LEAN FLOW

A quick guide to transform with lean digital

ELISA GRANHA LIRA

To the processes and people who transform us and make us flow. In particular, to my husband Rafa and my son Eduardo.

THE SLOWER BUT CONSISTENT TORTOISE CAUSES LESS WASTE AND IS MUCH MORE DESIRABLE THAN THE SPEEDY HARE WHO RACES AHEAD AND THEN STOPS OCCASIONALLY TO DOZE. THE TOYOTA PRODUCTION SYSTEM CAN BE REALIZED ONLY WHEN ALL THE WORKERS BECOME TORTOISES.

TAIICHI OHNO

ABOUT THE AUTHOR

Elisa Granha Lira is a professor of the Industrial Engineering Program at the Federal University of Rio de Janeiro (Universidade Federal do Rio de Janeiro - UFRJ) in Macaé. Based on her experiences in the industry as a Lean Engineer, she works today to extend lean philosophy beyond the university to transform people and organizations.

PREFACE

First, I believe it is essential to point out the objective of this book: to reach people from the most diverse backgrounds in accessible language. After all, we are all part of the digital transformation process that, directly or indirectly, has been revolutionizing work in both public and private organizations.

And what is lean manufacturing, lean production, or simply lean? Or, going more to the heart of the matter: what is lean digital? I highlight these questions because some readers may be wondering about this right now.

Therefore, before anything else, it is important to clarify that this book is not a dictionary, a step-by-step, etc. In a world where technology does the legwork, just Google it or ask ChatGPT, and better answers will surely be obtained to these superficial questions of knowledge.

My interest is in working on the submerged part of the iceberg, the part that, from my theoretical and practical experience, is the state of the art accessible to few people. Therefore, I immediately invite the reader to the space of knowledge construction in which we will develop these concepts in a didactic process throughout the book.

Starting off, then, let's work on the concept of lean. If you were to ask lean experts around the world what lean is, you would probably get different answers. For example:

"It is a waste reduction philosophy based on the Toyota Production System."

"It is a management system that seeks to increase efficiency and productivity by reducing errors in industrial production."

"It is a production system in which the customer's needs drive what must be produced."

These answers are not wrong, but they are clippings from different angles of lean manufacturing.

Lean basically involves three levels of abstraction (Figure 1). At a higher level of abstraction, it can be defined as a philosophy of production, that is, as a way of being and thinking. It is, after all, a philosophy aimed at developing a culture and values in which all employees naturally collaborate in continuous improvement in their day-to-day work through waste elimination.

At an intermediate level of abstraction, we can see lean as a method of planning and control: production system advocated (pull flow), that is, the customer's need must pull what should be produced; kanban control; production leveling; and stream sync.

Finally, lean is also a toolbox at a more tangible level, which is not necessarily exclusive to this management philosophy.

I think it is important to emphasize the existence of three levels to make it clear that, in this book, we will work with the highest level of abstraction, that is, the philosophical definition of lean, which aims to transform our way of thinking, acting, and being. After all, the more tangible levels can easily be found in reference books, on the internet, or with the help of artificial intelligence (AI) chatbots.

In other words, in my view, any reader can, for example, build a value stream map, but there are few who will effectively use the tool to generate real and sustainable changes in an organization and its members. So, I prefer to focus on this philosophical question: after all, how many lean experts, with years of experience in several companies and beautiful LinkedIn labels, are really lean when we penetrate the deepest layers of their being?

Lean, for me, is practically a therapy and involves (a lot of) self-criticism, humility, and collectivity. This is where the challenge begins in an increasingly superficial and individual world.

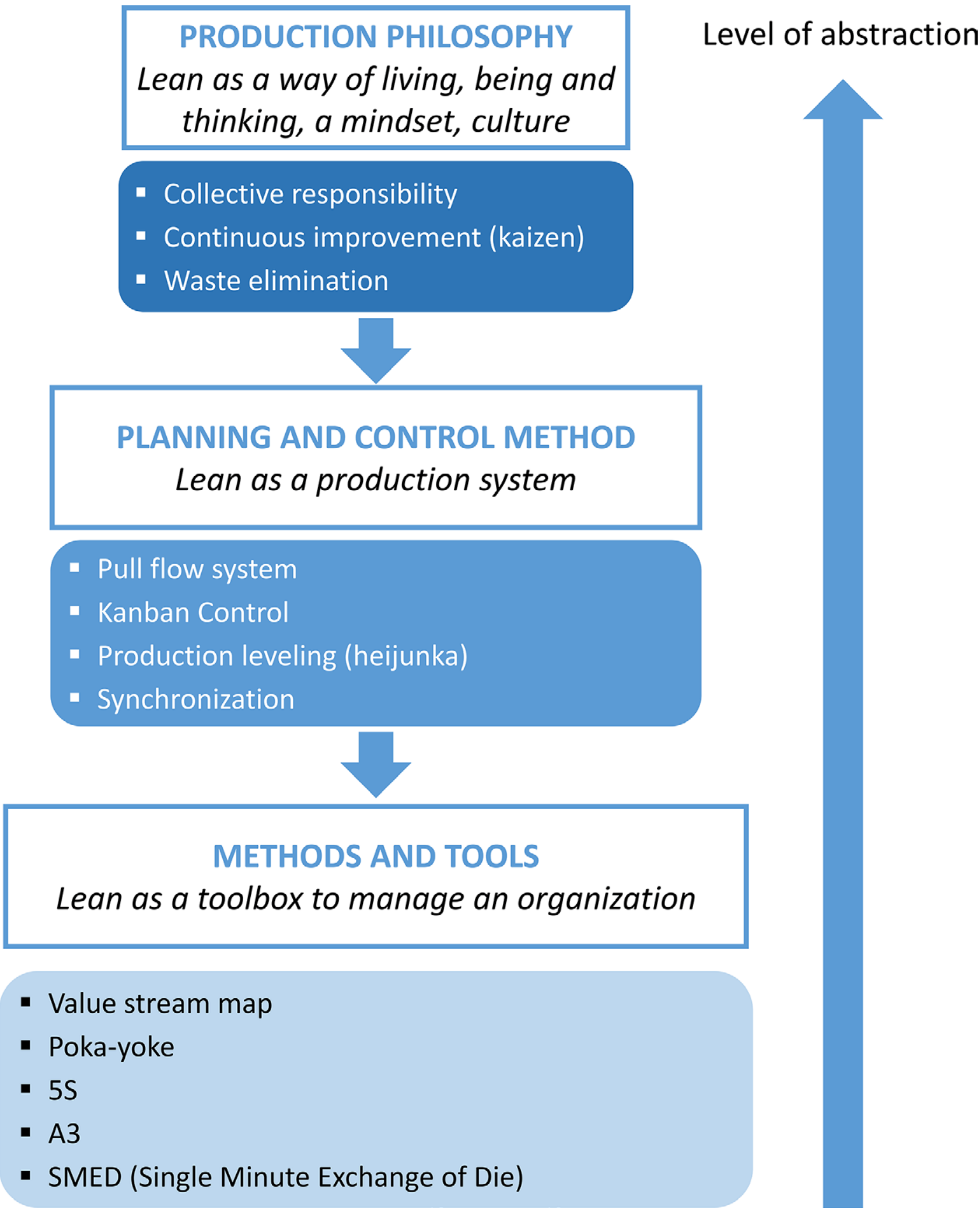

Figure 1 – The different abstraction levels of lean manufacturing

Let us finish with this reflection from the book “Toyota Production System”, by Taiichi Ohno, in which the author alludes to the parable of the tortoise and the hare:

“THE SLOWER BUT CONSISTENT TORTOISE CAUSES LESS WASTE AND IS MUCH MORE DESIRABLE THAN THE SPEEDY HARE WHO RACES AHEAD AND THEN STOPS OCCASIONALLY TO DOZE. THE TOYOTA PRODUCTION SYSTEM CAN BE REALIZED ONLY WHEN ALL THE WORKERS BECOME TORTOISES.”

We live in an increasingly fast-paced world, but is it increasingly efficient and productive? Taiichi Ohno, the genius behind the Toyota Production System (TPS) which inspired Lean philosophy, stated that the TPS would only work when all its employees became turtles. This reinforces that we must aim to increase productivity, but not just quantitatively. It is worth emphasizing that we must always consider the human being as a whole, taking advantage of all their potential and capabilities in a healthy way. So, the invitation remains: are we going on this journey of transformation together?

Starting then, the best way to understand lean is from the understanding of the flow in the processes. Niklas Modig and Pär Åhlström do this very well in their book “This is Lean: Resolving the Efficiency Paradox.” Thus, I will use the mindset of these authors as a basis to define what lean is and what is NOT, and then expand the concepts and reflections to the world of lean digital. However, it is important to point out that some terms of this book were changed to be more aligned with other lean and operations management references like the terms “flow efficiency,” “resource efficiency,” and “efficiency paradox.”

Going into more detail, first, I will introduce what lean is, which, in short, means having a lean process and focusing on satisfying the needs of the flow unit. Next, I will present guidelines to make a process flow – after all, they are the basis for making a process increasingly lean. The third chapter will deal with what is NOT lean: the waste of focusing on resources and the productivity paradigm. Finally, a digital transformation case study will be presented so that these concepts become more tangible and we can enter the lean digital world.

Elisa Granha Lira

WHAT IS LEAN?

In this chapter, we will define what lean is. In summary, when we focus on satisfying the needs of the flow unit, we naturally make our process as lean as possible.

It is learning to see waste

To understand a process, it is first important to understand what flows. We will, therefore, name these objects transformed by the process as flow units. These units can be materials, information, or people:

- **Materials:** in industries, materials are processed by machines and people until a product is produced (car, refrigerator, etc.).
- **Information:** administrative departments and services process information such as spreadsheets and other types of documents – whether physical or digital.
- **People:** in services, the flow units can usually be people who are being “transformed” by the process activities. For example: in a beauty salon, a person has their hair transformed by the hairdresser and beauty products. In a hospital, doctors and other health professionals cure a patient of any illness.

It is important to draw attention to this in order to show how lean is universal regardless of context. We often come across the terms lean healthcare, lean office, and lean digital, but the essence is always the same.

Another issue to point out is that being lean is not an attribute: a process is lean or not. In fact, it is a continuous variable in which we can always become leaner and leaner!

Therefore, when we want to define to what degree a process is lean, we are interested in knowing the time during which the flow unit receives value in relation to the lead time it spends in the whole process.

$$\text{Value added ratio} = \frac{\text{Time for activities that add value}}{\text{Lead time}}$$

That is, how lean a process is can be evaluated by calculating the value added ratio, which analyzes the amount of time it takes from identifying a given need to satisfying it. Thus, the higher the value added percentage (%VA), the leaner the process is.

It is also worth noting that an activity adds value when it meets the needs of the flow unit, while another that does not add value is a time-consuming activity that transfers no value to the flow unit.

Lead time refers to the time required to carry out a process from start to finish. Thus, during the lead time of a flow unit, it will be transformed by value added activities and will waste its time in non-value added activities such as transportation and inventory (Figure 2).

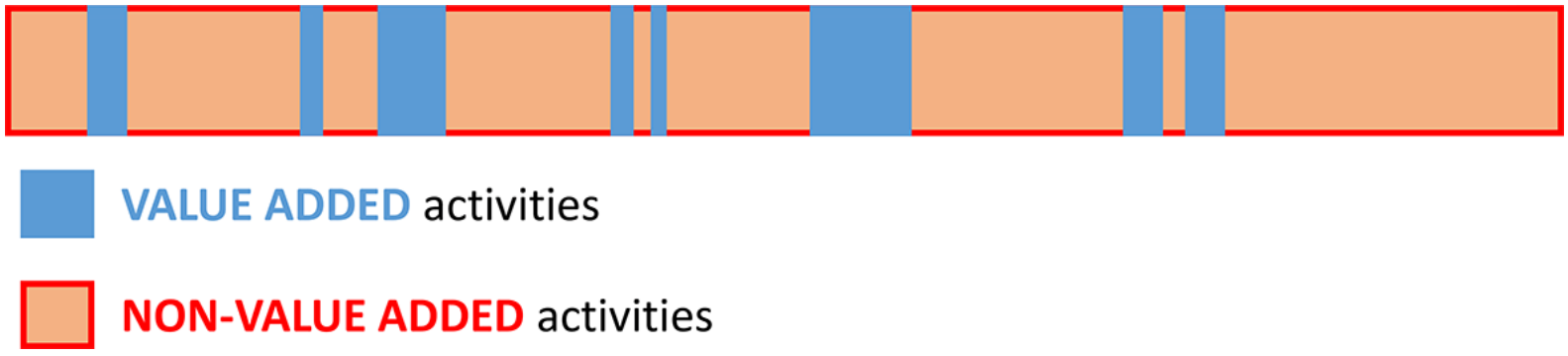

Figure 2 – *Value added and non-value added activities in lead time*

Note that to increase the percentage of value added, trying to increase the time of activities that add value to the process is not a good strategy. After all, this time is relatively small in relation to the lead time. In fact, the value transfer ratio should be maximized by eliminating waste.

