Received: 24 April 2020 | Accepted: 20 July 2020
DOI: 10.1002/joom.1115

JOM FORUM

WILEY

The lenses of lean: Visioning the science and practice of efficiency

Wallace J. Hopp[1] | **Mark S. Spearman**[2,3]

[1]Ross School of Business, University of Michigan, Ann Arbor, Michigan
[2]Strategic Project Solutions, Inc., San Francisco, California
[3]Project Production Institute, San Francisco, California

Correspondence
Wallace J. Hopp, Ross School of Business, University of Michigan, Ann Arbor, MI 48109.
Email: whopp@umich.edu

Handling Editor: Suzanne de Treville

Abstract

There is a significant gap between the descriptions of Lean used by industry practitioners and the various bodies of academic research that have studied the theory and application of Lean. There is also a gap between applied research on Lean and basic research in the mathematical, physical and social sciences. As a result, Lean practice is based largely on trial-and-error experience while potentially valuable research results remain locked away unused in archival journals. This paper attempts to close these gaps by describing four "Lenses of Lean," each of which aligns with a practical perspective and rests on a distinct body of conceptual research. Our hope is that this framework will provide a useful construct for Lean training and implementation and will also spur academic research that is relevant to advancing Lean practice.

KEYWORDS
efficiency, factory physics, lean, operations management

1 | WHAT IS LEAN?

The term "Lean" was coined by Krafcik (1988) and popularized by Womack, Jones, and Roos (1991). But despite its undeniable success as a label, "Lean" is actually a pretty old name for a very old concept. To begin with, Lean was explicitly derived from the Toyota Production System (TPS). Indeed, the original descriptions of Lean equated it with TPS. The seminal book on Lean, *The Machine that Changed the World*, stated "Toyota had fully worked out the principles of lean production by the early 1960s" (Womack et al., 1991, p. 68). The first paper on TPS in English appeared in the late 1970s (Sugimori, Kusunoki, Cho, & Uchikawa, 1977), but it took until the 1980s for real interest to be sparked by descriptions of TPS in popular books (Hall, 1983; Schonberger, 1982). By the 1990's, these had been digested and the Lean label had begun promoting the application of TPS beyond the automotive industry. These led to extensions in the definition and tools of Lean, to the point where now TPS is considered Lean but not all Lean implementations are considered TPS.

Just as Lean was rooted in TPS, TPS was rooted in an earlier system introduced at Ford (Ford, 1922). The Ford Production System revolutionized manufacturing by introducing the moving production line to complex products like automobiles. It also established a focus on efficient, waste-free flow that remains at the center of TPS and Lean systems. The Henry Ford Health System still uses the title Henry Ford Production System for their collection of practices that include Kanban, just-in-time, standardized work and other tools right out of TPS (Zarbo & D'Angelo, 2006).

Of course, Ford did not create his system from whole cloth either. Moving assembly lines existed before the Highland Park plant began using one to produce Model T cars in 1913. Efficiency methods in management can be traced backward through the Scientific Management movement around the turn of the 20th century, the American System of Manufacture in the 19th century, the (First) Industrial Revolution in the 18th century, the European Scientific Revolution in the 16th and 17th centuries, engineering and manufacturing innovations in

China during the first Millennium, and back at least as far as the construction of the Pyramids of Egypt beginning before 2,500 BC. For as long as humans have been producing goods and services, making efficient use of resources has been a priority.

The fact that Lean is not new does not diminish its importance. Enhancing efficiency increases productivity, which in turn elevates standards of living. Hence, it is not an overstatement to say that efficiency, the goal of Lean, is central to improving the human condition.

But what exactly is it? The Lean Enterprise Institute defines the core idea of Lean as "maximize customer value while minimizing waste." This definition is appealingly simple but decidedly unhelpful for implementation or research purposes. The list of tools associated with Lean and TPS (e.g., Kanban, 5S, kaizen, andon, heijunka, jidoka, etc.) contains many useful ideas but also falls short of a unifying definition.

To produce a working definition (actually, multiple definitions) of Lean that will serve our goal of bridging the gap between research and practice, we begin with three observations:

1. *History did not start with Toyota*. Lean may have started out as a generic label for TPS, but it but its roots predate TPS and its applications go well beyond it. Limiting Lean to the philosophy or practices at Toyota is arbitrarily narrow and unsuited to our purposes.
2. *There is a difference between a title and a term*. A title is simply a name that can be applied however one likes, but a term requires a definition. In the world of practice, Lean is generally used as title, which can refer to a collection of practices, a near-religious commitment to waste reduction, or just good operations management, rather than a well-defined term. Although such usage is maddeningly imprecise from a scholar's perspective, titles can be useful in practice. Henry Ford Production System, Michigan Quality System and Danaher Business System are examples of titles for collections of ideas and practices without rigorous boundaries or definitions, but which serve to focus and motivate activities within an organization. However, because Lean is used as a loosely specified title in practice, scholars that attempt to use it as a term are usually vague (when they fail to define Lean) or quarrelsome (when they do and then take on all the writings that conflict with their definition).
3. *All models are wrong, but some are useful*. This aphorism, usually attributed to statistician George Box, offers a way out of the vague vs. quarrelsome trap. If we admit the reality that every conceptual framework, including statistical mechanics, the theory of evolution and the entire discipline of Operations Management, is a simplification of reality, and hence "wrong," we can accept the pragmatic view that any definition of Lean that elevates practice or deepens our understanding of how to elevate practice is good.

With these in mind, we first describe what we feel is the broadest defensible definition of Lean, which delineates the potential scope of Lean research and practice. We then offer four narrower definitions of Lean that reflect the range of descriptions from the research and practitioner literature. Each, while admittedly incomplete, points toward a connection between practice and a particular body of research. Each is also appropriate as a practice guide in particular environments. This allows us to leverage these four definitions to describe four lenses through which Lean can be viewed productively and which provide the framework for this paper.

The maximal definition of Lean follows from examining what part of the Operations Management (OM) discipline it comprises. Obviously, if Lean encompasses all of OM it is redundant and unnecessary. Conversely, if Lean refers to a narrow set of practices in a limited set of environments, it is not very interesting from a scholarly standpoint or important from a practice standpoint. Therefore, we aim for something in the middle.

We do this by noting that a mainstream definition of Operations Management is the business function responsible for *effective delivery of goods and services to customers*. If we define goods/services and customers broadly enough, virtually any organization is engaged in delivering goods and services to customers. As such, OM is fundamentally about facilitating execution of a business strategy. However, what this means in a given setting depends on how "effective delivery" is defined. Fundamental dimensions, along which customer concerns can be described, include quality, time, variety and price/cost. Management of each of these defines an OM sub-field. For example, the quality dimension is the focus of the quality management area of OM, which has been addressed in practice under the headings of Total Quality Management and Six Sigma. The time dimension is the focus of the responsiveness management theme of OM, which has been addressed in practice as Time-Based Competition, Agile Manufacturing, and Quick Response Manufacturing (Suri, 1998). Variety is a key focus of the product management area of OM, which has been addressed in practice as Product Portfolio Management and (in part) by Material Requirements Planning.

Almost everyone would agree that Lean focuses primarily on efficiency, which means it addresses the price/cost dimension. Although one could further limit Lean within the efficiency space to certain environments (e.g., systems involving repetitive production or limited amounts of variability), this seems arbitrary to us. Therefore, for the purposes of this paper, we will equate Lean with efficiency management and regard anything that increases the efficiency of delivering products as a Lean practice. Under this definition, Lean is analogous to but distinct from quality, responsiveness and product management in the OM spectrum.[1] Lean practices, such as value stream mapping or Kanban are analogous to specific practices within these areas, such as fishbone diagramming and statistical process control in the quality management area.

We acknowledge that this broad definition begs the question of whether we need Lean as a term at all. Why not simply call it efficiency management? Why not indeed. Using industry titles, which are imprecisely defined and subject to change, as research terms is fraught with problems. For instance, consider the case of Six Sigma, which, like Lean, is a title that was applied to emergent industry practices. Six Sigma started out as a defect measurement tool at Motorola, where it was gradually expanded to include other statistical quality control tools. At that stage, it was clearly a methodological subset of the quality management area. But when GE adopted Six Sigma and added DMAIC, training with "belt" certification levels, and other features, it grew into a full-fledged management system that could be equated with quality management as a whole. Even if one does not regard the GE version of Six Sigma as encompassing all of quality management, it is only a matter of time until another firm defines it as such. Indeed, it is now common to find descriptions of Six Sigma that include TPS practices, which takes it even beyond the quality space into the efficiency space. Such shape shifting definitions make communication difficult and rigorous research impossible. Consequently, we would prefer to use generic terms like quality management and efficiency management, which can be rigorously defined and will always be with us, rather than imprecise industry labels like Six Sigma and Lean that will eventually be replaced by trendier titles.

