilities offer discounts for being allowed to shut off air conditioners for a limited time during periods of peak demand.

Although they offer obvious environmental advantages, adding private wind and solar power to the mix and requiring that utilities use this power in their grids makes the variability problem worse because they offsets demand on windy and/or sunny days but not on windless and/or cloudy days. Because the utility must maintain capacity to meet peak load without wind and solar power but loses revenue on days when wind and solar reduce demand, it is forced to increase cost per kW-hr to maintain financial viability.

If instead homes were equipped with a way to store power in high capacity batteries, the situation would improve markedly. By offsetting load during windless and/or cloudy days, this inventory of power would reduce the peak load for which the utility would need capacity. This in turn would reduce the capital expenditure of the utility and hence the cost to the consumer.

Had an assessment along these lines been done, it is possible that the government would not have focused on providing incentives to install solar and wind generation but would have invested instead in the development of inexpensive, high capacity batteries. Indeed, as longer lasting Lithium batteries have become available in recent years, some utilities have started installing them as an alternative to more costly power generation equipment (Johnson and DeCarolis 2019). The point here is that applying well-known operations insights from the manufacturing sector to the power sector can be useful. This is one of the benefits of having a unified operations framework that spans disparate sectors.

6. Relationships

Although a unified operations framework can be useful, the real value of operations science is the ability to predict the consequences of changes to a system and thereby serve as a decision guide in the design, control, and management of operations systems. To achieve such predictive power, a central challenge is understanding the relationships between demand and transformation and how these are affected by variability. As in the above power sector example, how variability is mitigated is an essential issue in the design and control of almost all operations system. The variability at issue could be due to randomness or lack of information and can affect the demand process, transformation process, or both. Regardless of the source or type, variability disrupts the synchronization between demand and transformation, which degrades system performance.

How variability affects performance of an operations system depends on how it is buffered. In our elementary operation, if we do not do anything proactive, the lack of synchronization between demand and transformation caused by variability will cause stocks to build up when the transformation rate exceeds the demand rate and backorders and/or lost demand to accumulate when the demand rate outstrips the transformation rate. We can reduce the amount of customer waiting and/or lost demand by increasing the inventory buffer by setting stocking high stock targets. Conversely, if we reduce the stock targets, and hence the inventory, the time buffer in the form of waiting and/or lost demand will grow. Finally, we can reduce both inventory and time buffers by increasing transformation capacity to enable us to keep up with demand surges or capacity contractions.

We noted in Hopp and Spearman (2008) that there are three ways to buffer variability—time, inventory and capacity—in an operations system. However, at that time, we were not able to fully describe the interactions between the buffers and did not recognize that two of the buffers are not independent. We can now make use of the primitive element described to explore the relationships between the three variability buffers. Since, as we noted earlier, we can represent a very wide set of production and service systems with combinations of the primitive element, these results are very broadly applicable.

To begin, we note that we can model interactions between the buffers by using *net-inventory* which is defined as the beginning net-inventory plus the total production minus the total demand up to a point in time. If net-inventory is positive, we have on-hand inventory; if it is negative, we have backorders and hence waiting time or lost demand. Since time and inventory are related in this rigid fashion, there are actually only two independent variability buffers, time-inventory, and capacity.

Having noted this, we simulated the net-inventory for a wide range of capacitated base-stock systems to understand how net-inventory is impacted by utilization and variability. Figure 6 presents sample paths for systems with four combinations of high and low variability (in both demand and transformation) and high and low utilization of the transformation process. Because it is a base-stock system, the maximum net-inventory is the base-stock level, which in this case was 10. But the minimum depends on both the variability and the utilization.

To use the data generated by our simulations to explore the relationship between variability and the two buffers, we note that the net-inventory, IN, of a base stock system with a base-stock level of S and a random demand, $D(T)$, during the random replenishment time of T is

$$IN = S - D(T)$$

Note that the standard deviation of the net-inventory is equal to the standard deviation of demand during the replenishment time, σ_{RTD}. It is well known that this standard deviation is a key driver of inventory performance measures such as on hand inventory, backorders, and lost sales (Zipkin 2000). Consequently, it makes a useful measure of the time-inventory buffer, which we define as.