Activities that do not add value are presented in lean as the 8 wastes (Figure 3), which consume time, resources, or space but do not necessarily contribute to satisfying the needs of the flow unit.

According to Taiichi Ohno, these eight types of waste are:

1. **Overproduction:** this waste occurs when the company produces more than it needs to serve the customer. It is considered the worst type of waste, the mother of all others, since when it is produced in excess, high inventories are generated, unnecessary movements are made, and resources are used unnecessarily, consuming productive capacity that could be used to meet customer needs. The cause of overproduction is the lack of predictability of internal processes and external relationships with customers and suppliers. For this reason, variability control should always be a goal to be achieved.

2. **Waiting:** the more continuous the production flow, the greater the efficiency and speed with which we can deliver the order to a customer. Thus, delays are related to people, machines, or information "waiting."

3. **Inventory:** they are problematic since they hide problems by making it difficult, for example, to detect defects. In addition, they represent immobilized capital, that is, money that is invested in inputs and whose return will only occur when the customer receives and pays for the product or service purchased. Therefore, one must define the necessary amount of inventory to be used to buffer possible variability in the process without creating excessive costs.

4. **Transportation:** it is a waste related to unnecessary transportation. This waste can be the result of truncated flows, intermediate inventories far from production lines, and inefficient supply strategies.

5. **Motion:** unnecessary handling, like searching for materials or information, is a waste of time. Thus, workstations, layouts, and inventories must be planned in order to optimize the movements that must be carried out by its employees. Unnecessary motion is not restricted to physical resources; information located on computers and other databases must be organized to optimize the company's knowledge management and allow its employees to access them quickly.

6. **Unnecessary processing:** deals with the waste of performing operations that, if eliminated, would not make the slightest difference to the customer. For example, excessive quality inspections can eventually generate more costs than benefits. In this sense, the unnecessary processing of information is also a waste, and many companies and people do not know how to deal with it. Thus, engineers receive high salaries to spend

most of their time reading and answering emails, preparing and watching presentations, updating indicators, filling out reports, and participating in meetings. Obviously, these activities are necessary, but they should not consume more than 30% of a person's work time. The truth is that engineers, for example, today spend up to 90% of their time on these activities that often do not add value. On the contrary, they should spend most of their time seeing and investigating problems in loco at the Gemba. This Japanese term means "real place" or "place where things happen," normally used for the production area of an industry or any place where work adds value to a product, a person, or information.

7. **Defects:** the ideal is always to perform "first time right". This waste concerns, therefore, committing financial, human, and time resources to redo, correct, or rework what was done wrongly.

8. **Nonutilized talent:** this eighth waste was not originally described by Ohno, but it should not be neglected for that reason. It is about the intellectual waste of not using the potential of a company's employees. The idea is not to think of employees as machines that must have as many projects and activities as possible allocated to them. This can be "genius" in the short term, but in the medium and long term, it can be suicide. If the employee is good and proactive, they will probably go to another company, or if they don't leave, it is possible that they will have physical and mental health problems that could compromise their employability. Thus, companies must hire and manage their human resources intelligently, encouraging a work environment that allows employees to fully develop and practice their capabilities and skills as human beings.

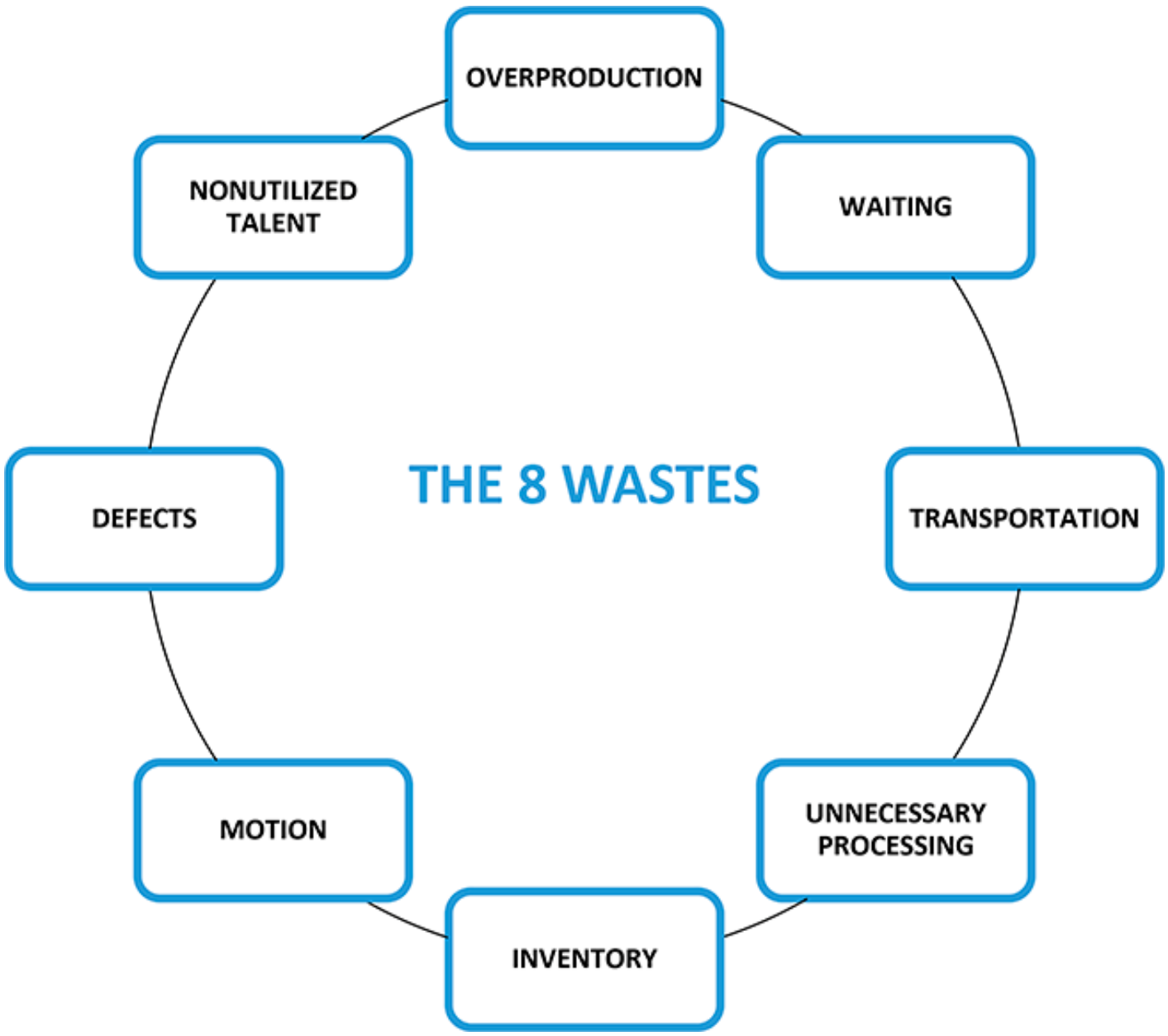

Figure 3 – OThe 8 wastes

It is to focus on the flow

The value added ratio shows the focus of a process. A process with a low percentage of value added is a process whose focus is on utilizing resources (Process A – Figure 4), i.e., the focus is on utilizing most of the organization's resources, such as equipment and people. Meanwhile, a process with a high percentage of value added is a process focused on satisfying the needs of the flow unit (Process B – Figure 4).

Even today, most organizations adopt such a perspective. There is a compelling reason why this approach dominates today's business world: it appears, at first glance, to make sense from an economic perspective, focusing on the internal productivity of resources, based on the opportunity cost. The issue is that if we focus too much on resources, we only achieve a local optimum, which is unlikely to be translated into a global optimum for an organization.

IF WE FOCUS TOO MUCH ON RESOURCES, WE WILL ONLY ACHIEVE A LOCAL OPTIMUM, WHICH WILL HARDLY BE TRANSLATED INTO A GLOBAL OPTIMUM FOR AN ORGANIZATION.

A good practical example is the cost reduction projects that companies often promote. How many improvement actions of these projects have their depart-mental financial gains directly evidenced in gains for the company as a whole?

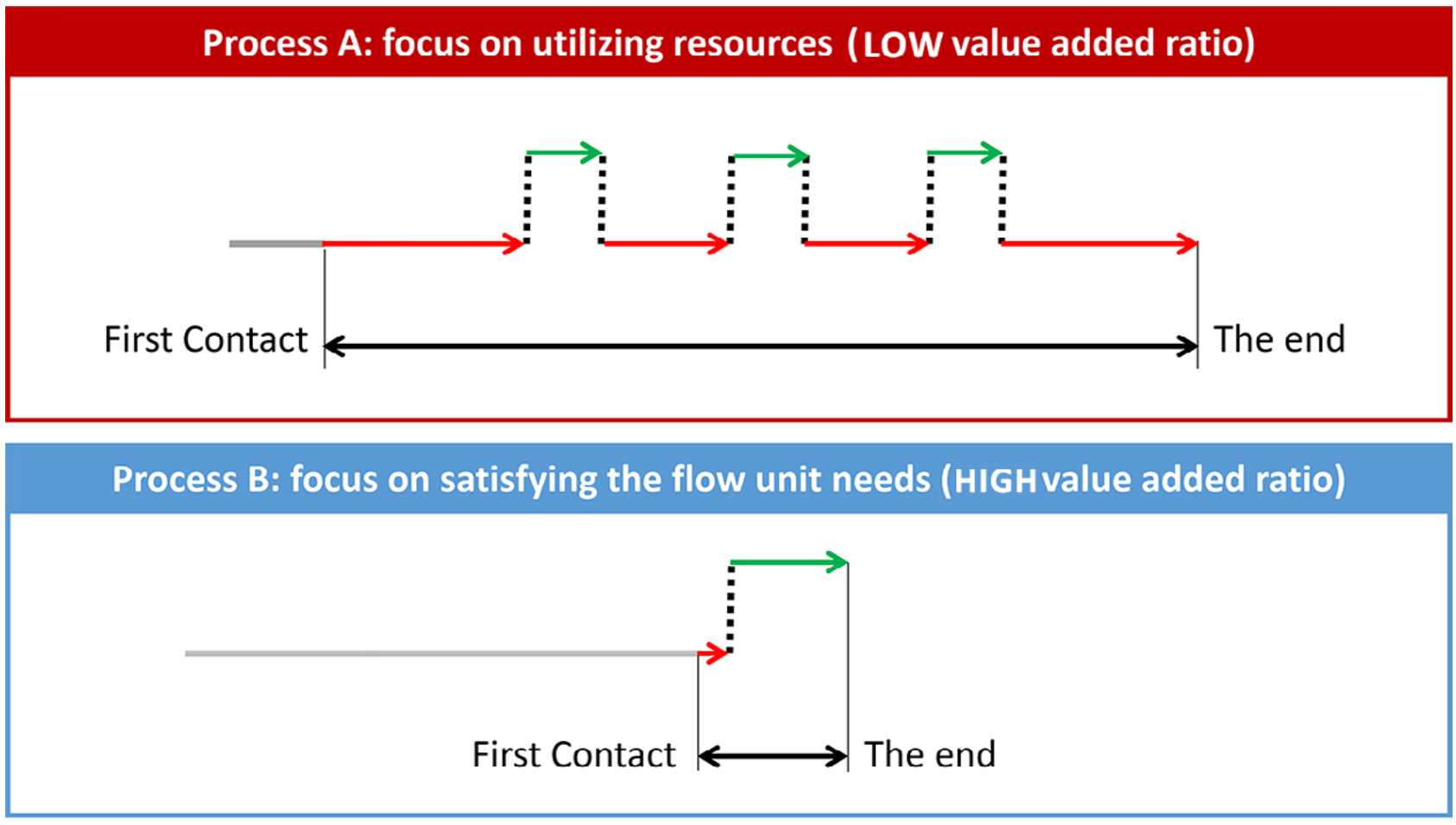

Figure 4 – Processes with different organizational focus

In conclusion, while in a resource-focused process, the flow unit adapts to the organization, in a process with a focus on the flow unit, the organization adapts to the unit being transformed to satisfy its needs as quickly as possible (Figure 5).

For example: in a decentralized breast cancer diagnosis process (with several independent organizations and functions being responsible for different stages of the process: local doctor's surgery, mammography/ultrasound, breast clinic, and cytology), if we go to an ultrasound clinic, we will see an equipment (resource) always busy with the intention of processing (performing tests) as many patients as possible. That is, it is a resource-focused process. However, there are specialized centers, mainly outside Brazil, in which a single organization centralizes all these activities. In this way, a patient can carry out all the activities in a single day, optimizing the time needed for the diagnosis and the results of his/her treatment. That is, in this second case, the focus is on satisfying the flow unit's needs (diagnosis of breast cancer in a patient), and not in occupying a resource as much as possible.