But, for the purposes of this paper, it does not really matter whether we use Lean or efficiency management. Either way, we are still only providing a heading for the problem area without any hint at structure or solution approaches. This is why writers in the academic and practice literatures have offered definitions of Lean that go beyond simply categorizing it as focusing on efficiency. The following list of Lean definitions is certainly not comprehensive, but it spans the gamut of ways, from simple to sophisticated, that Lean has been described by scholars and practitioners:

1. *Lean is the pursuit of waste elimination*: This is the goal implied by the Lean Enterprise Institute definition and is the essence of most descriptions of Lean in the practitioner literature. Some of these emphasize value in place of waste, for example by stating the goal as "creating more value for customers with fewer resources" (Lean Enterprise Institute, 2019). But, since waste is defined as anything that does not create value for customers, it is essentially the same as reducing waste. While indisputable as a goal, this definition only barely goes beyond the overarching efficiency management definition by introducing the concept of waste. Because it offers no explicit guidance on the causes or cures of waste, this definition tends to foster a focus on the most directly visible forms of waste and straightforward ways to eliminate them.
2. *Lean seeks to minimize the cost of excess inventory, capacity or time*: This definition invokes a basic insight from Factory Physics that variability must be buffered by some mixture of inventory, capacity and time (Hopp & Spearman, 2008). The implication is that variability is the main cause of waste beyond the obvious direct waste inherent in unnecessary activities. By emphasizing the cost of variability buffering, this definition forces a focus on both variability reduction and finding the most efficient mix of buffers for the variability that remains (Hopp & Spearman, 2004).[2] Framing Lean in terms of variability buffers highlights the underlying causes of waste and thereby helps to discover less obvious sources of waste than those likely to be surfaced by the first definition.[3]
3. *Lean is a systematic process for reducing the cost of waste*: Unlike the first two definitions, which define Lean in terms of its goal, this definition focuses explicitly on the improvement process. There are many books and articles in the practitioner literature that take this approach, generally by starting with the basic "production without waste" definition and then laying out a series of steps for operationalizing it. For example, Womack and Jones (1996) described a five-stage sequence consisting of Define Value, Map the Value Stream, Create Flow, Establish Pull, Pursue Perfection. By focusing on the improvement process, this approach to defining Lean promotes a more expansive system-wide perspective and a management focus that is absent from the first two definitions.
4. *Lean is an organizational culture that encourages continual reduction of the cost of waste*: This perspective

takes us fully into the realm of management by recognizing that it is people working within organizations that create and manage the systems for carrying out the waste elimination activities called for in Lean. Spear and Bowen (1999) and Liker (2004) are examples of practice-oriented descriptions of Lean that focus on creating cultures of waste elimination. For the most part, however, such descriptions are attempts to summarize characteristics of Toyota and other successful Lean organizations, rather than applications of behavioral and organizational science to guide the creation of a successful Lean culture.

We reiterate that none of these definitions is “correct” in any rigorous sense. But, in the spirit of “all models are wrong, but some are useful,” each has practical value. By providing four different “lenses” through which Lean can be viewed, these suggest modeling frameworks for scholars and problem-solving perspectives for practitioners. In the sections that follow, we examine each of these lenses, trace them to the relevant bodies of research, and describe the environments in which they are likely to be most useful in identifying the causes and remedies for waste in production and service systems. We will argue that the simple lens based on the first definition has been used (overused?) in practice but provides a limited basis for research, that the lenses based on the second and third definitions have been the basis for the most important research into Lean but may not have been as fully exploited by practitioners as they deserve to be, and that the lens based on the fourth definition is ripe with opportunity for scholars and practitioners alike. Furthermore, when taken together, this set of Lean lenses highlights opportunities for linking disparate streams of research to yield deeper insights into the science of efficiency and also suggests a staged structure for Lean implementations.

2 | THE PROCESS LENS

The first definition of Lean is “pursuit of waste elimination.” This goal has been behind every effort to develop a better way to produce a good or service since the dawn of civilization. But it was not formalized until the late 19th and early 20th century under the banner of Scientific Management. Frederic W. Taylor, the Father of Scientific Management, advocated “one best way” to perform a given work function and introduced the practice of time studies to discover such best practices systematically (Taylor, 1911). Fellow Scientific Management proponent Frank Gilbreth carried the process for finding best practices a step further through detailed motion studies of manual work (Gilbreth, 1911). In addition to establishing the concept of “standardized work” that would become central to TPS and Lean, the Scientific Management movement lead to creation of the field of Industrial Engineering (IE).

However, because the “pursuit of waste reduction” definition of Lean lacks any reference to underlying causes of waste, it basically leaves the user to focus on waste that is directly visible from observation of individual processes. For this reason, we term the lens implied by this definition the *Process Lens*. Taylor and Gilbreth focused on waste at the process level in their work during the Scientific Management era in the early 20th century. For example, the waste on which Taylor focused his time studies (e.g., shoveling with an improperly sized shovel) is appropriately termed *processing waste*. In contrast, the waste at which Gilbreth aimed his motion studies is logically termed *motion waste*, since it involves unnecessary motions (e.g., excessive reaching and lifting bricks). Contemporary descriptions of Lean have added five more categories of waste to those addressed by Taylor and Gilbreth to produce following list of seven types of waste:

1. Defects
2. Overproduction
3. Transportation
4. Waiting
5. Inventory
6. Motion
7. Processing

Although this set of waste categories is ubiquitous in the Lean literature it is neither precise nor valuable. For example, an integrated steel mill that can only produce steel in 250-ton batches (or “heats”) will have leftover slabs any time a customer orders less than 250 tons. These will pile up in storage waiting for another order for that particular blend of steel. Is this overproduction waste? Inventory waste? Does it matter? We argue that the label is irrelevant to the elimination of the waste. What matters is that the large and rigid batch size makes it impossible to match supply with demand efficiently. Reducing this waste—whatever it is called—requires changing the process to facilitate smaller and more flexible batch sizes or finding less costly ways to accommodate the current batch size. Unfortunately, nothing in the Process Lens or the list of waste categories it has engendered guides us to this conclusion.

Because it lacks guidance on how to diagnose and remedy waste, the Process Lens is best suited to identifying waste that is created directly in the process itself, rather than as a by-product of issues or activities outside the process. As such, it can help us identify ways to improve efficiency by improving specific steps in a process. Such improvements can be facilitated by a classic industrial engineer (with a clipboard and stopwatch) or by workers with sufficient training to do classical IE analysis of their operations. For example, a 5S reorganization of a workstation on an assembly line can be very effective in revealing and reducing motion and transportation waste. For organizations just beginning their Lean journey, this very simple view of waste reduction can be helpful in attacking obvious waste. But this lens is less helpful in identifying waste that propagates from other parts of the process or beyond it. It is also not terribly useful in enabling organizations to drive out waste on a continual basis because it does not provide any framework for prioritizing different sources of waste. Firms seeking to go beyond the most basic level of Lean need a more detailed framework than the Process lens.

The Process Lens is also limited as a frame for research. Since it lacks any perspective on the means for eliminating waste, it cannot draw on bodies of basic research for guidance. This leaves direct research into Lean as the only avenue for research suggested by the Process lens. For example, scholars have looked for empirical evidence of the impact of Lean practices on firm performance (Belekoukiasa, Garza-Reyesb, & Kumarc, 2014; Lopes Negrão, Godinho Filho, & Marodin, 2017; Moraro, Lemstra, & Nwankwo, 2016) and on the workforce (Parker, 2003; Ulhassan, von Thiele Schwarz, Thor, & Westerlund, 2014; Vidal, 2007). But to go further and learn from research other than direct studies of Lean systems, we need a richer definition of Lean.

3 | THE FLOW LENS

The second definition of Lean is "*to minimize the cost of excess inventory, capacity or time*." The explicit identification of exactly three types of buffers is essential to Lean because the core challenge in delivering goods or services to customers is *efficiently matching supply with demand*. The Process Lens focuses on the efficiency element of this challenge by seeking to drive out waste in execution caused by activities that do not add value to the customer. While this is fine for highlighting waste due to unnecessary motion or process errors, it cannot help us identify waste that is the result of variability in the supply and demand processes. The second definition of Lean zeroes in on variability-induced waste. Because the effects of variability are manifested in the flows of people, materials, dollars or other entities, we term this second lens the *Flow Lens*.

Variability is defined as any deviation from absolute regularity (see Hopp & Spearman, 2008, Chapter 8). Fluctuations in customer preferences, order sizes, seasonal trends and many other factors contribute to demand variability, while fluctuations in production rate, product yield, rework rate, staffing, delivery times and many other factors contribute to supply variability.[4] Variability may or may not be accompanied by uncertainty. For example, planned shifts in production output constitute variability but are not uncertain, while unplanned reductions in productive capacity due to equipment failures result in variability that is also uncertain.

Omitting variability from discussions of Lean, as is often the case in descriptions based on the Process Lens perspective, is a huge oversight because (a) all real-world systems involve variability and (b) addressing indirect waste caused by variability is more difficult than addressing direct waste from unnecessary activity. Managing production without variability is like farming without weather: It would be easy if it existed. Just as managing the impacts of weather fluctuations is essential to farming success, managing the indirect waste due to variability is fundamental to business success.