$$B_{TI} = \sigma_{RTD}$$

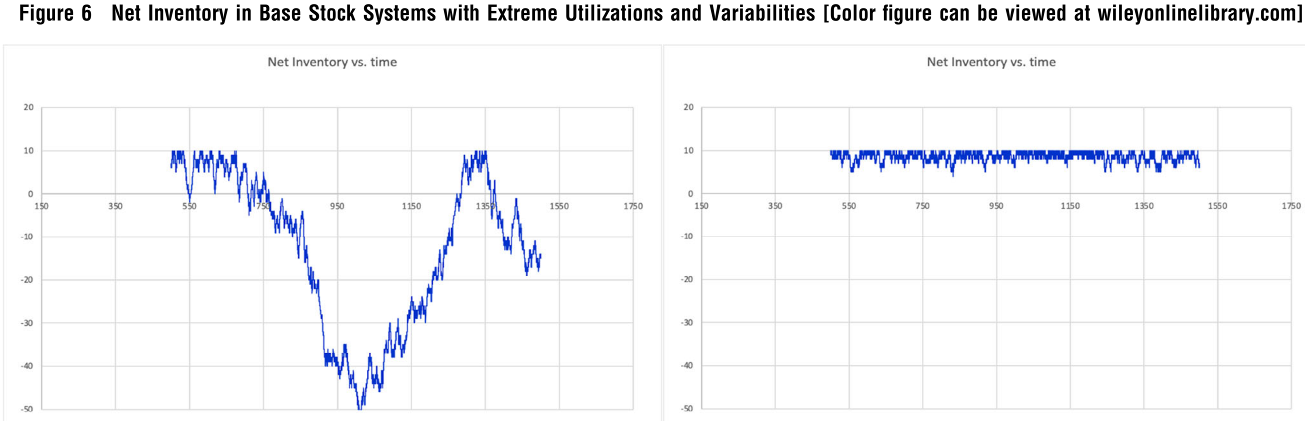
The capacity buffer is the difference between the mean capacity and mean demand, which we express as.

$$B_C = \mu - \lambda$$

The fundamental questions of interest are: (1) How do these two buffers interact? (2) How are they related to the variability in demand, $\sigma_D^2 = \psi^2\lambda$, and variability in transformation, $\sigma_P^2 = c^2\mu$? The high variability/high utilization and the low variability/low utilization cases behave as we would expect exhibiting alternative net-inventory sample paths that have high and low variability (see Figure 6). The interesting cases are the high variability/low utilization and the low variability/high utilization because their net-inventories appear to have similar sample paths and, in particular, similar variances (Figure 7).

To examine how variability and the two buffers interact, we take a cue from physics and hold one parameter constant while varying the others to see if a relationship can be determined. Specifically, we set the two variance to mean ratios (i.e., ψ^2, c^2) equal to a constant, V, and then vary V and the capacity buffer (B_C) while keeping the time-inventory buffer (B_{TI})

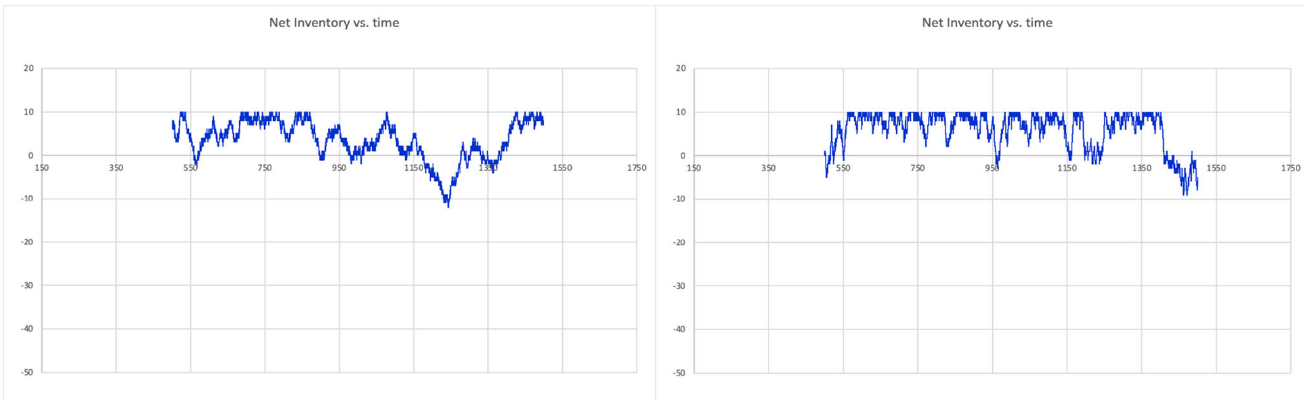
Figure 6 Net Inventory in Base Stock Systems with Extreme Utilizations and Variabilities [Color figure can be viewed at wileyonlinelibrary.com]