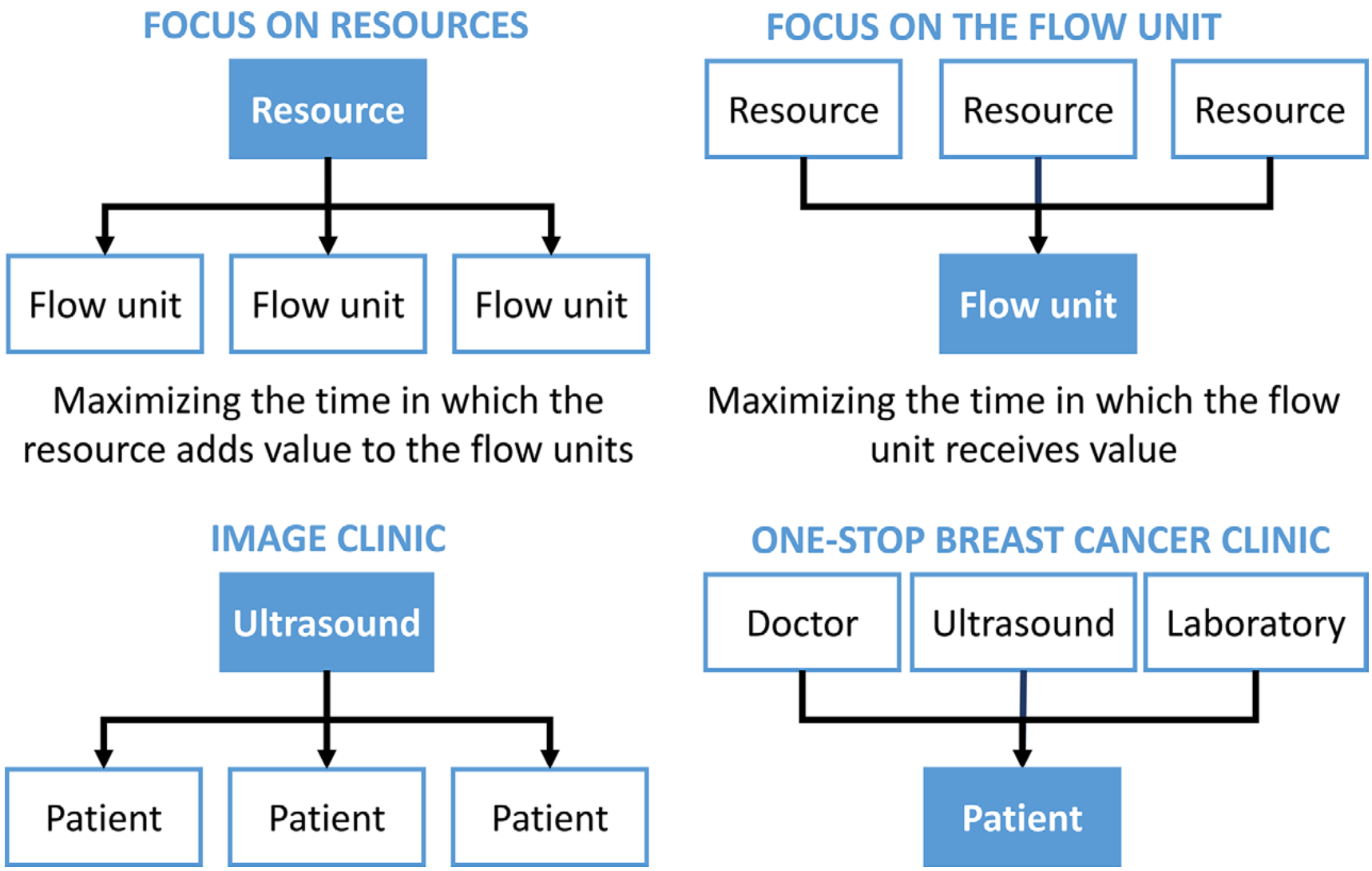

Figura 5 – Processes with different organizational focus (example of a breast cancer diagnosis)

The following chapter will present the principles behind the processes since they are crucial to understanding why focusing on resources compromises the value added percentage of the whole process.

HOW TO ACHIEVE A LEAN FLOW

The processes of any organization are formed by activities that create the route that will transform the flow unit, aiming to satisfy its needs. Thus, to understand how to make a process increasingly lean, it is necessary to understand the principles that govern the processes. These "laws" are universal, as they are applicable to any type of flow unit (materials, information, or people).

As seen previously, we can determine how lean a process is through the value added ratio, which depends on the lead time. Thus, we are interested in understanding what affects the time required to go through a process.

Little's Law

The first principle is Little's Law, which is important insofar as, when managing the flows of a process, three important questions must be answered:

- How many flow units pass through the process given one unit of time (throughput)?
- How long does a flow unit spend, on average, within the process boundaries (lead time)?
- How many flow units are present at the process boundaries (Work in Progress - WIP)?

This law relates these three concepts:

$$Lead\ time = \frac{Work\ in\ Progress\ (WIP)}{Throughput}$$

As previously presented, the lead time refers to the time required to carry out a process from start to finish, from the limits defined for the system.

In-process flow units are all units that started the process but have not finished it yet. It is the work in progress of flow units within the process boundaries.

The throughput refers to the rate at which flow units pass through the process boundaries, which is given by flow units per unit of time.

Figure 6 illustrates Little's Law logic.

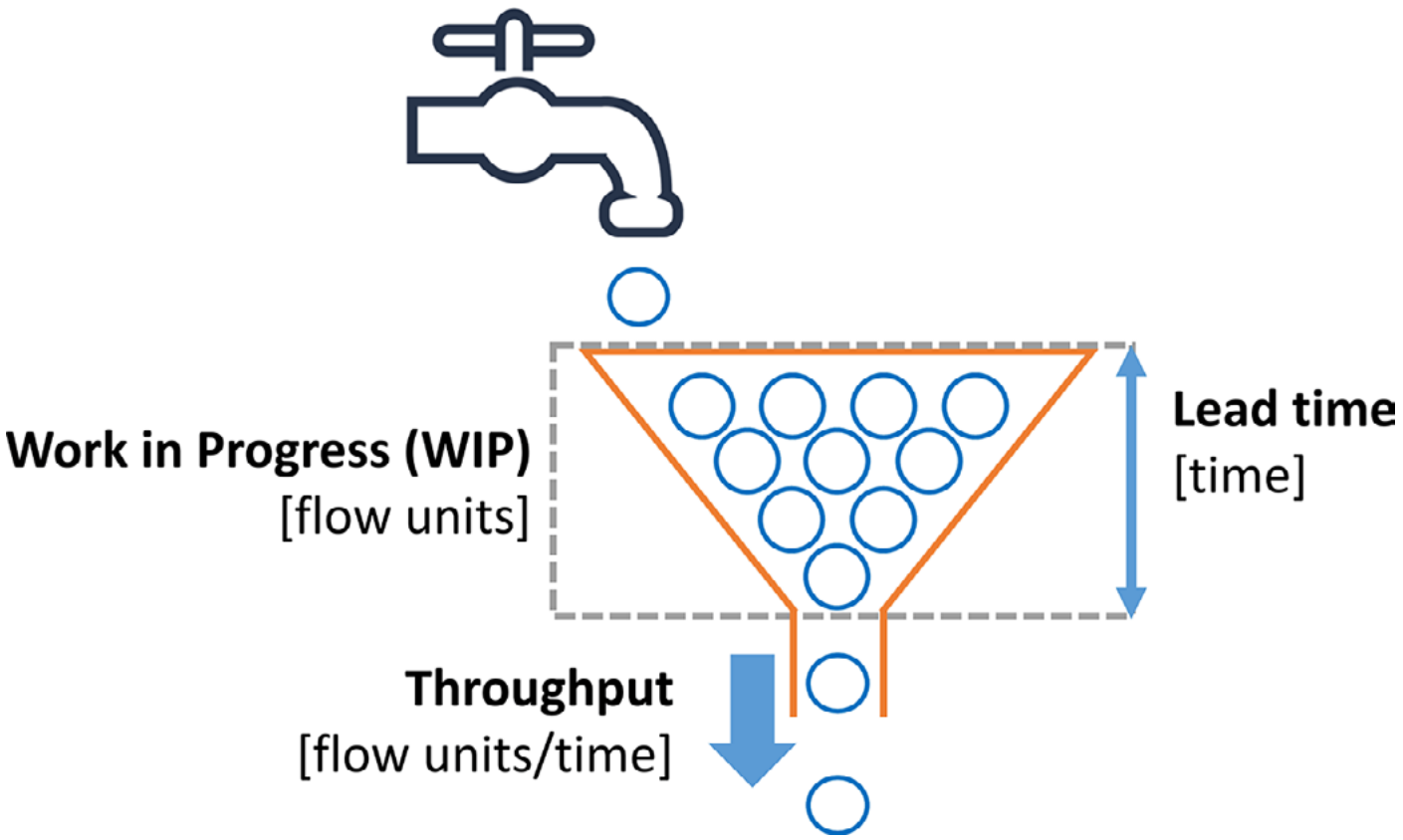

Figure 6 – Little's Law

Let's illustrate Little's Law using the example of a bank. If ten people are in a queue (WIP) at this bank and the cashier attends to one customer every two minutes (2 minutes per customer = 0.5 customers per minute), the lead time is 20 minutes:

$$Lead\ time = \frac{10\ customers}{0.5\ customers\ per\ minute} = 20\ minutes$$

That is, the tenth customer in that queue will wait, on average, 20 minutes to be attended by a cashier (Figure 7).

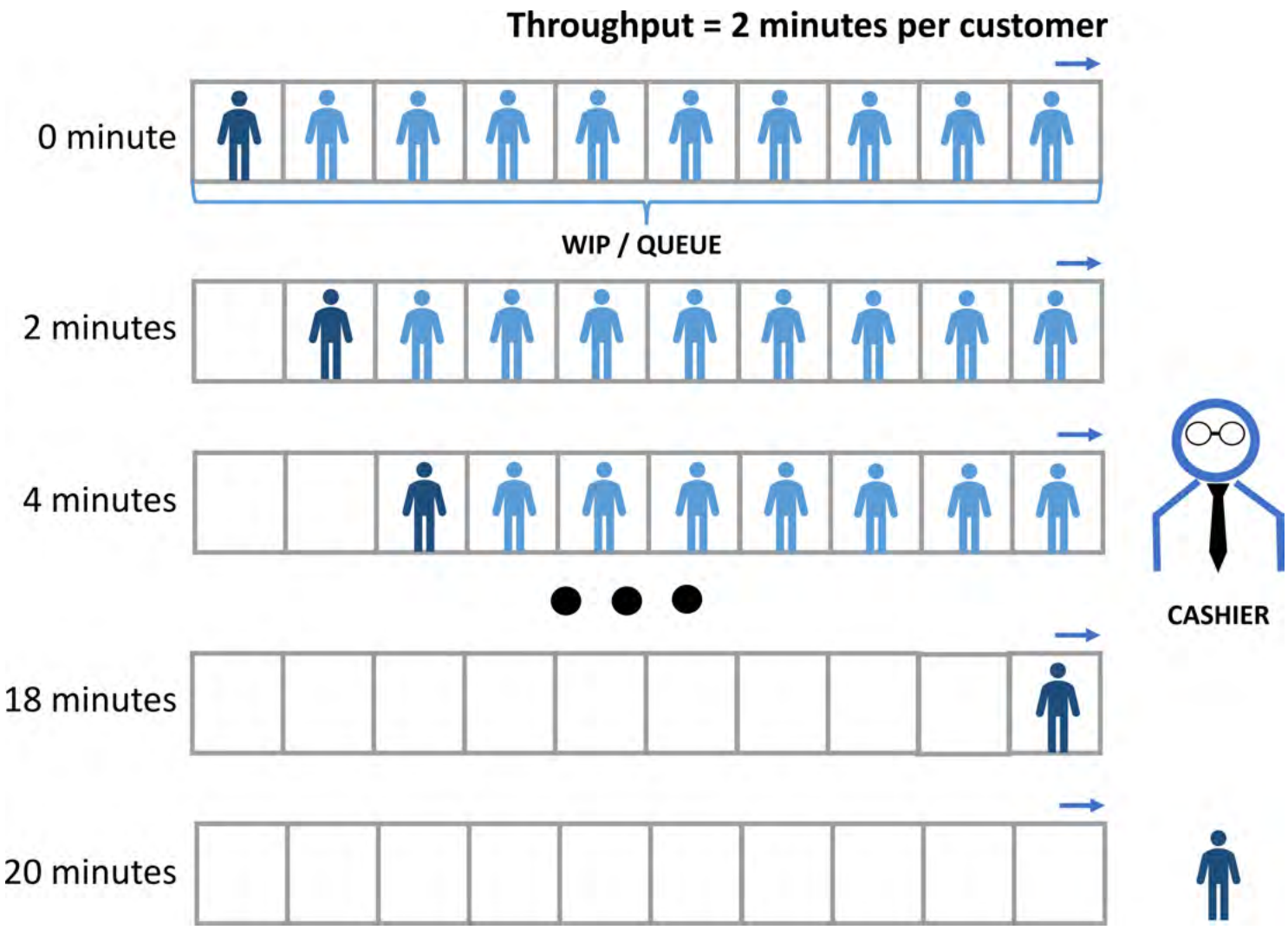

Figure 7 – Little's Law (Bank example)

Little's Law demonstrates, then, that the lead time is affected by two variables: the number of flow units in the process (WIP) and the throughput. Consequently, the higher the WIP and the lower the throughput, the greater the lead time.