The fundamental law of Factory Physics that describes how variability causes waste can be stated as follows (Hopp & Spearman, 2008, p. 309):

> ***Variability-Buffering Law***: *Variability in a production system will be buffered by some combination of inventory, capacity and time.*

To make the concept of buffers concrete and to show how they relate to the various types of waste, we consider a sandwich station within a retail bakery. Demand for sandwiches is variable due to factors such as time of day, weather, and the whims of individual customers. Supply (i.e., the pace at which customers are served) varies due to product differences (e.g., a hot sandwich takes longer to make than a cold one), stockouts (e.g., the sandwich station runs out of Provolone cheese and the clerk needs to go in back to slice more), and other factors. As a result, there will be intervals in which the demand rate exceeds the supply rate and a backlog of customers builds up.

The queue that results when supply lags demand constitutes a *time buffer* since it involves customers waiting for their sandwich. The most straightforward way to reduce this customer waiting is to put more clerks on sandwich making duty. The excess capacity of the clerks during slow periods represents a *capacity buffer* since it is capacity held in reserve to deal with demand spikes. An alternative for reducing customer waiting is to make up common sandwiches (e.g., turkey and swiss on white) in advance. These stocks of pre-made sandwiches, which will not stay fresh as long as the individual ingredients and hence are at risk of being discarded, represent an *inventory buffer*. The more variability in demand and/or supply, the more inventory and/or capacity and/or time buffering there must be.

In the language of the seven types of waste from the Process Lens, time buffering is a form of waiting waste and inventory buffering is a form of (what else?) inventory waste. The other five types of waste (defects, overproduction, transportation, motion, processing) could be capacity buffers if they are the result of variability (e.g., an unexpected spike in demand leads to defects due to rushing or extra motion to get an ingredient that has run out) or if they provide protection against future variability (e.g., overproduction produces extra inventory that can be used to meet a demand spike). But they are often simple waste due to poor design or execution (e.g., a badly laid out station that forces clerks to walk around to do their work). But the point of the variability focus of the Flow Lens is not to take us back to the waste labeling approach of the Process Lens. Instead it is to help us trace the causes of variability and provide a guide to minimizing the total cost of variability buffering.

The Variability-Buffering Law indicates two candidates for reducing the cost of variability buffering: (a) reduce variability, and (b) alter the mix of variability buffers. To do this in a systematic way, we need to understand how variability buffers interact.

We begin by noting that time and inventory buffers are clearly mirror images of the other. If a part is produced before its demand, it waits in stock as an inventory buffer. If it is produced after the demand, the customer must wait, incurring a time buffer. Variability causes a lack of synchronization between demand and production, and hence results in both types of buffers. But we can adjust the balance between them via an inventory control policy. By increasing the target stock level, we will increase the inventory buffer but, because it lowers the likelihood of stockouts, decrease the time buffer. The mechanics of this time-inventory buffer have been well described in inventory control theory (see Zipkin, 2000 for an excellent summary).

We can also describe the relationship between the time-inventory buffer and the capacity buffer. Specifically, as shown in Spearman and Hopp (2019), the product of the time-inventory buffer and the capacity buffer is a constant that is monotonically increasing in the variability of both the demand and the supply processes. This means, for example, that if we double the capacity buffer, the time-inventory buffer will be halved. This inverse relationship describes a "diminishing returns" to buffering. Each time we double the capacity buffer, the absolute increase gets larger, but each time we halve the time-inventory buffer the absolute decrease gets smaller and smaller. In economic terms, this relationship shows diminishing returns when adding buffers. This, in turn, implies an optimal buffer configuration for almost any cost structure for the buffers.

It is important to note that, as emphasized by underlining in the Variability-Buffering Law, these buffers are not optional. If there is variability, there will be buffers. A small capacity buffer will lead to larger fluctuations of net-inventory that create either larger stocks or more backorders. The *only* way to reduce the total amount of buffering (as measured by the product of the buffers) is to reduce total variability.

It is worth noting that TPM practices include a host of variability-reduction techniques, such as, total quality management, preventive maintenance, production smoothing, standardized work, and others. Each of these techniques serves to smooth demand and supply flow so that they can be matched more efficiently. Although almost never described in this manner in the Lean literature, TPS can be thought of as a giant variability-reduction machine.

To provide more structure to the search for ways to reduce the cost of variability buffers, it is helpful to characterize variability in terms of predictability and customer value. Variability can be predictable (e.g., planned downtime for maintenance) or unpredictable (e.g., emergency outages). This distinction matters because predictable variability can be buffered only when needed, while unpredictable variability must be buffered continuously. This implies that predictive analytics techniques that convert unpredictable variability into predictable variability can reduce buffering costs. For example, processing data from wear monitors to reliably determine when a machine needs preventive maintenance and thereby prevent random failures will reduce the amount of inventory buffering needed to maintain downstream flow during outages.

A second useful characterization of variability is in terms of customer value. Variability can be the result of errors (e.g., machine failures), which have no customer value, or the result of actions that produce customer value (e.g., providing product variety). The former are simply targets for elimination, while the latter pose tradeoffs to be addressed through management decisions. For example, an automaker can reduce variability and cost by limiting the number of vehicle variants. Ford famously did this by painting all Model Ts black. Modern car companies address this variability vs. variety dilemma by building many different models on a single vehicle platform. Managing variability rooted in customer value can be complex.

To translate this conceptual description of variability and buffers into practice, we return to the bakery sandwich-station scenario. In this system, we could reduce demand variability by having customers call in sandwich orders so they can be pre-made during slow intervals. We could reduce supply variability by stocking the sandwich station more carefully to prevent disruptions due to stockouts. Reducing either demand or supply variability will reduce the amount of buffering required and hence the cost.

Since supply variability is largely self-inflicted, the bakery can reasonably strive to reduce it to near zero through better ordering and stocking execution. It could also make demand variability predictable by requiring customers to pre-order their sandwiches or make it minimal by having them sign up for time slots (e.g., like a dentist does). But both of these would be burdensome to the customer and hence competitive disadvantages. Consequently, demand variability in this scenario, as in many other service environments, has positive customer value.

This means the bakery will always have to buffer demand variability. The challenge is to find the most economical mix of capacity (sandwich clerks), inventory (pre-made sandwiches), and time (customer waiting). Although sufficiently detailed data to do a formal cost optimization is probably unavailable, it is certainly feasible to implement a satisficing solution that sets a customer waiting-time target that will satisfy the market and then find by trial-and-error a staffing/inventory mix that achieves it.

We can make variability buffering more efficient by invoking the following useful corollary to the Variability-Buffering Law (Hopp & Spearman, 2008, p 313):

> ***Buffer-Flexibility Corollary***: *Flexibility reduces the amount of variability buffering required in a production system.*

There are two reasons for the above corollary: (a) flexibility facilitates variability pooling, and (b) flexibility facilitates synchronization of supply and demand variability.

Variability pooling takes advantage of the statistical property that a sum of random variables is less variable, percentage wise, than the individual random variables in the sum. An example of pooling in the bakery sandwich station scenario would be replacing three separate stocks of turkey and swiss sandwiches, one with mustard, one with mayo and one with both, with a single stock of turkey and swiss sandwiches bagged with a packet of mustard and a packet of mayo so customers can add either or both. Because combined demand for the three types of sandwiches is more predictable than the demand for each individual sandwich type, we can carry less safety stock in the pooled case than in the un-pooled case. This is an example of flexible inventory.

Synchronization of supply and demand variability makes it possible to reduce buffering because it is actually only unsynchronized variability in these two processes that creates the need for buffering. If capacity increased and decreased in exact proportion to demand, customer orders would be filled just in time without excess inventory or capacity. Of course, this never occurs in the real world because we lack sufficient information to perfectly match capacity to demand on a real time basis. However, we can use flexibility to partially synchronize capacity and demand. An example in the sandwich station scenario would be to cross-train workers from other areas of the bakery, such as the bakers or service counter staff to help make sandwiches when demand spikes. By using the flexible capacity provided by these cross-trained people only when needed, we will require less capacity buffering than if we could only assign sandwich specialists who would be on duty during idle periods as well as busy periods.

Of course, cross-training is more than simply a variability-pooling technique. It also impacts the worker experience by changing the nature of jobs. This impact can be positive or negative. Negative impacts of Lean practices have been documented in studies such as Parker (2003) and Vidal (2007). But we observed a positive example in a consulting engagement in which we were seeking ways to increase the capacity of a very expensive bottleneck. As we often do, we put the question to the workers themselves in a workshop. The workers in a non-bottleneck unit quickly came up with a plan to swap positions with the bottleneck team after their lunch break and before the bottleneck team took their lunch break. This effectively shifted 30 minutes of labor capacity per shift (for three shifts) from the non-bottleneck to the bottleneck. And, much to our delight, it made the workers in both units happier to have a greater variety of work during the day. We will drill more deeply into the connection between the physics and psychology of flow in the next section.