High Utilization, High Variability

Low Utilization, Low Variability

Figure 7 Net Inventory in Base Stock Systems with Alternately High and Low Utilizations and Variabilities [Color figure can be viewed at wileyonlinelibrary.com]



High Utilization, Low Variability   Low Utilization, High Variability

constant. Doing this over a wide range of cases with different levels of variability and utilization, revealed that the following relationship that fit the data very closely.

$$B_C \times B_{TI} \approx \frac{\sigma_D^2 + \sigma_P^2}{2}$$

By substituting the expressions for B_C and B_{TI} we can express this relationship as

$$\sigma_{RTD} \approx \frac{\sigma_D^2 + \sigma_P^2}{2(\mu - \lambda)}.$$

We acknowledge, of course, that simulating an imaginary system does not constitute a rigorous scientific result. Nevertheless, it is evocative of a potentially deep and fundamental relationship of operations science. The above equation suggests that the capacity and time-inventory buffers have a simple inverse relationship that involves the total demand and transformation variance. Since σ_D^2, σ_P^2, and σ_{RTD} are observable, this relationship represents a testable hypothesis of the very type we need to evaluate in order to create a legitimate descriptive science of operations.

7. Conclusions

We have made a case for creation of an operations science (OS) that is related to but fundamentally different from IE/OR. The main difference between OS and IE/OM is that OS is meant to be a descriptive science, whose mission is to understand how operations systems behave, while IE/OM is a prescriptive discipline, whose mission is to find ways to improve the design, control and management of operations systems. We envision OS supporting and enhancing IE/OR education, practice, and research in an analogous manner to which mechanics underpins mechanical engineering and psychology and sociology support organizational management.

In the course of making our case, we have offered some comparisons with other fields for inspiration and guidance. We have offered a sketch of a unified framework of operations as a parsimonious but general description of the common core of operations systems. We have given examples of empirical testing of the validity of specific axiomatic models to real-world situations (one of which refuted a result in our own book!). And we examined simulated data for an elementary operations system that revealed an intriguing testable hypothesis about a fundamental relationship between essential parameters. But whether this result is a seminal breakthrough, an important result in a limited set of environments, or a misguided blunder can only be determined by empirical testing against real-world data.

This is the real point of the paper. As Popper eloquently described, the scientific process is one that requires constant testing and refutation of ideas and frameworks. For a field to be scientific, scholars must make testable hypotheses and subject them scrutiny using real-world data. Because of our understandable focus on finding prescriptive solutions to practical problems, our IE/OM discipline has devoted far too little time to such hypotheses or empirical tests. We have relied too heavily on axiomatic models that are mathematically elegant but too simplistic to capture important aspects of actual system behavior. And we have been contemptuous of research that is purely descriptive in nature, relies on simulation rather than analytic models, or builds too directly on previous work. The result is that we have produced a literature filled with clever but idiosyncratic analyses of individual problems. Because they lack an integrative spine, the whole of this body of work is not greater than the sum of its parts.

But it should be. Operations facilitate the functioning of all organizations from the artistic and religious to the scientific and commercial. Human progress from the pyramids to the moon landings to the creation of COVID-19 vaccines is powered by operations. If ever there was a field that warranted pure scientific study, operations is it.

In addition to our illustrative examples, we have cited some exceptional research studies, including the classic bullwhip paper by Lee, Padmanabhan and Whang (1997), that exemplify the type of work we need to create a true operations science. But we need much more of these. To cultivate them, the IE/OM community must develop a more receptive culture for descriptive scientific work. Specific steps we recommend include:

1. Authors should write and editors should publish papers that summarize the big questions of operations science (i.e., issues that cut across broad categories of organizations), and note which have been answered and which have not.
2. Authors should write and editors should publish papers that make empirically testable claims about the behavior of operations systems.
3. Editors should push authors to compare mathematical models with empirical data. Showing that a result holds in one case but not another adds nuance and practicality to research results. It also provides stimulus for research progress.
4. Journals should establish processes for reviewing papers that claim paradigm changing results. Such papers often depart from established norms. "Vote counting" review processes may be ill suited to recognizing the importance of such papers. A better process might be one in which a set of senior editors scans papers that claim to make profound changes in the paradigm for operations science research.
5. Leading journals in our field should catalyze descriptive research by using special issues to attract and promote papers that make or test hypotheses about real-world operations systems.