This explains why the focus on resources compromises the process flow. In order to make maximum use of the resources, a "queue" of flow units should be created before these resources, reducing their idle time. The logic is that it is the flow unit and not the resources that have to wait, which, in turn, increases the lead time and reduces the value added ratio. Another important point is that the resources are so worried about being busy that they do not even have time to think (waste of nonutilized talent) and critically reflect on any rework that this logic generates in their work routine.

For example, in a bank, we are often faced with a queue of customers (flow units) waiting to be attended by a manager (resources). Thus, the logic of a resource-focused process is that these customers should wait while the manager should always be busy. This overcrowding often generates quality issues and rework for managers since the focus is on serving as many customers as possible and not on assertively solving people's needs.

The Theory of Constraints

Bottlenecks are process stages or activities that restrict the flow of units. Restricting this flow impacts the lead time since the stage with the lowest flow rate limits the other resources.

Bottlenecks are a natural phenomenon of processes, given their sequential execution and the variation in the flow rate in the process stages or sub-processes. In the book "The Goal: a Process of Ongoing Improvement," Eliyahu Goldratt and Jeff Cox illustrate how to identify a bottleneck from the following symptoms (Figure 8):

- Before a bottleneck, there is always a queue/WIP: when the flow unit is physical, this is readily visualized. But when it is digital, the bottleneck can be camouflaged and go unnoticed.
- After the bottleneck, process stages or activities have idle capacity.

Therefore, as bottlenecks generate queues of flow units to be processed, the Theory of Constraints demonstrates, therefore, that these bottlenecks increase the lead time and reduce the value added ratio.

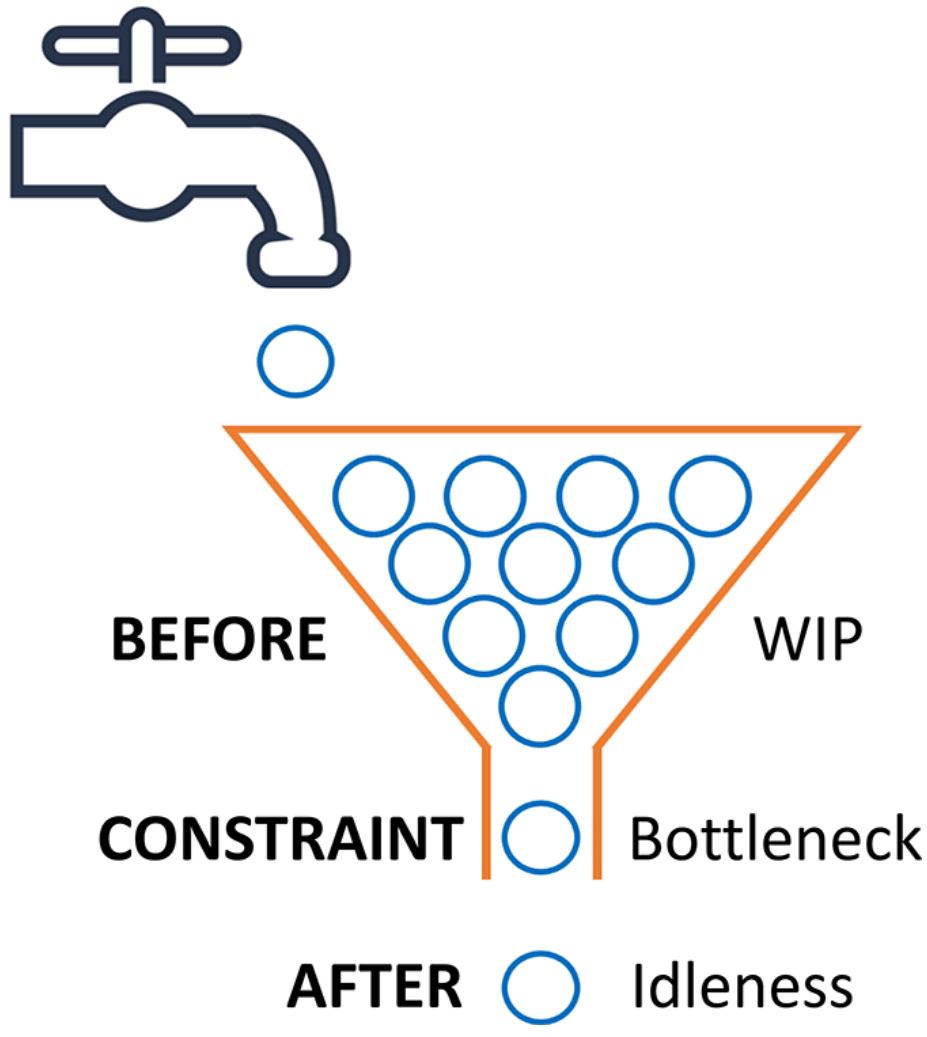

Figure 8 – Theory of Constraints

The variability

The Japanese word mura can be translated as “unevenness,” “fluctuation,” or “variation” and is a major source of waste since it generates both muda (a Japanese word that can be translated as “waste”) and muri (a Japanese word that can be translated as “overburdened”). These three sources of waste (muda, mura, and muri) are known in lean as the “3Ms” (Figure 9) due to their initial letter.

Ideally, processes should be respected in their ideal capacity, without muda, muri, or mura. That is, without overloads, underutilization, and with the least possible variability. Consequently, costs would be optimized, and the risks of quality and safety issues would be reduced.

Variability is also inherent in processes and affects the lead time. A good example is vehicle traffic in large Brazilian cities. If all cars on a large avenue stayed in the proper lane and at the same speed, traffic problems would be avoided. But the reality is that the avenues are often chaotic: the speed of each vehicle varies, and vehicles are constantly changing lanes, not to mention that unexpected situations with pedestrians, bicycles, or animals can occur. These variabilities

end up limiting the flow of vehicles and generating queues in the traffic lanes.

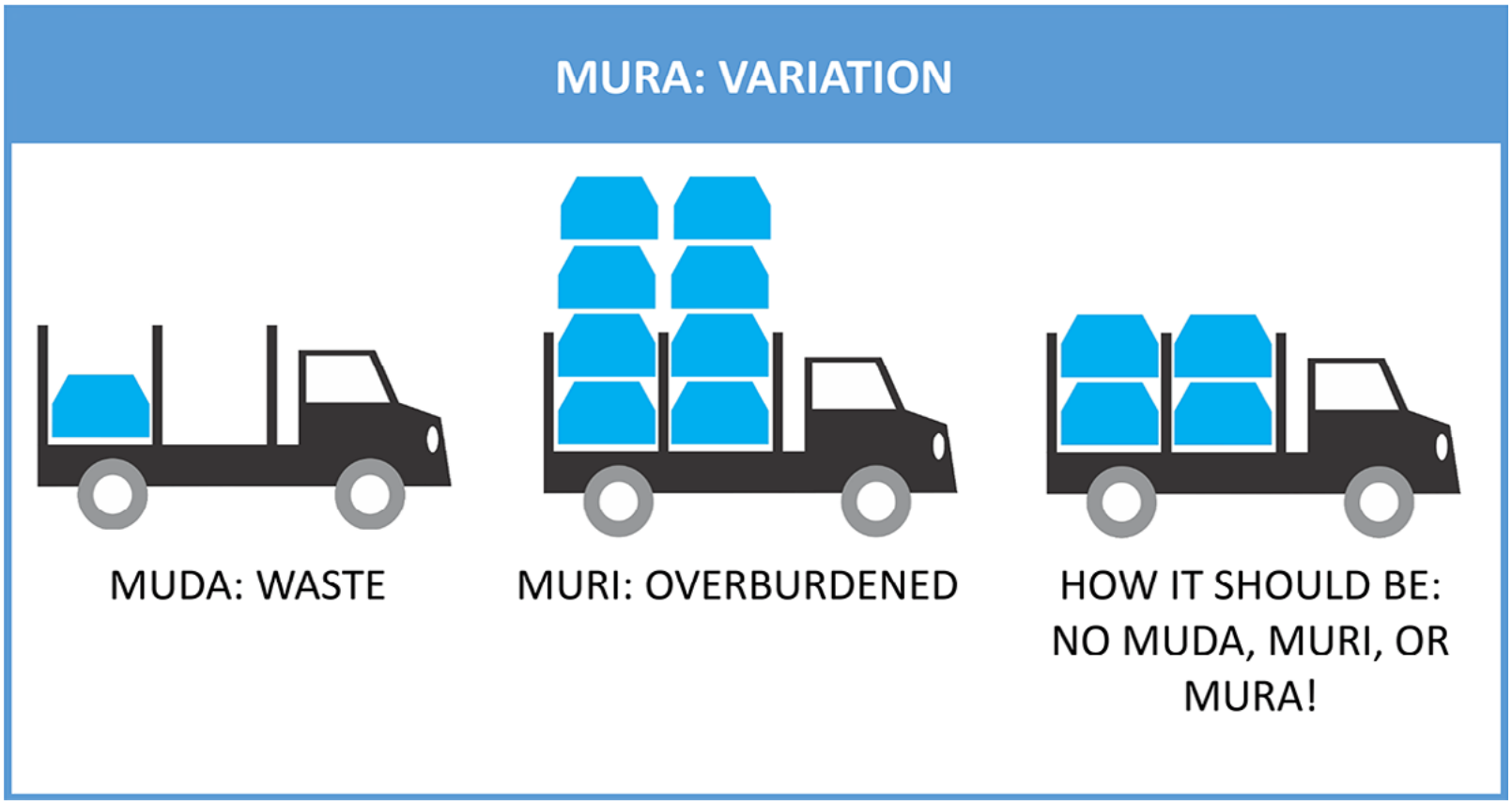

Figure 9 – The 3Ms

Thus, the greater the variation in the process, the greater the lead time. When the lead time increases, the process becomes less lean.

Creating a lean process based on the three principles

Therefore, based on the three guidelines presented, the lead time is affected by the following process variables: number of flow units (WIP), throughput, bottlenecks, and degree of variation (variability).

We can, therefore, develop a leaner process from the following actions:

- Reducing the number of flow units in the process boundaries (queues/ WIP).
- Increasing the throughput, i.e., processing more flow units per unit of time.

- Identifying and eliminating bottlenecks.
- Controlling the sources of process variability.

These principles also demonstrate the impossibility of simultaneously focusing on the flow unit's needs and the resources. Imagine that a manager aims to achieve 100% utilization of their equipment; that is, the manager wants the equipment to produce good parts 100% of the time.

To achieve this, the natural logic is to work with a queue of flow units to be processed. The greater the process variation, the more WIP is needed to buffer variability and keep resources at maximum capacity. In addition, bottlenecks also create additional rows of flow units.

According to Little's Law, a greater number of flow units within the process limits (WIP) increases lead time and reduces the percentage of value added. Thus, when one focuses on increasing the use of resources, one compromises the process flow efficiency.

Another problem with focusing mainly on the use of resources is that superfluous work, which does not add value to the process, is generated from the creation of secondary needs. This is known as the productivity paradigm:

IT IS BELIEVED THAT IF ONE MAXIMIZES RESOURCE UTILISATION, IT IS MORE PRODUCTIVE. HOWEVER, AS MANY OF ITS ACTIVITIES ARE UNNECESSARY OR AVOIDABLE, IT IS ACTUALLY AN UNPRODUCTIVE RESOURCE. IT IS THE FAMOUS AND SHALLOW POP PRODUCTIVITY: THE MORE I DO, THE BUSIER I AM, THE MORE PRODUCTIVE I AM. AM I?

In economics, productivity refers to how much output can be produced with a given set of inputs. However, in a deeper analysis, it doesn't make sense to think about productivity without linking this concept to other ones such as efficiency and effectiveness. If I produce 10 parts per day, but two of them have quality issues, I am less productive than a person who produces 9 good parts per day.

After all, who is more productive: an employee who works day and night to meet their goals or another who delivers the same results within their regular workday? The productivity paradigm can be explained using the logic of the Overall Equipment Effectiveness – OEE performance indicator (Figure 10). Aiming to maximize the utilization of our resources, we try to occupy them as much as possible. That is, we focus only on decreasing their availability loss. However, this blindness leads to performance and quality losses (defects, reworks, secondary needs) which compromise our effectiveness. We don't see these issues because they are the hidden and intangible part of the iceberg: it is easy to see whether a person is working, but it is not easy to check whether this work adds value or is a waste.