There is a time analogy to the capacity matching made possible through cross-training. We came across an example of this years ago in a manufacturer of customized institutional cabinets. The manufacturer had a brochure that promised 10-week lead times for delivery of an order. However, because one of their competitors had begun quoting 4-week lead times, the firm was looking to shorten their quotes. We suggested shifting to dynamic lead times that would take into account the work backlog at the time the order was received. Because 10 weeks was chosen as a lead time, they were almost completely confident they could meet, it meant every customer received a "worst case" quote, even when the work backload was light. By moving to dynamic quotes, the firm was able to continue quoting times they were nearly certain to meet and quote an average lead time significantly shorter than 10 weeks.[5] As in the cross-training case above that made use of demand information to adjust flexible capacity in relation to demand, the dynamic lead-time quoting policy made use of demand information to adjust lead times in relation to demand. As such, it was an example of a flexible time buffer.

Finally, an aspect of variability that has been largely overlooked in both the practitioner and academics literatures on Lean is that variability occurs over different time scales. Minute-by-minute variation in supply and demand lead to queueing delays. Day-by-day or week-by-week fluctuations in product mix can shift bottlenecks. Rare but extreme events can trigger crises. The Variability-Buffering Law and its Buffer-Flexibility Corollary apply to variability no matter what the time frame. Hence, a well-designed Lean program should go beyond addressing the minute-by-minute variability we usually think of (when we think of variability at all) in the context of Lean. This requires thinking about the interactions between the buffers used for variability at different time scales.

To see this, consider a set of franchised bakeries in a metropolitan area. These independently owned bakeries make use of cinnamon chips they each purchase from a common supplier. Most of the bakeries receive replenishment orders every other week. Because the holding cost is relatively low and the cost of stocking out is high (because these chips are needed in some popular products), Newsvendor logic argues for ordering enough cinnamon chips to ensure they almost never run out.

But what does “almost never” mean? Surely it means having enough to cover plausible demand spikes and the occasional wastage due to burned bread or other mistakes. But does it mean having an extra supply in another room in case the regular supply gets destroyed by water from a roof leak? Does it mean having enough stock to protect against a fire at the supply plant that cuts off deliveries for 6 months? Probably not. Holding so much safety stock of every important ingredient would require a great deal of space and would hardly qualify as Lean.

The key tradeoff here is one of cost versus responsiveness. Physical buffers such as inventory and capacity are expensive because they must be maintained at a cost constantly. But they are ready for instant deployment. Two strategies that are less expensive and less responsive are *contingency planning* and *emergency response*. Contingency planning involves anticipating events and defining plans for dealing with them, while emergency response involves building some institutional readiness for events without detailed plans. Contingency planning is generally less expensive to install and maintain than physical buffers, and emergency response.

Is usually even cheaper.[6] If the objective is to minimize expected cost of variability, the more costly measures should be used where the likelihood and consequences of disruption are high.

Figure 1 provides a schematic for mapping variability-response strategies to different scenarios. Short-term variability that is highly likely to occur in a given interval is suited to buffering with inventory, capacity, or time, possibly in combination with the flexibility strategies discussed earlier. In the cinnamon chip example, this would translate into each of the franchised bakeries stocking enough chips at their stores to ensure a high service level. Medium-term variability that is too rare to justify the constant presence of a buffer but still likely enough to occur to warrant advance attention can make use of a contingency plan. For example, the bakeries could implement an inter-store sharing agreement under which they agree to lend cinnamon chips to one another in emergency situations such a store that has chips damaged by a roof leak.[7] Long-term variability that is too rare to justify even a contingency plan must rely on emergency response. For the bakeries, this might involve seeking out a backup cinnamon chip supplier who could provide chips if the primary supplier were disrupted. Of course, all of these responses are costly, which implies they should only be used where the benefits outweigh the costs. If not, then no action (e.g., do not bake cinnamon bread and try to divert customers to substitute products like cranberry-walnut bread) may be the optimal choice.

The variability-response mechanisms for different time scales are very different, but they impact one another. For example, if a bakery holds more safety stock to cover routine variability, it will be more likely to be in a position to share if another bakery has an emergency. If the bakeries collectively hold more combined safety

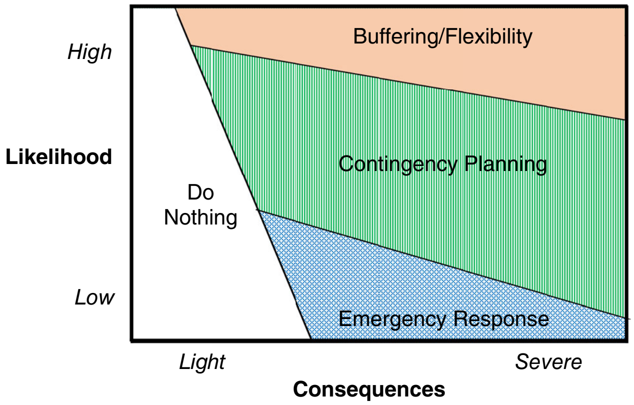

FIGURE 1 Addressing variability on different time scales. *Source:* Adapted from Hopp (2011) [Color figure can be viewed at wileyonlinelibrary.com]

stock, they will all be able to keep making product for a longer time while seeking a new supplier in the event their regular supplier is disrupted. The Flow Lens focus on minimizing the cost of buffering implies that we should think about variability on different time scales, and the interconnections between the different types of buffers.

Note that this sort of layered response to variability on different time scales is not limited to production systems. An example of a service system is a hospital emergency room. Normal demand variability is buffered by staffing to accommodate a high percentile of the demand distribution. Demand spikes above this level may be addressed by calling in off duty staff members. Even more extreme demand scenarios, due to a mass casualty event for instance, may be dealt with by diverting some patients to other emergency rooms in the area. Again, a Lean strategy is one that seeks to minimizes the total cost of buffering over time, which will require using and coordinating different buffering mechanisms.

If you have done much reading of the Lean literature, the majority of which is based on the Process Lens, you may be thinking that Figure 1 and the discussion of it are not Lean at all. Indeed, even serious scholars such as Yin, Stecke, Swink, and Kaku (2017) have been careful to distinguish between Lean, which deals with fixed buffers in low variability environments (e.g., inventory to maintain flow in assembly lines), and Agile, which deals with flexible buffers in high variability environments (e.g., cross-trained capacity to respond to demand spikes). But there is no scientific reason to distinguish between Lean and Agile on the basis of the amount of variability in the environment and the type of buffers used. The Variability-Buffering Law and the Buffer-Flexibility Corollary apply to both scenarios. And, as noted by Yin et al., a number of practices (e.g., cross-training and lead-time reduction) are associated with both Lean and Agile.

However, as Yin et al. also note, one can certainly distinguish Lean from by focusing on strategic intent. If Agile is defined as the science of responsiveness, then it is indeed different from Lean, which we equate with the science of efficiency. But as long as we are focused on efficiency, then it should not matter whether variability is high or low or buffers are fixed or flexible. Browning and Heath (2009) recognized the danger of an overly rigid distinction between Lean and Agile, noting that Lean "may provide even greater value by incorporating some aspects of agile manufacturing." Finally, however, even if the cases represented in Figure 1 represent a continuum of Lean from a scientific standpoint, it could still be useful to use different titles (Lean and Agile?) in industrial settings as a way to focus attention on different aspects of the system. Evaluating the utility of such titles requires a behavioral perspective, which we introduce in the fourth Lean Lens.

Semantics aside, the Flow Lens is of great value to researchers because it describes Lean in a way that makes it amenable to the vast range of probabilistic modeling tools in the OM arsenal. It is also highly useful to practitioners who have already taken elementary steps, such as using value stream mapping and 5S to root out obvious waste and are now seeking to find and eliminate more subtle forms of indirect waste.

4 | THE NETWORK LENS

The third definition of Lean is "a systematic process for reducing the cost of waste." Unlike the first two definitions, which help us to enumerate types of waste, the focus on a systematic process forces us to think about where and how to reduce waste to achieve maximum cost efficiency.

In a production line or a simple supply chain, the number of choices may be sufficiently limited to allow the Flow Lens to achieve much of the potential of Lean. But most production and service systems consist of many interconnected flows. Examples include a plant with thousands of product routings that share processes and people, a supply chain in which producers use multiple suppliers and suppliers serve multiple customers, and a service system (e.g., a hospital) that serves many types of customers (patients) that require different but overlapping sets of resources. All of these can be represented conceptually as networks of flows, so we label this the *Network Lens*.

One approach for addressing complex flow networks is to separate and simplify the flows, and then apply the Process or Flow Lenses to identify improvement options. This is essentially what cellular layouts, supplier consolidation and focused factories do. While appealing where practical, it can sometimes be very expensive to break up complex networks into separate cells or factories because this can require duplicate capacity. Hence, to develop a systematic process for reducing the cost of waste, we need a means for diagnosing and improving complex flow networks directly.