These are things we can and should do collectively. And if we do, maybe we can finally put an end to papers like this calling for more science in management.

References

Adan, I., J. Van der Wal. 1989. Monotonicity of the throughput of a closed queueing network in the number of jobs. *Oper. Res.* **5**(6): 953–957.

Bertrand, J. W. M., J. C. Fransoo. 2002. Operations management research methodologies using, quantitative modeling. *Intern. J. Oper. Prod. Manag.* **22**(2): 241–264.

Brahimi, N., N. Absi, S. Dauzere-Peres, A. Nordli. 2017. Single-item dynamic lotsizing problems: An updated survey. *Eur. J. Oper. Res.* **263**(3): 838–863.

Buffa, E. S. 1987. *Modern Production/Operations Management*, Wiley, New York.

Cox, D. R. 1962. *Renewal Theory*, Methuen & Co. Ltd., London.

Donohue, K., Ö. Özer, Y. Zheng. 2020. Behavioral operations: Past, present, and future. *Manuf. Serv. Oper. Manag.* **22**(1): 191–202.

Fuller, S. 2004. *Kuhn Vs. Popper: The Struggle for the Soul of Science*, Columbia University Press, New York.

Harris, F. W. 1913. "How many parts to make at once." Factory: The Magazine of Management 10(2), 135–36. Reprint. *Oper. Res.* **38**(6): 947–950.

Hopp, W. J. 2011. *Supply Chain Science*, Waveland Press, Long Grove, IL.

Hopp, W. J., M. L. Spearman 2008. *Factory Physics, 3rd edn*. McGraw-Hill Irwin, Burr Ridge, IL.

Hopp, W. J., W. Lovejoy 2012. *Hospital Operations: Principles of High Efficiency Health Care*, FT Press, Upper Saddle River, NJ.

Hopp, W. J., D. Simchi-Levi. 2020. Management Science: The legacy of the past and challenge of the future. *Management Sci.* Articles in Advance : 1–11. Available at http://pubsonline.informs.org/journal/mnsc

Inman, R. 1999. Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems. *Prod. Oper. Manag.* **8**(4): 409–432.

Johnson, J., J. F. DeCarolis. 2019. Utilities are starting to invest in big batteries instead of building new power plants. Available at https://theconversation.com/utilities-are-starting-to-invest-in-big-batteries-instead-of-building-new-power-plants-110961 (accessed date February 22, 2019).

Kahneman, D. 2011. *Thinking, Fast and Slow*, Farrar, Straus and Giroux, New York.

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

Lee, H. L., V. Padmanabhan, S. Whang. 1997. Information distortion in a supply chain: the bullwhip effect. *Management Sci.* **43**(4): 546–558.

Little, J. D. C. 1961. A proof of the queuing formula: L = lW. *Oper. Res.* **16**(3): 651–655.

Little, J. D. C. 1992. Are there laws of manufacturing? J. A. Heim and Compton, W. D. (eds). *Manufacturing Systems: Foundations of World-Class Practice*, National Academy Press, Washington, DC, 180–188.

Ohno, T. 1988. *Toyota Production System: Beyond Large-Scale Production*, Productivity Press, New York.

Popper, K. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge, London.

Schmenner, R. W. 1990. *Production/Operations Management: Concepts and Situations*, Macmillian, New York.

Schmenner, R. W., M. Swink. 1998. On theory in operations management. *J. Oper. Manag.* **17**(1): 97–113.

Smiddy, H. F., L. Naum. 1954. Evolution of a "Science of Managing" in America. *Management Sci.* **1**(1): 1–31.

Thaler, R. 2015. *Misbehaving: The Making of Behavioral Economics*, Norton, New York.

Wagner, H. M., T. M. Whitin. 1958. Dynamic version of the lot size model. *Management Sci.* **5**(1): 89–96.

Zipkin, P. H. 2000. *Foundations of Inventory Management*, McGraw-Hill, New York.