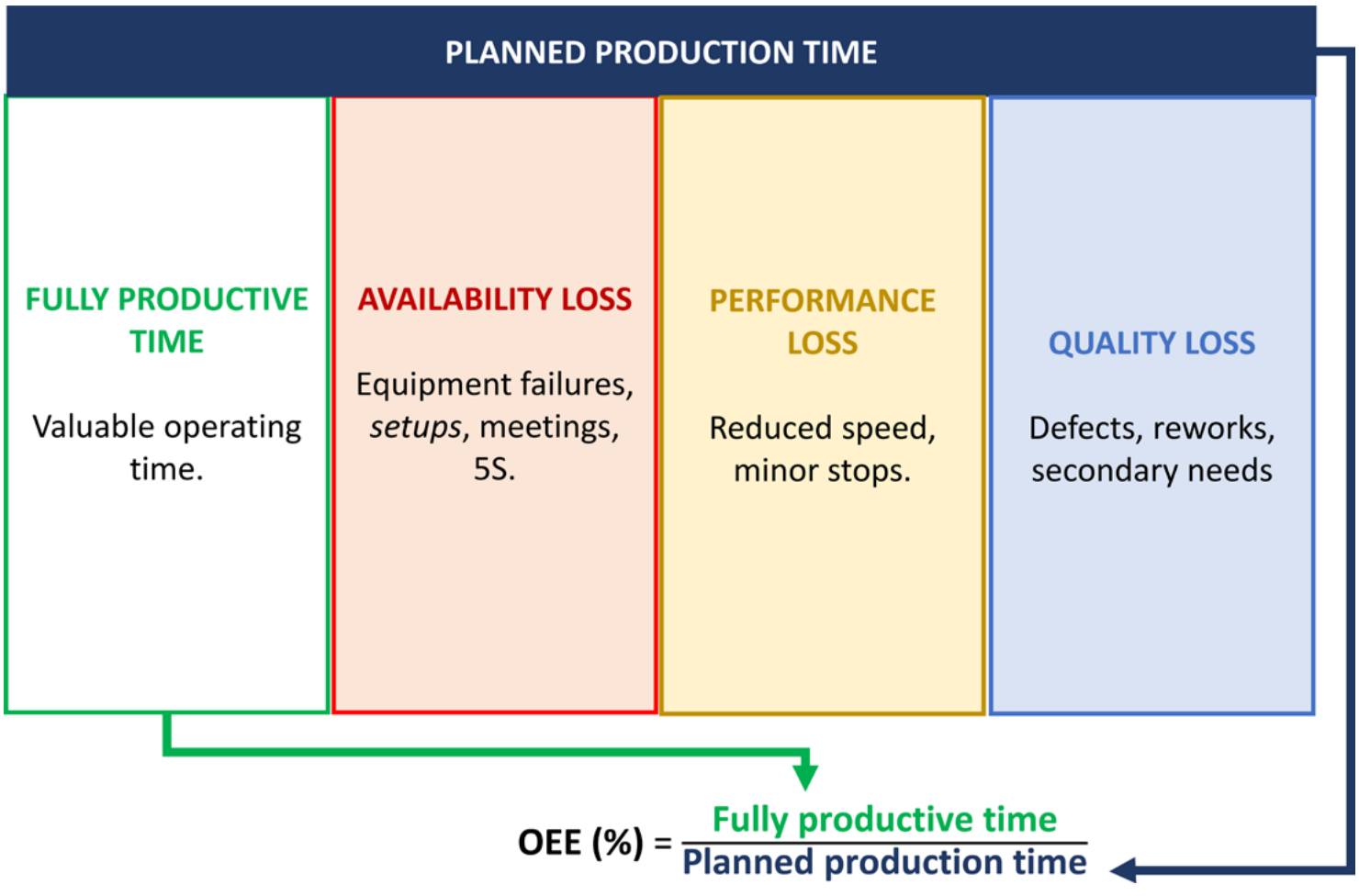

Figure 10 – OEE

Believe it or not, being lean starts with ourselves. Not to mention the potential harm of this overallocation of resources in terms of equipment and people in the medium/long term. Quoting once more Taiichi Ohno: "Speed is meaningless without continuity. Just remember the tortoise and the hare. Moreover, we cannot fail to notice that machines not designed for endurance at high speeds will have shortened lifespans if we speed them up".

WHAT ISN'T LEAN?

As we have seen, when one focuses only on resources without looking at them in relation to the whole process, the efficiency of the entire flow that runs through these resources is affected. This is because primary actions generate secondary needs, which, despite apparently being necessary and adding value, are often unnecessary and represent financial and time waste for organizations. Thus, although many organizations recommend increasing the use of their resources as a primary objective, much of this work would be unnecessary if the organization focused on meeting the flow unit's needs right away. This chapter will present the problems of focusing on resources and the productivity paradigm.

The problem of not learning to see

The exaggerated focus on resources can be easily diagnosed from three symptoms that are sources of inefficiency, according to what was seen about the principles to achieve a lean flow:

- Long lead time.
- Many flow units (high WIP).
- Many restarts per flow unit.

These three symptoms, in turn, create waste in the system in a domino effect that compromises the flow efficiency (Figure 11) since they consume resources without focusing on satisfying the flow unit's needs.

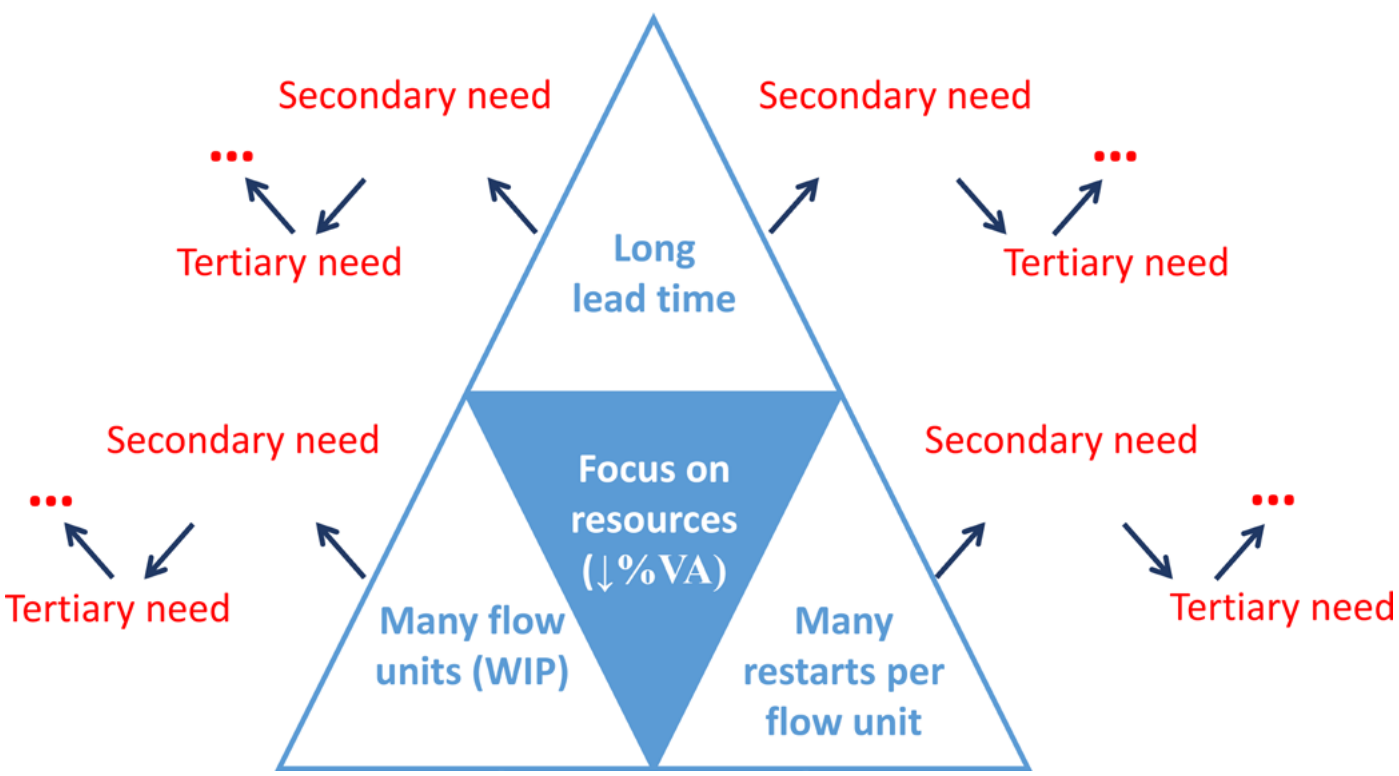

Figure 11 – Domino effect of waste

Table 1 presents the inefficiencies created in a process that focuses on resources without looking at the whole, based on each of these three symptoms.

SOURCES OF INEFFICIENCY
LONG LEAD TIME…
Generates secondary needs
Closes important windows of opportunity
MANY FLOW UNITS (HIGH WIP)…
Require additional resources
Generates secondary needs
Hide problems
Trigger stress
Stimulate loss of control
MANY RESTARTS PER FLOW UNIT…
Generate secondary needs
Require mental set-up time
Generate defects
Generate frustration

Table 1 – Sources of inefficiency

These inefficiencies generated by a resource-focused process will be illustrated using different examples.

The long lead time generates secondary needs and closes important windows of opportunity. Let's take as an example the typical activity of organizing receipts during a trip for future reimbursement requests. This is a common reality for many of us: we keep adding receipts in a disorganized way in our wallet, and usually, when we return from the trip and have free time, we take the opportunity to fill in a spreadsheet with this information, scan the receipts and enter the refund request in the system.

The point is that many secondary needs arise in this process, from organizing receipts to finding them all, for example. See: if, as soon as we had lunch, we would take a photo of the receipt and enter the amount into a refund system via an app, for example, it would not be necessary to fill out a spreadsheet, organize the receipts, pick them up, and there would be no risk of losing them. For this reason, not inputting this data right away, in addition to creating secondary needs, makes us miss the window of opportunity to do "first time right."

The symptom "many flow units in the process" is harmful because it generates secondary needs, hides problems, and stimulates loss of control. Thus, the high WIP ends up demanding additional resources to overcome these inefficiencies.

An example of this is the excess of information that we need to deal with nowadays via email, app messages, and social networks. Have advances in information technology made us more efficient? High volumes of information make it difficult for us to have an overview of what is a priority and what is not. As a result, we waste a lot of time looking for and organizing information (secondary needs). Not to mention that quality problems are generated, which tend to be hidden by the high WIP.

In addition, this second source of inefficiency (many flow units in the process) creates productivism stimuli, overloading existing resources. Doing too many things at the same time increases the risk of losing control, which leaves people frustrated and stressed, causes communication problems, and creates secondary needs (meetings, search, organization, and action plans). If even machines are limited resources, why shouldn't humans be?

Returning to the example of the excess of information that we deal with nowadays, its negative effect on people's physical and mental health is a fact. We are always busy and continuously receiving new demands. As a result, we end up carrying out a series of reworks because instead of thinking and planning, we are stuck in a "do" loop. Sometimes, it is important to ask ourselves: is it worth it?

In the end, many routines are created by organizations to deal with the problems that arise from the high number of flow units in process, which would not be necessary if the organization focused on meeting the needs of the flow unit.

The third source of inefficiency – many restarts per unit of flow – is the famous "start and stop". When we restart a task, mental preparation time is required. In addition, when we shift focus from one task to another repeatedly, we mentally tire ourselves out more, which can lead to a continuous increase in the time needed for mental preparation.

For example, when we deal with a large number of emails, we often have to read the same email more than once, especially when the message has many details and a higher level of complexity. Thus, we end up preferring to postpone this task to a more opportune moment. Restarts, therefore, generate secondary needs for rereading, searching, and organizing these emails, in addition to the risk of forgetting and reworking. It also increases the risk of transferring responsibility in case of vacations, maternity leave, or employee turnover.

If we need time to remember and mentally prepare ourselves for an activity that we restart, imagine when it is a new person who will be responsible for it... This responsibility reallocation generates distortions in information and communication problems, which, in turn, can lead to defects and quality problems. That is, this source of inefficiency ends up being converted into rework and has negative impacts on customer satisfaction.

As can be seen, inefficiency propagates in a domino effect, in which each of these three sources of inefficiency influences the others. As previously discussed, the strategy of focusing on the local optimum, prioritizing resources over flow units, has as symptoms many process flow units and long lead time, which are directly linked by Little's Law. These two symptoms end up generating the third source of secondary needs: the need to continually restart the same task.