The core concept for understanding the behavior of networks of flows, and thereby identifying points of maximum leverage, is that of a bottleneck, which we define as follows (Hopp & Spearman, 2008, p. 231):

> ***Bottleneck Definition***: *The bottleneck in a production or service network is the resource (node) with the highest long-term utilization.*

Queueing theory tells us that bottlenecks will produce the majority of backups and delay (inventory and waiting waste). The intuition behind this is that the closer a resource is to full utilization, the smaller the fluctuation in demand or capacity needed to overload the resource. Small and transitory overloads that happen with high frequency cause normal queueing behavior (e.g., stores of work-in-process inventory in a factory or lines of patients waiting in a hospital). Large and extreme overloads that happen rarely cause major backups and shortages (e.g., jet planes grounded in airports or people waiting for emergency supplies in a hurricane). Regardless of time scale, bottlenecks must be key focal points in a systematic effort to reduce the cost of waste.

Furthermore, queueing theory also tells us the effect of utilization is nonlinear (exponential), while the effect of variability (as measured by squared coefficient of variation) is linear. This implies that we should first look for ways to reduce utilization (e.g., by adding capacity and/or eliminating unnecessary demand, such as that from rework) and then look for ways to reduce variability (e.g., by smoothing flow into bottlenecks and/or reducing variability in bottleneck processes).

Finally, in addition to bottleneck and queueing logic, we need to assess costs and benefits of candidate improvements, as we would in development of any business case. Taken together, these tools comprise the conceptual basis for the Network Lens of Lean.

In the world of practice, this perspective has appeared in various ways. The best-selling book *The Goal* (Goldratt & Cox, 1986) highlighted the importance of bottlenecks but in the context of simple flows rather than in complex networks. Schmenner and Swink (1998) brought the concepts of bottlencks and variability together in the theory of swift, even flow, and Yin et al. (2017) applied this theory to the Japanese *seru* system to describe how it can elevate both efficiency and responsiveness. Even basic Lean tools, such as Value Stream Mapping and Fishbone/Ishakawa/Cause-and-Effect/Five Whys diagrams, which are useful in breaking down and assessing complex environments, are more effective when paired with the bottleneck/queueing/cost–benefit analysis perspective of the Network Lens.

One reason for the gap between the conceptual perspective of the Network Lens and the Lean tools used to address complexity in systems is that much of the academic literature related to bottleneck and queueing analysis assumes stationarity. That is, over the long run resource utilizations are stable, which implies that a unique bottleneck (or at least a small set of bottlenecks and near bottlenecks) will govern system behavior. But in practice, it is possible, even likely, that neither demand nor capacity are stationary. Demand may be predictably seasonal, subject to grow, disrupted by occasional shocks, or affected by other factors. Capacity may be altered by learning, staffing changes, investments or other factors. As a result, bottlenecks may shift over time or even be in constant flux. In such cases, static capacity calculations and queueing analyses may provide little useful guidance.

This suggests that there may be a need for more dynamic representations of production and service networks that make use not only of bottleneck and queueing analysis, but also of real-time data and network-structure analysis to evaluate vulnerabilities and effectiveness of improvement options. An example of research that makes use of a data-driven network approach to analyze and improve a production or service system include Gokpinar, Hopp, and Iravani (2010), who used networks to represent product designs and organizational communication patterns to identify waste inducing mismatches. An example of a real-world attempt to incorporate a dynamic, system-wide perspective into flow management of a complex system is the electronic dashboard adopted by the Johns Hopkins Hospital (Martinez et al., 2018). Leveraging research from queueing, network theory, data analytics and machine learning in dashboards like this may open opportunities for Lean execution that have so far eluded systems as complex as those in global product design systems and research hospitals. Because these system-wide efficiency initiatives require much more sophistication to pursue than the direct waste elimination activities promoted by the Process Lens or the indirect waste elimination via variability reduction and buffering promoted by the Flow Lens, the Network Lens is generally the purview of advanced Lean practitioners.

5 | THE ORGANIZATION LENS

The fourth definition of Lean is “an organizational culture that encourages continual reduction of the cost of waste.” The focus on organizational culture forces us beyond the physics of flows that were at the center of the first three Lean Lenses. Although a physics focus might be sufficient for a production or service system run entirely by machines (e.g., a true “lights out” factory), the reality is that all business systems involve people. To account for this, we require the fourth and most expansive perspective on Lean which we term the *Organization Lens.*

Some of the Lean practice literature has recognized the need for an organizational perspective. For example, Spear and Bowen (1999) noted that a Lean culture requires attention to training, communication and other human behaviors within the organization. Liker (2004) emphasized developing people and organizational learning in his 14 principles of the TPS. While these are valid observations about the culture at Toyota, they do not leverage modern behavioral science. In order make full use of the Organization Lens we need to incorporate human behavior more scientifically.

A few authors (e.g., de Treville & Antonaki, 2006; Hopp, 2018; Parker, 2003) have recognized that the literature on the psychology of work is relevant to Lean because it is ultimately implemented by people with jobs. Designing those jobs to make work rewarding and motivational can have a significant influence over the success of Lean implementations. But, so far, operations scholars have made only limited use of the insights from psychology.

A more extensive, and largely untapped, resource is the wide array of cognitive research into heuristics and biases that has been developed by behavioral and decision scientists since the 1970s. Kahneman (2011) uses the concepts of "System 1" and "System 2" to describe this body of work. System 1 ("fast thinking") refers to the involuntary, automatic, effortless responses our brains make to the world around us.[8] System 2 ("slow thinking") refers to the methodical, logical and energy intensive responses we make to selected situations. Because each of us makes some 35,000 decisions every day, it is impossible for us to use System 2 thinking for more than a tiny fraction of our choices. Fortunately, because it has been shaped by evolutionary forces, System 1 usually works well. But, because our world differs from the one in which it evolved or for other reasons, System 1 heuristics sometimes lead to unconscious bias or other irrational decisions.

Recent research shows that the occasional irrationality from System 1 thinking is often predictable (see for example, Ariely, 2008). A few common biases that affect our everyday thinking and almost certainly come into play when people make decisions about the design and operation of Lean systems include:

- *Hindsight Bias*—We judge decisions by outcome rather than soundness of method. For example, suppose a bakery owner is clueless about how much bread to bake on the day before Thanksgiving and uses guesswork and poor logic to justify making a very large amount of bread. If she gets lucky with an unusually high demand day, she is apt to say "See! I knew it all along" instead of recalling the uncertainty and controversy surrounding her decision.
- *Confirmation Bias*—We look for evidence that confirms our beliefs. For example, a bakery owner who, because of a past lucky decision or another reason, feels the bakery should be making more bread each day will notice and remember the days that confirm her belief more than the days that refute it.
- *Loss Aversion*—The motive to avoid losses is stronger than the motive to achieve gain. For example, a bakery owner may be more upset by a day on which many loaves of bread are thrown away than a day on which bread ran out by mid-afternoon. The former is a measurable loss, since expensive ingredients were wasted. The latter is a foregone gain (the bakery could have sold more bread and made more profit). If the owner over emphasizes the loss and under emphasizes the gain, she will order less than the profit maximizing amount of bread. Indeed, experimental evidence has shown that people have a tendency to do precisely this, provided the profit margin is sufficiently high (Schweitzer & Cachon, 2000).[9]

If the hindsight bias extended by confirmation bias leads the bakery to make too much bread it will result overproduction waste. If the loss aversion bias leads the bakery to make too little bread it will lead to the waste of excess waiting on the part of customers who must come back another day if they want to get bread. The time buffer represented by this waiting could be particularly costly, because it may frustrate customers and decrease their likelihood of making future purchases. Either way, the result will be a wasteful departure from the goal of Lean.[10]

A possible implication from the research into psychological biases is that people need more training in probabilistic thinking to deal with problems involving uncertainty. Another conclusion could be that planning under uncertainty in many situations should not be done by intuitive feel at all, but instead should make use of a data-based decision support system. A small business like a bakery has neither statistically trained personnel nor a decision support system. As a result, they presumably make errors like the one described above routinely. The "newsvendor" scenario presented by the bread planning example is a very simple case where biases lead to waste.

Large organizations using sophisticated ERP systems presumably have the computational power to address the newsvendor aspect of their inventory planning problems. But even they often use highly visible sales to represent demand, while omitting the invisible lost sales. As a result, decisions such as inventory planning, which are subject to uncertainty, may be compromised by decision biases in large, as well as small, firms.

Subtler situations related to Lean implementations where incentives and/or decisions may also be distorted by similar biases include choosing training activities to carry out, choices made during kaizen events, and conclusions drawn from Gemba walks. In each of these, we are prone to focus on outcomes rather than processes, see evidence that we want to see, and inconsistently emphasize avoiding losses at the expense of passing up gains.

We see examples of faulty decision making in Lean implementations all the time. However, as is always possible when uncertainty is involved, sometimes bad (i.e., not based on logical consideration of the information available) decisions lead to good outcomes. When this occurs, the two of us are fond of shaking our heads and saying "oh well, it's better to be lucky than smart." But, while this is tautologically true if lucky means getting good results and smart means making well-reasoned decisions, it may not be true in real life when we consider cognitive biases and their impacts on future decisions. As Kahneman points out, even a single observation can create an associative memory that System 1 can mistake for a pattern. This plus confirmation bias and hindsight bias can result in a lucky outcome distorting decision making for a long time. For example, if the bakery owner makes a bad decision and bakes too much bread but gets lucky and sells it, the associative memory reinforced by confirmation bias could lead to many more bad decisions to overbake in the future.