The productivity paradigm

As we have seen, the focus on resources creates “productive” islands, where the needs of flow units are subdivided into minor steps that are performed by resources with high utilization. However, each island sees only itself and has no vision of the whole. Consequently, the process is suboptimized with multiple local optima, while the global optimum is compromised with this approach, creating secondary needs and reducing its value added ratio. This is what we call the productivity paradigm, that is, the illusion of believing to be productive when, in fact, it is not, since activities that apparently “add value” are camouflaged waste.

It is worth noting that this productivity paradigm exists both at the individual and organizational levels in all contexts: in private services, industries, and even public offices.

For example, imagine a metallurgy company in whose plant we have several departments such as steel mill, hot rolling, and wire drawing. These departments are also subdivided into subprocesses with specific managers or coordinators. If the organization is not concerned and does not create an effective effort for global improvement, the manager of each department will naturally favor local interests, which are not necessarily the best for the organization as a whole.

Thus, in order to optimize a system, all those involved must understand its overall objective and start thinking differently, as local interests do not always match global interests. In addition, it is crucial to have engaged leadership and to share the performance metrics among departments/functions.

From the individual perspective, this waste with superfluous work also often goes unnoticed: we think we are adding value by being continually busy, but the reality is that much of our work is avoidable waste. The work of an engineer, for example, often involves considerable time reading emails, collecting information for performance indicators, filling out reports, and participating in meetings.

Obviously, some of these activities are necessary, but does all this work add value to the organization? That is why we must always ask ourselves:

HOW MUCH OF OUR STAFF/ EQUIPMENT WORK TIME IS SPENT ON ACTIVITIES THAT DO NOT NECESSARILY ADD VALUE AND WERE CREATED OUT OF SECONDARY NEEDS?

One can be sure that there is always hidden waste that can be eliminated, and self-questioning is the first step in dealing with the productivity paradigm. When questioning ourselves about this productive paradox, we create the challenge of looking at the flow throughout the process and aiming for global gains. In a system whose focus is on satisfying the flow unit's needs, the lead time is short, and few flow units are within the process boundaries (low WIP). This ends up avoiding unnecessary restarts. Secondary needs and superfluous work are therefore avoided. Literally, everything flows.

Lean is a strategy that aims to optimize the efficiency and effectiveness of the process. We need only recall its philosophy as presented in Figure 1: we seek to eliminate waste by adopting a global vision in which everyone is responsible for continuous improvement, to the detriment of individual thoughts and stimuli, based on customers' real needs.

WHAT IS LEAN DIGITAL?

In this chapter, we will illustrate the concepts previously presented in a digital transformation case study. It is worth noting that, as the name implies, this is a transformation in which a physical process gradually incorporates digital technologies. In this sense, lean digital is nothing more than using these digital technologies to make the process increasingly lean.

Case study: what is NOT lean digital?

A certain organization created a project with the objective of digitizing its process. The idea is that in the future, analysts will be able to carry out their work remotely, which is called telecommuting. Thus, services that were previously provided in person are now being carried out at a distance. According to those involved, the change will represent a considerable gain in productivity for the organization, as less physical structure will be required. In addition, with a more "controlled" work environment, with less interference, it is expected that these analysts will be able to analyze a greater number of requirements per day.

In order to create productivity targets for the analysts, their activities were mapped, and chrono-analyses were carried out to assess the time needed to carry them out. Therefore, the process was evaluated, and the time before and after digitization was measured.

Figure 12 represents how the process was performed before its digitalization: the applicant scheduled the service at one of the organization's units by telephone or Internet. Upon arriving at the place on the scheduled date, he had to wait to be called. When it was his turn, he was attended directly by an analyst, who already analyzed the case in the presence of the user. If any pending documentation were detected, the analyst would immediately request the document from the applicant. Most of the time, the applicant had the document in hand and, thus, already handed it over to the analyst to continue the process. In this case, the person who requested the service spent 28 minutes in activities that "added value" (data input, problem-solving, and conclusion) in a total of 68 minutes, which was the total time that the person remained in that place. The remaining 40 minutes the person spent waiting for service.

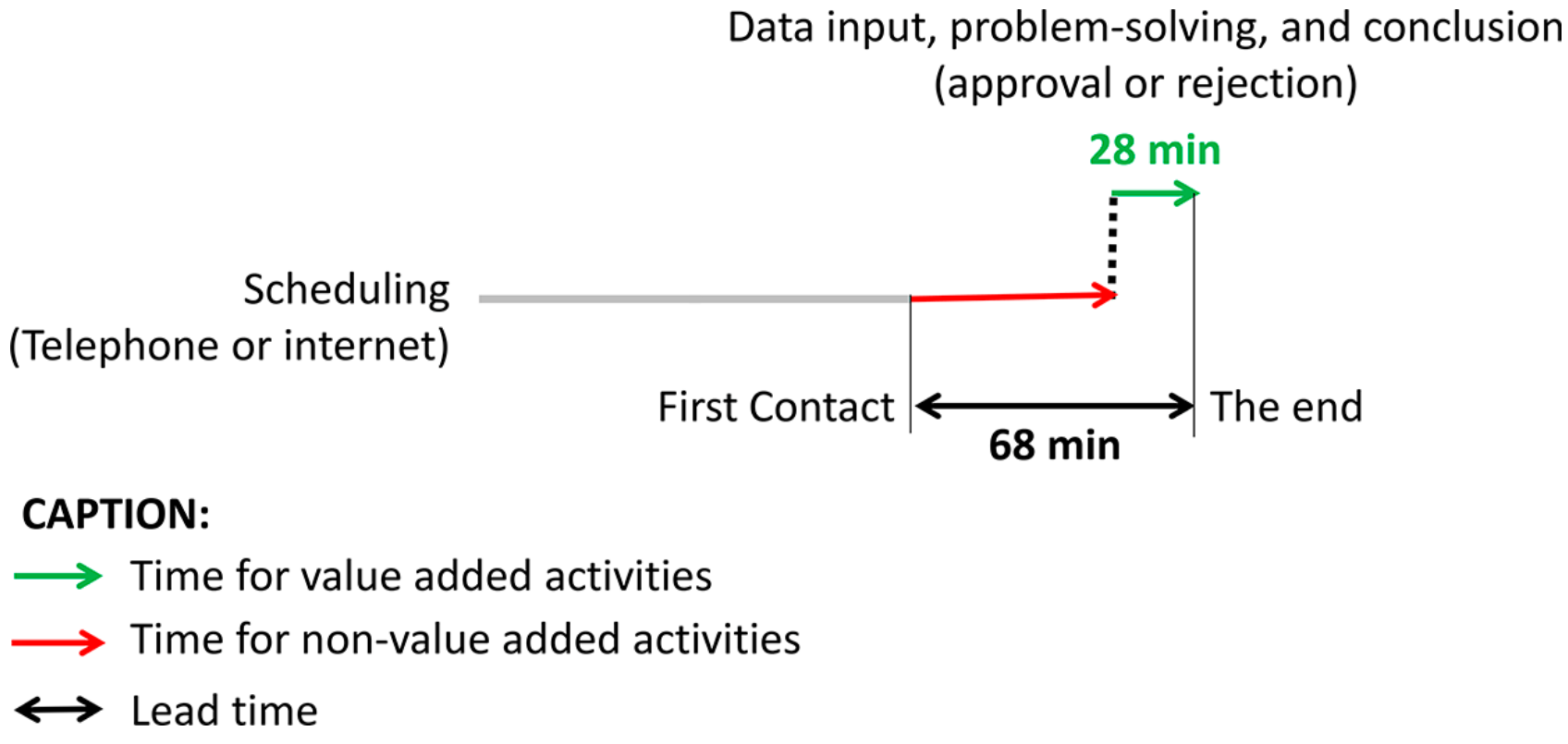

Figure 12 – Process BEFORE digitization

Figure 13 illustrates the process after digitization. In the future, the applicant will also have to schedule an appointment by telephone or Internet, but on the scheduled date, they will be attended by an intern, who will be responsible for assisting them and digitizing all the documents necessary for the analysis of their case. A predetermined time of 20 minutes was established for each appointment. After this initial service, the application will enter a virtual application queue to be analyzed. In Figure 13, the green arrows represent the time for value-added activities performed by the analysts. The red arrows, on the other hand, show the moments in which the application was stopped in the system, waiting for the pending documentation and the next analysis. In this second case, it was necessary for the analyst to request the resolution of pending issues twice. Thus, the analyst's total work time was 90 minutes in relation to the total time from their first analysis to the date of completion of the process (18 days).

It is worth mentioning that when some necessary document is missing, the analyst makes a demand for the applicant, which must be met so that the analysis of the application can proceed. This is a very curious activity. As soon as a case is pending, the analyst writes tips and relevant information about it in a notebook for future reanalysis. Another point that draws attention is that there is neither visual management nor a friendly interface in the system they use. In this way, they cannot easily distinguish which cases are waiting for a first analysis or have already returned from an issue resolution. Therefore, the analyst always sets aside part of the day to look for cases that solve their issues based on their notes in the notebook (through the famous CTRL+F) and eventually

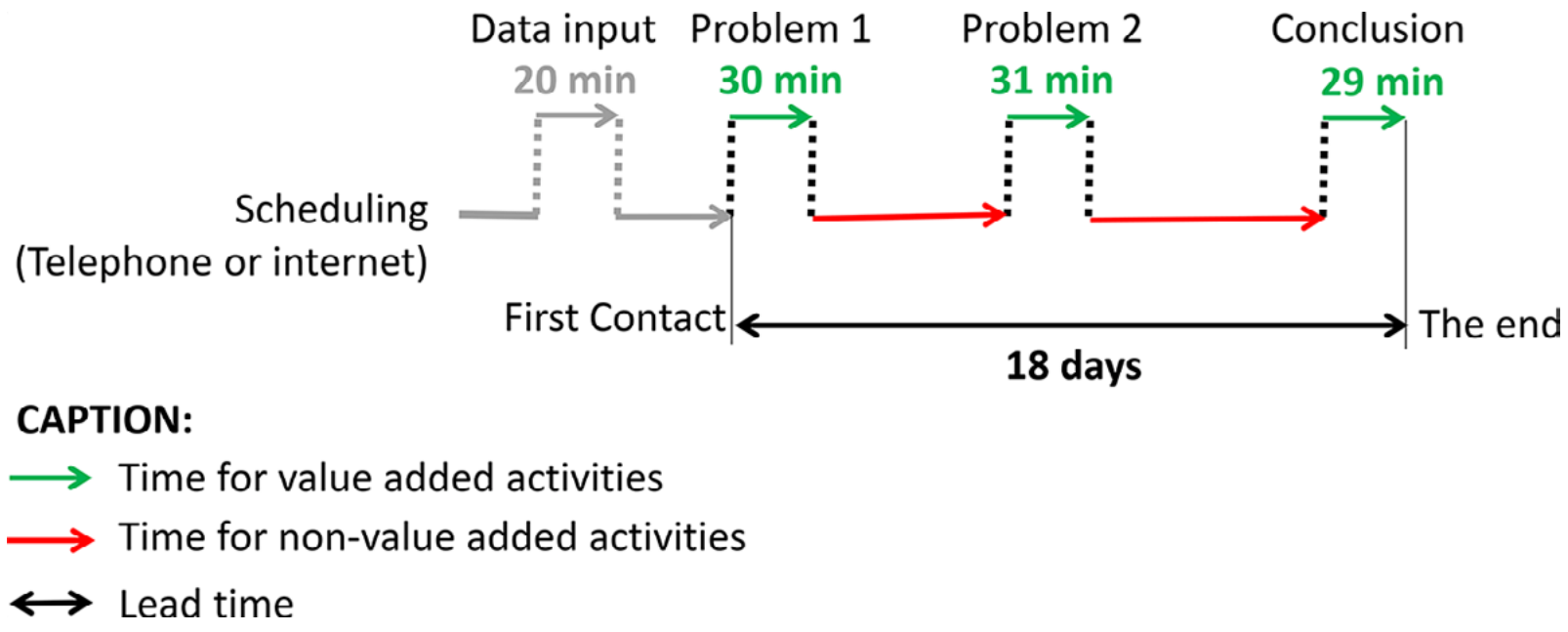

Figure 13 – Process AFTER digitization

Case analysis

First, it is interesting to compare the two cases in relation to the value added ratio (% VA) from the first contact of the analyst resource until the completion of its analysis:

- Calculation of the value added ratio BEFORE process digitization:

$$\%\,VA_{\,BEFORE} = \frac{28\ minutes}{68\ minutes} \cong 0.41 \rightarrow \mathbf{41}\%$$

- Calculation of value added ratio AFTER digitization:

$$\%\,VA_{\,AFTER} = \frac{30 + 31 + 29\ minutes}{18\ days \times 24\ \frac{hours}{day} \times 60\ \frac{minutes}{hour}} =$$

$$\frac{90\ minutes}{25920\ minutes} \cong 0.0035 \rightarrow \mathbf{0.35}\%$$

This is a curious example because we notice that there was a reduction in the value added ratio of the process. That is, the process before digitization had a greater focus on satisfying the flow unit's needs compared to the situation after the changes were made. After digitalization, there was a low value added ratio (< 1%), which characterizes a process that focuses on the use of its resources. That is, if we evaluated their resources, in this case, the analysts, we would conclude that they analyze many requests per day but finish few in the same period, which means they are "efficient" reworkers. This exaggerated focus on occupying resources can be evidenced in the form of the three symptoms that a non-lean process presents:

- Long lead time: an increasing time required to analyze an application from its entry into the system until its approval or denial.
- Many flow units: long virtual queue of applications to be analyzed;
- Many restarts per flow unit: lack of documentation creates the secondary need to reanalyze the same case many times. Moreover, every time a request returns with a pending issue resolved, the analyst spends considerable time re-studying it, as, due to the high number of requests analyzed per day, one no longer remembers much about the one that just returned.

Table 2 presents in detail the secondary needs and other problems that each of these symptoms generates in this process.

SOURCES OF INEFFICIENCY
LONG LEAD TIME FOR AN ANALYSIS…
Generates secondary needs: redundant reanalyses and writing e-mails requesting new documents
Closes important windows of opportunity: previously, one could request missing documents directly from the applicant
GROWING VIRTUAL QUEUE OF CASES AWAITING TO BE ANALYZED (HIGH WIP)…
High WIP demands additional resources (more analysts)
Generates secondary needs: writing in the notebook data about interrupted cases and tips to speed up the next analysis
Hides problems: camouflaged quality and even ethical issues are generated during the analysis
Generates stress: analysts overloaded with rework
Stimulates loss of control: some cases return with the missing documents, but the analyst takes weeks to notice this return
ANALYST INTERRUPTIONS TO REQUEST ADDITIONAL DOCUMENTS…
Generate secondary needs: analysts take part of the day to look for cases that came back to be reanalyzed
Demand mental preparation time: analyst needs to remember cases they analyzed months before
Increase the risk of transferring responsibility: another person is going to reanalyze the case if the analyst asks for vacation, maternity leave, or changes company.

Table 2 – Sources of inefficiency (case study)

This case illustrates well the inefficiencies of a non-lean process: the long lead time closes important windows of opportunity for the analyst to request the missing documents directly from the applicant, which would facilitate communication and even allow to solve the problem in the first meeting – after all when this happened before digitization, customers often had them in their hands. That is, the activity of requesting additional documents can be seen as a secondary need.

Thus, the delay in meeting a primary need created a chain reaction of secondary needs. The various requests for missing documentation contributed to an increase in the lead time since the applicant was not always able to promptly deliver the pending documents. Every time a decision gate closes in the process, waiting for pending documents, variability is invited to act. That is, every time "the ball is passed" for the applicant to present the missing documents, many things can happen until they effectively manage to resolve the pending issue. The long lead time generates superfluous work that consumes existing resources and generate the need to acquire acquiring new resources.

EVERY TIME A DECISION GATE CLOSES IN THE PROCESS, WAITING FOR AN ISSUE TO BE RESOLVED, VARIABILITY IS INVITED TO ACT.

Exemplifying the problems generated from the source of inefficiency "many flow units in the process," in this case study, every time additional documentation was requested, the analyst used to write in a notebook information about the respective requirement (secondary need) because the system made it difficult to visually and automatically identify the cases that had pending issues resolved. Conclusion: many times, a pendency was already resolved, but the analyst would take a few days or weeks to realize this and proceed with the analysis. Not to mention that in the end, after this change in the process, the analyst works more; that is, they are subjected to greater stress, with no real gain in productivity as a counterpart. After all, he/she reworks more and completes fewer analyses. This is the productivity paradigm in practice. Those who see the analyst often believe that he has become more productive after

the digitization of the process, as they become busier. However, most of their work is rework, that is, hidden waste.
Returning to the parable of the tortoise and the hare, the analyst is clearly the hare that runs the most but loses the race. After all, it is increasingly busy re-examining cases where documentation is missing, but it finalizes fewer of them. In relation to the constant reanalyses, these also generate secondary needs, such as the analyst's activity to check which cases could eventually have returned at the end of the day. Furthermore, when the analyst begins a second analysis, it is as if he/she were analyzing it for the first time. After all, despite the hints in the notebook, it has been a long time since they first looked at that case. Finally, there is the risk of transferring responsibility. That is, the analyst can go to another company or ask for maternity leave, retirement, etc. In this case, an eventual second analysis will be carried out by another person, and obviously, the lead time will be even longer.

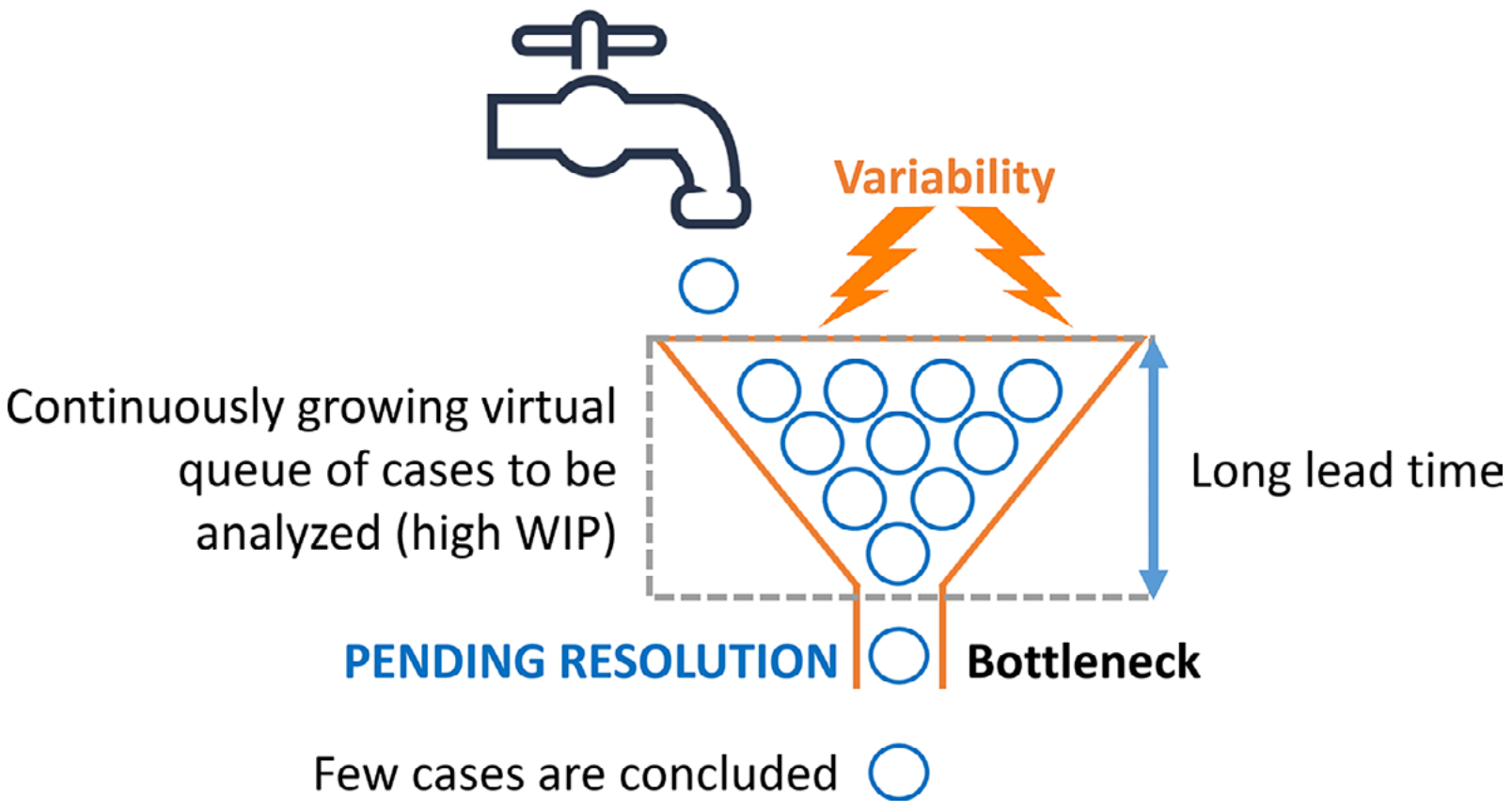

Figure 14 – Principles of the processes (case study)

This example demonstrates that a poorly planned digital transformation can have the potential to make a process less lean. Sure enough, there was plenty of real evidence that something had gotten worse, like the growing queue of cases to be analyzed, the fact that analysts finish fewer analyses, and the fact that the bottleneck in this process is always the resolution of pending issues (Figure 14). This, by the way, is an activity that should never have existed if data entry had been done correctly – and that, because it was not, ended up

generating reworks. Thus, it is worth emphasizing that this activity, as well as other types of waste, should not be included in the work instructions. We must always carry out kaizen (continuous improvement) already on paper.

It sounds obvious when an example like this is presented in a chewy and didactic way for people, but is it so easy to see what is happening when you don't have the vision of the whole but only a picture of a part of the process? After all, the virtual queue is continuously growing, encouraging firefighter-type actions, which solve only the symptoms and create other problems.

For example, the objective of this project was to create productivity targets for analysts: if the problem that arises is that one is analyzing more cases but concluding fewer, why not create a target number of cases to be finished per day? This famous naive "management" reduces the quality of the analysis and can even generate unethical situations by encouraging the rejection of certain cases to simply hit the goal. Thus, it is worth reinforcing:

TOP MANAGEMENT IS RESPONSIBLE FOR THE ENTIRE ORGANIZATION'S COMMITMENT TO SATISFYING THE NEEDS OF THE FLOW UNIT SO THAT CORPORATE GOVERNANCE IS FOCUSED ON THIS SINGLE PURPOSE.

Now, we are going to investigate the root causes of this problem to think of actions that could easily prevent this situation from happening, ensuring that real productivity gains were achieved in this digital transformation project.

As seen earlier, the low percentage of value added is related to the waits and rework that arise from not doing it right the first time. The bottleneck of the process is found in the pending resolution activity since, before that, a growing virtual queue of cases waiting to be analyzed is created, and, in the sequence, it is verified that few cases are finalized. If the guideline to do "first time right" was followed, this activity should not even exist.

Thus, this bottleneck, in addition to considerably increasing the lead time of the process, creates secondary needs that represent a "waste" of time for the analyst. Consequently, poor data entry quality triggers serial reworks, which in turn generate a high backlog of requests awaiting to be analyzed (long virtual queue), long lead times, and overloaded resources that are highly efficient in reworking (Figure 15).

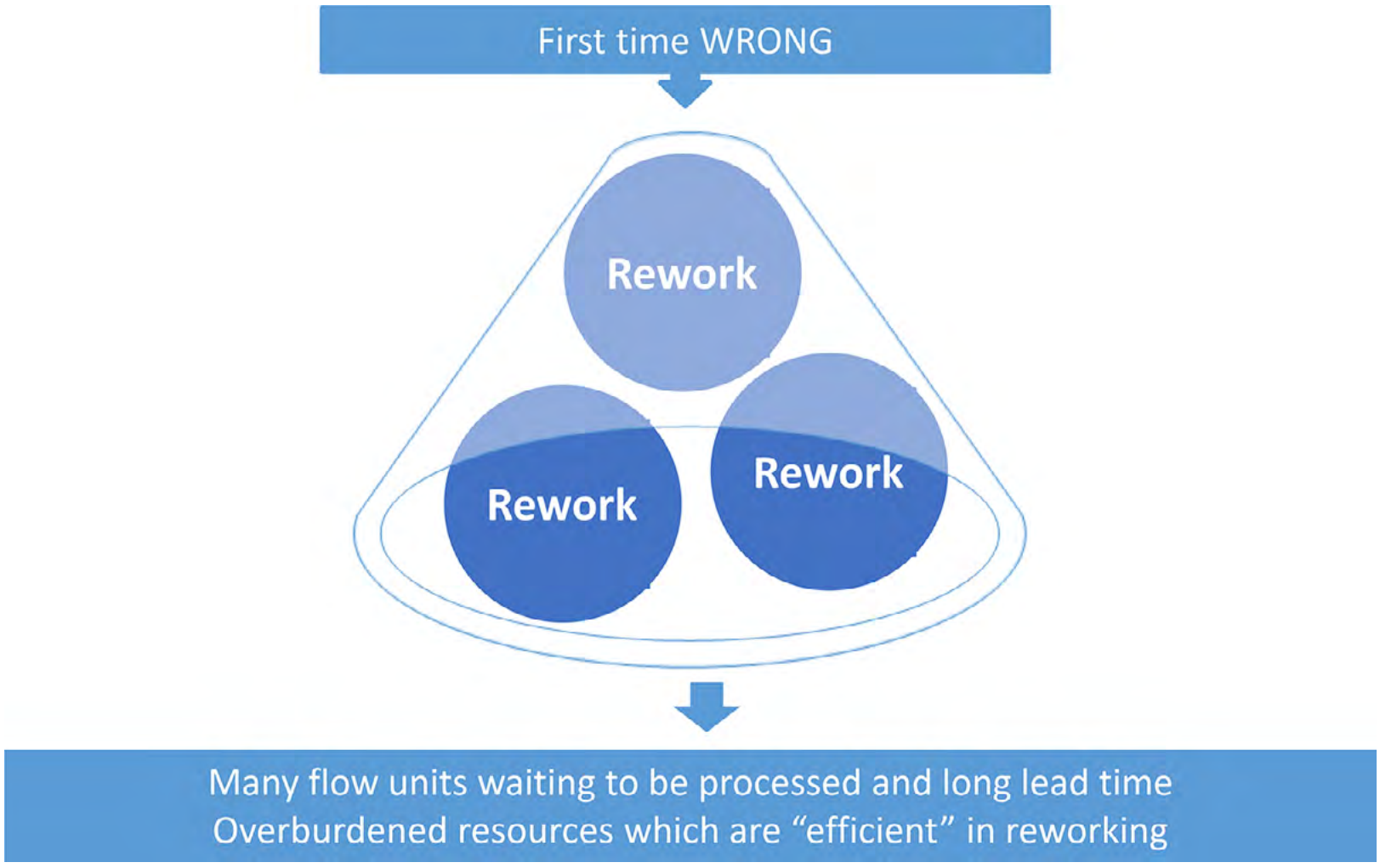

Figure 15 – Problems generated by doing “first time wrong”