These and other biases can certainly affect management decision making related to Lean, such as choosing production quantities, incentives, or motivational activities. But they can also affect behavior in the execution of Lean. An illustrative example occurred in the bread-slicing operation of a small family-owned bakery. The basic bread production process is to bake, cool, slice, bag, and sell the bread. However, because some customers prefer their bread unsliced and there is uncertainty about the proportion of customers who will want unsliced bread on any given day, there is an inherent challenge in matching bread to customer preferences. In this particular bakery, this challenge had been addressed for as long as anyone could remember by bagging unsliced bread once it was cool and then un-bagging and slicing it for any customer who wanted sliced bread. This resulted in extra work for the staff to unbag and re-bag the bread (processing waste), as well as delay while the customer waited for the bread to be sliced (waiting waste). Moreover, because it required slicing one loaf at a time, rather than slicing all the loaves in a repetitive batch mode, it involved even more processing waste due to the loss of an economy of scale.

Recognizing this, the new store owner suggested slicing at least some of the bread before bagging it. This led to a surprisingly difficult discussion with the staff. Some of them focused on specific customers who strongly preferred unsliced bread (What if Mrs. Smith cannot get her unsliced loaf!). Others agreed that slicing some of the bread made sense but could not agree on how much. When asked what percentage of customers wanted their bread sliced, the staff could not say with any precision but agreed that it was well over half. Nevertheless, it took a good deal of cajoling and reminding to get the staff to agree to slice half of the bread bagging as an experiment. When it quickly became apparent that this policy never resulted in a stockout of unsliced bread but almost always required un-bagging and slicing some of the unsliced bread, the staff recognized on their own the need to increase the proportion of bread sliced before bagging. They soon converged on a policy of slicing 80% of the bread and agreed that the new system was clearly better for them and for the customers.[11]

The point of this example is two-fold. First, even in very simple execution-level activities, uncertainty can present decision problems for which our System 1 thinking processes are ill-equipped and prone to errors. Second, implementing Lean improvements requires more than laying them out in logical fashion for people to adopt. People need to be engaged in the implementation in a way that engages their System 2 thinking and overcomes their System 1 biases. This may be why the Lean literature is replete with claims (Durin, 2018; Liker & Rother, 2011) that implementation failures are due to factors such as lack of involvement by top management, unclear goals, or inadequate training. What these plausible but ill-defined reasons may really be saying is that the Lean implementations failed to take into account the way real people think and behave. It may also be behind the success of bottom-up problem-solving processes like Kaizen (Shingo, 2007) that engage those responsible for implementing Lean policies in the design of them. However, we need to be careful about putting too much faith in the simple act of involving workers in designing improvements. We are all prone to cognitive biases that can blind us to effective alternatives. Therefore, how we involve people in the search for ways to implement Lean, what data are provided, how questions are posed, and how well people are prepared to think about problems that involve uncertainty, all matter in the effectiveness of bottom-up problem solving.

Psychological biases may even influence the effectiveness of continual improvement processes like Kaizen. For example, the bias toward loss aversion can lead people to avoid setting high-achieving goals. An unachieved goal is a loss, while exceeding a goal is a gain. Hence, a low goal makes it easier to avoid a painful loss, albeit at the expense of a reduced likelihood of the satisfaction of a high-level outcome. Moreover, once a goal is set, people have more incentive to reach it than to exceed it. Kahneman (2011) explains that in many cases, people will reduce their efforts once they have reached a specific goal because they see no point going above and beyond.

The loss aversion bias makes us more concerned about avoiding losses than about striving to achieve gains.

Finally, underappreciating the nature of System 1 and the biases it introduces can lead management to conclude that implementing a new policy should be simpler than it actually is. Managers often assume that if the logic of a new way of doing things is manifest then this should be enough to get people to adopt it (e.g., how could anyone not appreciate the efficiency benefits of slicing bread before bagging to avoid unnecessary un-bagging and re-bagging?). But this neglects the inertia created by associative memory and confirmation bias. As in the case of the bread slicing and bagging example, an effective Lean implementation plan needs to be designed in a way that encourages staff to engage their System 2 thinking processes and overcome their unconscious biases.

Because little has been done to make use of cognitive science in academic studies of Lean, and almost none of the insights from behavioral science have been incorporated into Lean implementations in industry, there is a huge opportunity to leverage the insights about psychological biases in both Lean research and practice.

6 | LINKING THE LENSES

Each of the above lenses can serve as a research perspective on Lean and can lead to insights that help guide Lean practice. But fully embracing Lean as part of a comprehensive change management program requires more than making use of individual lenses. It also requires sequencing and integrating the perspectives of the lenses in a way that facilitates continual improvement. To illustrate we invoke the famous "two-shift" practice of Toyota.

In the 1980s when almost all auto manufacturers ran their assembly plants on a three-shift basis, Toyota made use of a two-shift system consisting of two 8-hr shifts per day separated by two 4-hr preventive (PM) maintenance periods. On the surface, it appeared that Toyota was sacrificing 8 hr/day of capacity in very expensive facilities. However, when viewed through the lenses of Lean, it becomes clear that they were doing something much more profound.

A key difference between Toyota's practices and those of their contemporaries and many firms implementing Lean today is the sequence in which they employed the lenses. Other firms have tended to start with a Process Lens focus on obvious waste (e.g., with 5S reorganizations of individual processes), progress to a Flow Lens focus on simple flows (e.g., by using Value Stream Mapping to identify inefficiencies), and finally adopt a Network Lens focus on the overall production system (e.g., by reorganizing into modular layouts. Most firms never get to the Organization Lens. This is probably not too surprising given that the Lean literature is devoted primarily to Process Lens descriptions, with some coverage from the Flow and Network Lens perspectives and almost entirely lacking in an Organization Lens focus.

In contrast, Toyota adopted a Network Lens focus on flow simplification very early on. In the assembly plants using the 2-shift schedule, this led to mixed-model production to simplify flow in final assembly and cellular manufacturing in component lines (see Hall, 1983; Schonberger, 1982; Shingo, 1985 for a description of these and other TPS methods). Spear and Bowen (1999) described this emphasis on reducing complexity in flows as the DNA of the TPS. The first three of the four rules they used to encapsulate TPS were:

> *Rule 1 – All work shall be highly specified as to content, sequence, timing, and outcome.*
> *Rule 2 – Every customer-supplier connection must be direct, and there must be an unambiguous yes-or-no way to send requests and receive responses.*
> *Rule 3 – The pathway for every product and service must be simple and direct.* (Spear & Bowen, 1999)

Rules 1 and 2 are prerequisites to the simple flows called for in Rule 3. Note that Spear and Bowen gave this Network Lens perspective primacy over the Process Lens emphasis on waste reduction and Toyota tools (Kanban and 5S).

Toyota also adopted an early Flow Lens focus on variability reduction. In particular, they made clever use of their variability-buffering strategy to make variability visible and only then made use of Process Lens methods to eliminate the variability and the waste it causes. The two-shift strategy was essential to this stage of Toyota's Lean evolution because, in addition to providing time for maintenance, the PM periods could be used for overtime if needed to meet the daily production quotas. As such, the PM periods represented a very large capacity buffer against variability in the production or demand rate. This permitted Toyota to reduce inventory (raw materials, work-in-process, finished goods) buffers without inflating time buffers (delays in meeting customer demands).

Reduced inventory helped to identify sources of variability in the production system, as Toyota described with the well-known analogy of lowering the water in a river to uncover the rocks. In less poetic terms, without excess WIP in the system, each glitch in the system would be immediately apparent because of the disruption in flow it would cause. Toyota further highlighted glitches with an

Andon system that allowed workers to stop the line whenever they encountered a problem either with the quality of the product or with the procedure being used. To torture the water analogy a bit, the Andon system was analogous to a lookout on the ship watching for rocks, while the available PM periods were analogous to having the capability to dig channels around the rocks so that they could be uncovered without harming the ship (factory) or passengers (customers).

With this systematic process for making problems visible in place, all of the Process Lens techniques become more effective. Without it, efforts at waste reduction can be "for show rather than dough." That is, as we have seen many times, Lean initiatives can lead to inventory reductions that have little (or even negative) impact on costs, capacity enhancements that affect only non-bottlenecks, or internal flow enhancements that do not improve customer service, all of which do not serve the core strategy of the organization. Making problems visible helps waste seekers find the strategically important waste. The fact that Toyota premised their direct waste elimination efforts with production-network rationalization and flow-variability reduction phases may be an important reason they were able to practice Lean more effectively than most of their rivals for decades. Furthermore, in the early 2000s, after decades of variability reduction, Toyota evidently concluded they no longer needed large capacity buffers to protect customer deliveries and began switching assembly plants to three-shift schedules.

Finally, although neither Toyota nor anyone else has fully exploited the behavioral focus of the Organization Lens, Toyota has always devoted time to the culture of Lean.[12] For example, they use kaizen as a way to gain worker buy-in to the continual improvement process and to inject a worker perspective into any job-design decisions made as part of the Lean transformation.

Figure 2 provides a high-level schematic of the staging of lenses in the Toyota scenario. Note that we have included reverse arrows to allow for repeated cycling through the various perspectives. The reason for this is that, given persistent change in product volumes, mix, customer preferences, technology, and many other factors, it is impossible to permanently optimize a production system. A continually evolving and adapting system is needed. To achieve this, it is essential to punctuate ongoing Process Lens vigilance with periodic reviews of the bigger pictures provided by the Flow and Network Lenses.

An example of this type of iteration was on display when one of us took a sabbatical at Motorola. The company had recently announced completion of their celebrated Six Sigma initiative and had launched a "10X Cycle Time" initiative with a goal to radically reduce (by a factor of 10) their design, production, and fulfillment cycle times. This ambitious goal forced the firm to eliminate process stages, and even entire processes. When this became apparent, a Vice President remarked that he wished they had done 10X Cycle Time before Six Sigma, because they could have avoided a great deal of (Process Lens) improvement work on process that were removed through the (Network Lens) review of the 10X program. But he may have been holding himself to too high a standard. Anticipating the need for speed at such a detailed level a decade earlier would have required remarkable foresight. In a changing world iteration will always be necessary.

We can summarize the insights from this discussion captured in Figure 2 as:

1. *Think big before thinking small.* A system review that simplifies the production network and a flow focus that identifies bottlenecks helps to focus waste elimination efforts on processes that matter to overall performance.
2. *Adjust buffers to facilitate exploration and exploitation.* Optimizing variability buffers is a vital part of Lean implementation. But buffers can also be adjusted to reveal the sources of variability. This is done by reducing WIP buffers, so that disruptions affect flow quickly, while increasing other buffers (e.g., capacity or finished goods) to protect customer service.

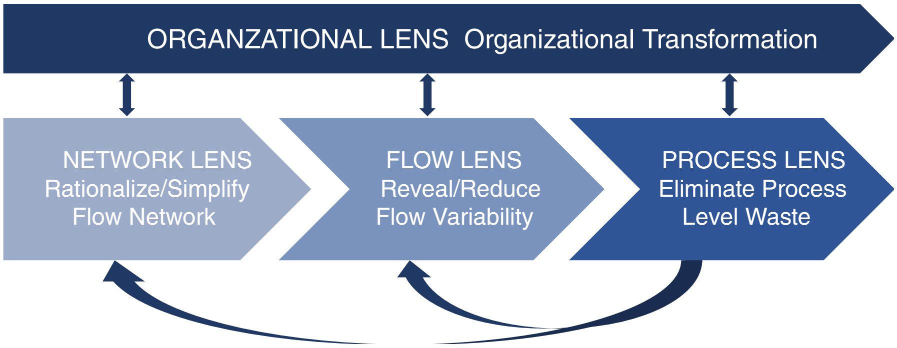

FIGURE 2 Staged use of lean lenses in practice [Color figure can be viewed at wileyonlinelibrary.com]

3. *Pursue physics and psychology in parallel*. Each step in the Lean process must be executed by people and each change that is made affects someone's job. It is therefore essential to think about both the physics of flows (variability, bottlenecks, etc.) and the psychology of work (job design, motivation, cognitive bias, etc.) at each stage of Lean implementation.
4. *Iterate, iterate, iterate*. Although Lean as a title may have an expiration date, the pursuit of efficiency is a never-ending journey. Production systems must be adapted constantly to changes in the market and production environments and organizational cultures must evolve to support and accommodate these changes.

Although the schematic in Figure 2 is meant primarily as a practice guide, it can also serve as a basis for identifying research opportunities. For example, Lean scholars could make empirical studies of the effectiveness of different activity sequences in Lean implementations to provide clearer guidance to firms adopting Lean. They could study the impact of WIP level on organizational learning to see whether the approach used by Toyota with their two-shift system is broadly effective. They could examine the impact of cognitive bias on specific Lean practices as a prelude to finding better ways to coordinate the physical and psychological sides of Lean. Many other research opportunities are likely to arise from examining Lean through the combined perspectives of the four Lenses.

7 | CONCLUSIONS

Explicitly addressing Lean progressively through the Process, Flow, Network, and Organization Lenses can help practitioners from getting stuck at the lowest level of Lean execution. It is unfortunate that so much press has been given to the 7 types of waste and specific practices such as Kanban and 5S. Neither categories of waste nor individual practices lead to an understanding of the underlying causes of waste. The “5 Whys” approach from Toyota seeks to expose the primary causes of waste, as do more formal techniques like fault-tree analysis, root-cause analysis and fishbone diagramming. But all of these are generic methods, which depend on a knowledge of the behavior of flows on the part of the user to be effective in diagnosing waste in production and service systems.

To provide explicit guidance for Lean implementation we need to go beyond classification and practices implied by the Process Lens. This is what the Lenses introduced in this paper do. The Flow Lens leverages the science of variability to describe the causes of waste in a flow. The Network Lens makes use of a network representation of a production or service system, and particularly the concept of bottlenecks, to understand causes of waste and to identify areas of maximum leverage in complex systems of interconnected flows. The Organization Lens draws on behavioral science to anticipate human reactions to Lean policies and to guide designs that work with human tendencies rather than against them.

Taken together, the four Lenses offer a perspective on Lean research that leverages rapidly advancing fields like Data Analytics, Network Science, and Cognitive Psychology, and extends the range of questions amenable to empirical and analytical studies. The Lenses also offer a framework for enhancing Lean management systems by making of the insights from these fields, and by balancing implementation efforts across the levels of the enterprise from process mechanics to organizational culture. As such, these Lenses can help both scholars and practitioners see the opportunities of Lean more clearly and promote the pursuit of efficiency more effectively.

ORCID

Wallace J. Hopp https://orcid.org/0000-0003-1586-8606

ENDNOTES

[1] The separation of OM sub-fields by strategic focus does not mean they do not overlap. For example, a central tool of efficiency management is variability reduction and an important means of variability reduction is quality management. Hence, efficiency and quality management are often pursued in concert, as evidenced by the blurring of the two under the popular heading of *Lean Six Sigma*.

[2] Although this buffering framework was introduced in the 1990's as part of the Factory Physics framework, Taiichi Ohno, the father of the Toyota Production System, recognized the central role of variability as a cause of waste much earlier in his tri-dimensional description of waste as muda (waste), mura (variability) and muri (overburden) (see Hopp, 2018 for a discussion).

[3] If you asked us to give the single most useful definition of Lean this would be it. By leveraging the most powerful insights of Factory Physics into the nature and causes of waste, it provides a robust modeling framework for research and a practical diagnostic tool for practice. However, tempered by the mantra “all models are wrong, but some are useful”, we recognize that alternate definitions can be useful too. Therefore, we present this definition as only one among four, rather than as the Holy Grail of Lean.

[4] Note that variability propagates across measures. For instance, variability in quality impacts variability in the time and quantity of production by impacting yield and rework. Variability in the time and quantity of output results in variability of costs and revenues. Consequently, variability is an important connection between operational and financial metrics.

[5] Many of the quoted lead times would also be 4 weeks or less and hence, unlike the fixed 10-week lead times, would be competitive with those of the rival firm.

[6] A clever example of creating the institutional readiness to implement an emergency response is the "chaining" structure proposed by Jordan and Graves (1995). By equipping factories with overlapping capabilities, this system enables adjustment of product-to-plant assignments to accommodate wide fluctuations in demand mix. Although few would describe this as a Lean initiative, it facilitates meeting demand with less installed capacity, which is clearly an efficiency measure. So, in the terminology of this paper, it is a means to Lean.

[7] Another example of a contingency plan is the tailored base-surge policy proposed by Allon and Van Mieghem (2010) in which an efficient supplier is used to provide a constant supply, while a responsive (but more expensive) supplier is used to meet surges in demand. de Treville, Cattani, and Saarinen (2017) address the challenge of the responsive supplier by describing how producing a portfolio of time-sensitive and time-insensitive products can make effective use of capacity and allow competitive production in a high-cost environment. Bakeries use a version of the tailored base-surge approach for commonly available items (milk, eggs, chocolate chips) with a wholesale supplier as the efficient supplier and a grocery store as the responsive supplier.

[8] Gigerenzer and Goldstein (1996) coined the more descriptive "fast and frugal" label for System 1 thinking in order to emphasize that it relies on heuristics that limit both information search and computation.

[9] Experiments indicate that people tend to over-order in settings like that of the bread baking scenario when the profit margin is low, which was originally thought to violate the behavior predicted by prospect theory. However, Long and Nasiry (2015) pointed out that if the decision maker defines gains and losses relative to an aspiration level other than the status, as Kahneman and Tversky (1979) allowed for, then the observed behavior is consistent with prospect theory.

[10] If bread seems to mundane an example of how cognitive biases can distort operational decisions, consider the work of Gray, Esenduran, Rungtusanatham, and Skowronski (2017), which studied companies that offshored their production to reduce cost only to reverse their decisions and bring production back home at great expense. By scrutinizing these decisions in light of the research on heuristic decision making, the authors concluded that the firms fell prey to well-known decision biases that led them to rely on an overly simplistic "lowest per-unit landed-cost" that omitted important cost considerations such as quality issues and intellectual property loss.

[11] The cognoscenti are likely to point out that finding an optimal percentage of bread to pre-slice is a Newsvendor problem that requires the (underage) un-bagging and re-bagging cost to slice and unsliced loaf, the (overage) cost of a disappointed customer who wants an unsliced loaf but must accept a sliced loaf or go without, and the distribution of the demand for unsliced bread. But this is probably too much System 2 thinking to devote to this problem when there are bigger problems (e.g., determining bake quantities and optimizing the advertising strategy) that can generate more value from analytic solutions.

[12] Of course, Toyota does not use the generic term Lean. Instead, their culture building initiatives are made under the eponymous Toyota Production System banner. It is not clear that Lean, whether defined broadly as efficiency management or more narrowly as some subset of the efficiency space, can serve as a motivational title. It may be more effective in practice for firms to create their own titles ("X Production System" or "Y Business System") to describe a collection of Operations Management ideas and practices (from the Lean/efficiency area, as well as other OM subfieklds) combined with insights from other management fields (e.g., strategy, marketing, organizational behavior), and firm-specific goals and practices. Whether firm-specific titles are more motivational than generic terms is an open research question.

REFERENCES

Allon, G., & Van Mieghem, J. (2010). Global dual sourcing: Tailored base-surge allocation to near- and offshore production. *Management Science*, *56*(1), 110–124.

Ariely, D. (2008). *Predictably irrational*. New York, NY: HarperCollins.

Belekoukiasa, I., Garza-Reyesb, J., & Kumarc, V. (2014). The impact of lean methods and tools on the operational performance of manufacturing organisations. *International Journal of Production Research*, *52*(18), 5346–5366.

Browning, T., & Heath, R. (2009). Reconceptualizing the effects of lean on production costs with evidence from the F-22 program. *Journal of Operations Management*, *27*(1), 23–44.

de Treville, S., & Antonaki, J. (2006). Could lean production job design be intrinsically motivating? Contextual, configurational, and levels-of-analysis issues. *Journal of Operations Management*, *24*(2), 99–123.

de Treville, S., Cattani, K., & Saarinen, L. (2017). Technical note: Option-based costing and the volatility portfolio. *Journal of Operations Management*, *49-51*, 77–81.

Durin, M. 2018. Why most companies are failing with Lean implementation? Kaizen Institute Blog. Retrieved from https://www.kaizen.com/blog/post/2018/07/05/why-are-most-companies-failing-with-lean-implementation.html

Ford, H. (1922). *My life and work*. New York, NY: Doubleday, Page & Co.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Gilbreth, F. (1911). *Motion study*. New York, NY: D. Van Nostrand Co.

Goldratt, E., & Cox, J. (1986). *The goal: A process of ongoing improvement*. New York, NY: North River Press.

Gokpinar, B., Hopp, W., & Iravani, S. (2010). The impact of misalignment of organization structure and product architecture on quality in complex product development. *Management Science*, *56*(3), 468–484.

Gray, J., Esenduran, G., Rungtusanatham, M., & Skowronski, K. (2017). Why in the world did they reshore? Examining small to medium-sized manufacturer decisions. *Journal of Operations Management*, *49-51*, 37–51.

Hall, R. (1983). *Zero inventories*. Homewood, IL: Dow Jones-Irwin.

Hopp, W. (2011). *Supply chain science*. Long Grove, IL: Waveland Press.

Hopp, W. (2018). Positive lean: Merging the science of efficiency with the psychology of work. *International Journal of Production Research*, *56*(1–2), 398–413.

Hopp, W., & Spearman, M. (2004). To pull or not to pull: What is the question? *Manufacturing & Service Operations Management*, *6*(2), 133–148.

Hopp, W., & Spearman, M. (2008). *Factory physics* (3rd ed.). Long Grove, IL: Waveland Press.

Jordan, W., & Graves, S. (1995). Principles on the benefits of manufacturing process flexibility. *Management Science*, *41*(4), 577–594.

Kahneman, D. (2011). *Thinking, fast and slow*. London, England: Penguin Books.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Krafcik, J. (1988). Triumph of the lean production system. *Sloan Management Review*, *30*(1), 41–52.

Lean Enterprise Institute. (2019). What is lean? Retrieved from https://www.lean.org/WhatsLean/

Liker, J. (2004). *The Toyota way: 14 management principles from the world's greatest manufacturer*. New York, NY: McGraw-Hill.

Liker, J., & Rother, M. (2011). Why lean programs fail. Lean Enterprise Institute. Retrieved from https://www.lean.org/Search/Documents/352.pdf

Long, X., & Nasiry, J. (2015). Prospect theory explains newsvendor behavior: The role of reference points. *Management Science*, *61*(12), 3009–3012.

Martinez, D. A., Kane, E. M., Jalalpour, M., Scheulen, J., Rupani, H., Toteja, R., … Levin, S. R. (2018). An electronic dashboard to monitor patient flow at the Johns Hopkins Hospital: Communication of key performance indicators using the Donabedian model. *Journal of Medical Systems*, *42*(8), 133.

Moraro, J., Lemstra, M., & Nwankwo, C. (2016). Lean interventions in healthcare: Do they actually work? A systematic literature review. *International Journal for Quality in Health Care*, *28*(2), 150–165.

Lopes Negrão, L., Godinho Filho, M., & Marodin, G. (2017). Lean practices and their effect on performance: A literature review. *Production Planning & Control*, *28*(1), 33–56.

Parker, S. (2003). Longitudinal effects of lean production on employee outcomes and the mediating role of work characteristics. *Journal of Applied Psychology*, *88*(4), 620–634.

Schmenner, R., & Swink, M. (1998). On theory in operations management. *Journal of Operations Management*, *17*, 97–113.

Schonberger, R. (1982). *Japanese manufacturing techniques: Nine hidden lessons in simplicity*. New York, NY: Free Press.

Schweitzer, M., & Cachon, G. (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, *46*(3), 404–420.

Shingo, S. (1985). *A revolution in manufacturing: The SMED system*. Portland, OR: Productivity Press. (Original Japanese edition, *Shinguru Dandori*, published by the Japan Management Association, Tokyo, 1978).

Shingo, S. (2007). *Kaizen and the art of creative thinking*. Bellingham, WA: Enna Products Incorporated. (Original Japanese edition published by Shobo Publishing Company, Tokyo, 1959).

Spear, S., & Bowen, H. (1999). Decoding the DNA of the Toyota production system. *Harvard Business Review*, *77*(5), 96–106.

Spearman, M. L., & Hopp, W. J. (2019). Fundamentals of operations science. *Working paper, University of Michigan, Ross School of Business.*

Sugimori, Y., Kusunoki, K., Cho, F., & Uchikawa, S. (1977). Toyota production system and kanban system materialization of just-in-time and respect-for-human system. *International Journal of Production Research*, *15*(6), 553–564.

Suri, R. (1998). *Quick response manufacturing: A companywide approach to reducing Lead times*. Portland, OR: Productivity Press.

Taylor, F. W. (1911). *The principles of scientific management*. New York, NY; London, England: Harper & Brothers.

Ulhassan, W., von Thiele Schwarz, U., Thor, J., & Westerlund, H. (2014). Interactions between lean management and the psychosocial work environment in a hospital setting – A multi-method study. *BMC Health Services Research*, *14*, 480–490.

Vidal, M. (2007). Lean production, worker empowerment, and job satisfaction: A qualitative analysis and critique. *Critical Sociology*, *33*, 247–278.

Womack, J., & Jones, D. (1996). *Lean thinking: Banish waste and create wealth in your corporation*. New York, NY: Simon & Schuster.

Womack, J., Jones, D., & Roos, D. (1991). *The machine that changed the world: The story of lean production*. New York, NY: Harper Perennial.

Yin, Y., Stecke, K., Swink, M., & Kaku, I. (2017). Lessons from *seru* production on manufacturing competitively in a high cost environment. *Journal of Operations Management*, *49-51*, 67–76.

Zarbo, R., & D'Angelo, R. (2006). Transforming to a quality culture: The Henry Ford Production System. *American Journal of Clinical Pathology*, *126*(Suppl 1), S21–S29.

Zipkin, P. (2000). *Foundations of inventory management*. New York, NY: McGraw-Hill.

How to cite this article: Hopp WJ, Spearman MS. The lenses of lean: Visioning the science and practice of efficiency. *J Oper Manag*. 2021;67:610–626. https://doi.org/10.1002/joom.1115