Investigating the root cause of this problem (doing “first time wrong”) through the 5 whys methodology (Figure 16), two interesting facts can be noticed. First, often, the people who perform the service do not have enough know-how since employees with less experience are allocated to this function as interns, although the documentation requirement logic is complex. Thus, many problems that could be foreseen during the service go unnoticed and end up becoming pending issues later in the process.

Second, appointments are made in the system every 20 minutes, which is already a very short time – even to carry out basic activities. In this way, the intern who provides the service often knows there is a pending issue, and the applicant may have the missing document at that moment. However, they can spend a maximum of 20 minutes per customer because otherwise, they will

generate cumulative delays. That is, this “20 minutes” internal rule generates many delays later due to quality issues.

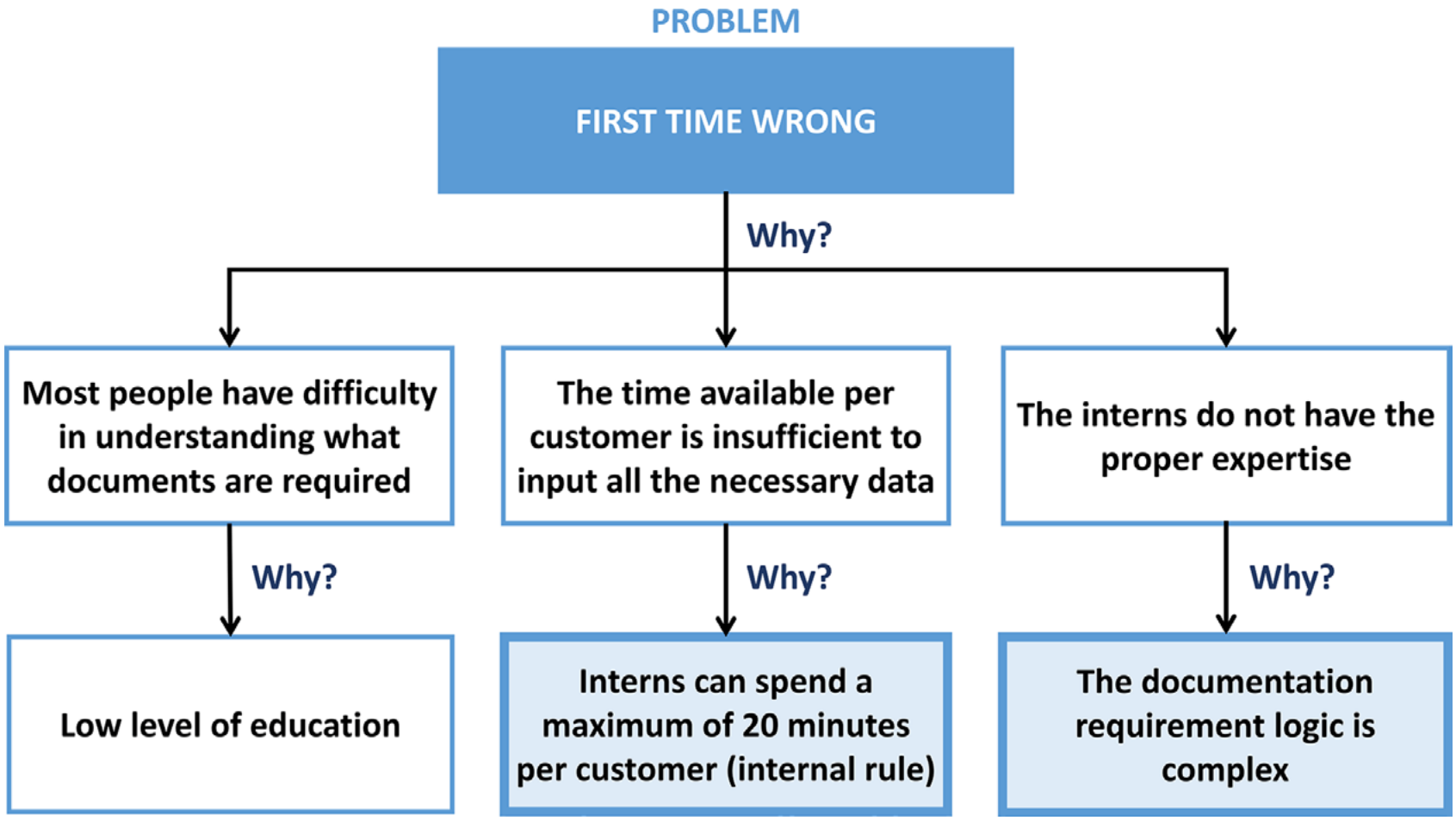

Figure 16 – 5 Whys root cause analysis

It is worth noting that the low level of education of the customers is a contextual problem. Therefore, we will not address this issue in depth or propose improvement actions on the subject. However, it is important to emphasize that the process must be designed with a friendly interface for any type of user.

This is an example that illustrates the productivity paradigm well. A poorly planned project created “efficient” islands where the customer’s needs were subdivided into redundant activities performed by resources with high utilization. In this case, the islands would be the analysts who are always busy processing many requests per day but finalizing a few. Thus, each island sees only itself and has no vision of the whole. The process, therefore, becomes suboptimal with several local “optima,” while the global optimum is compromised with this approach by creating secondary needs and reducing the percentage of value added.

IF PERFORMANCE MANAGEMENT PRIORITIZES A LOCAL OPTIMUM, THE GLOBAL OPTIMUM CAN BE COMPROMISED!

And how could we treat these two root causes? The problem is obviously not the interns but the fact that they accept requirements with missing documents – either due to lack of knowledge or due to their performance targets that encourage them, for example, to prioritize a 20-minute service even if the documentation is incomplete. Regarding the pre-established time, the question is simple: what can cause more problems, meeting the standard time of 20 minutes or extrapolating this time when necessary to avoid future reworks? Based on the process principles, not doing it right the first time leads to a domino effect of rework. So basically, the interns must be retrained to acknowledge that documentation should be prioritized over time.

WE DO NOT HAVE TIME TO DO "FIRST TIME RIGHT," BUT INCREDIBLY, WE ALWAYS HAVE TIME FOR THE REWORK THAT ARISES FROM THAT BAD CHOICE.

Regarding the complex logic of documentation requirements, it would be possible to map all the necessary documents in an analysis and create a checklist in the system (Figure 17). This would "oblige" the interns to only enter the data if all the necessary documentation was presented. If documents were missing, the interns would guide the applicants about the pending issues and schedule a new appointment to complete data entry. That is, the first data input must be a filter that only allows the entry of requirements with all the necessary documents.

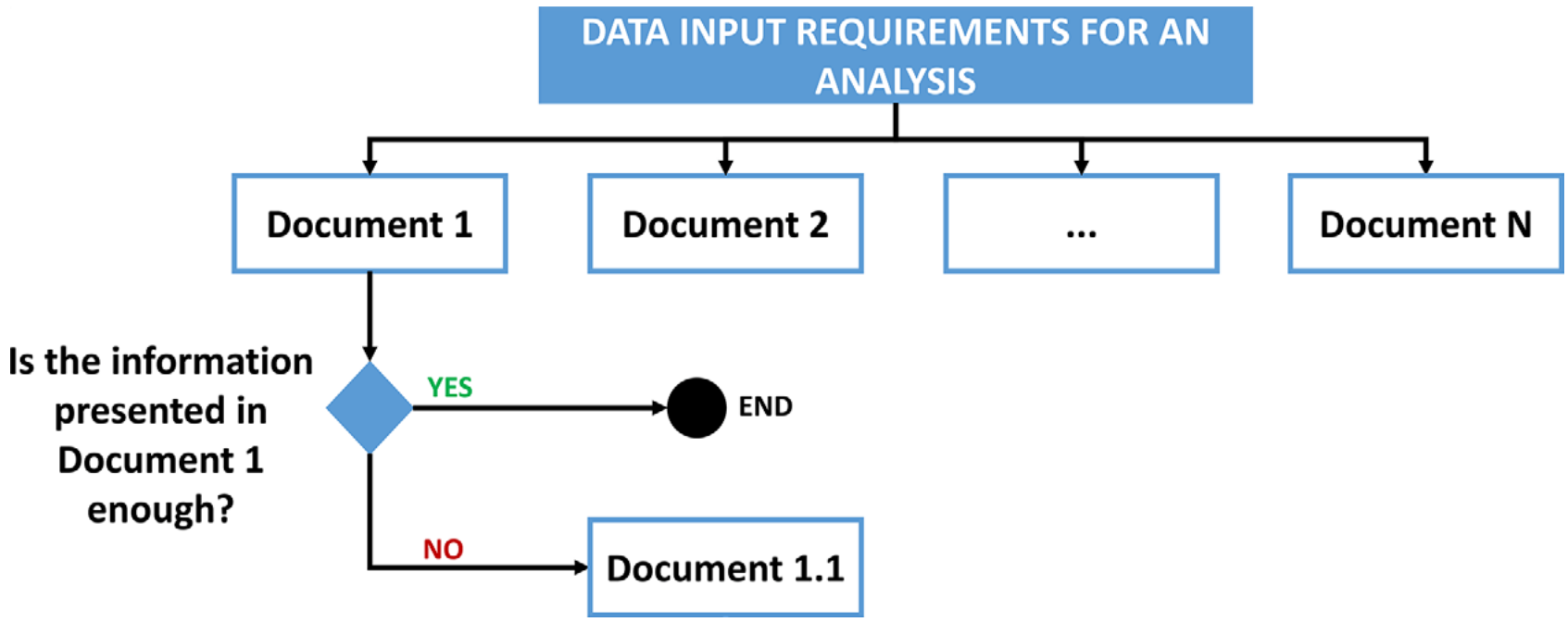

Figure 17 – - Flowchart of data input requirements

In conclusion, digital transformation really has the potential to bring many benefits to the process. However, if not properly planned, it can happen as in the case under analysis, in which the transformation brought another change: the quality deterioration of the input data. Therefore, the famous guideline of doing the “first time right” wasn’t followed, and the process was inflated with camouflaged waste, increasing the lead time to satisfy the flow unit’s needs.

TOWARDS A LEAN DIGITAL WORLD

Digital transformation, like “being lean,” is a gradual process of improvement in which digital technologies are being incorporated into a process. For example: in the case study presented, improvement actions could be made – the applicant could himself upload the documents necessary for the analysis. However, critical thinking must be adopted to effectively achieve productivity gains and make the process leaner. The presented case study is a good example of this. It is about digital transformation, but a case that made the process less lean. Hence the importance of submerging beyond the superficial concepts of lean and digital transformation, to really understand in depth what it means to be lean digital.

DIGITAL TRANSFORMATION, LIKE “BEING LEAN”, IS A GRADUAL PROCESS OF CONTINUOUS IMPROVEMENT.

Now, we are going to go a step further, as we are working with processes in continuous transformation, in which more and more digital technologies are incorporated. Therefore, it will be presented how this same case study would be if the digital transformation led to a process with 100% value added. We conceptualize lean digital from a counter-definition; nothing fairer than presenting the opposite extreme in the sequence.

Databases with personal information are increasingly integrated and available online. Soon, analysts will not be needed to study an applicant’s information and reach a conclusion of approval or denial. It will be unnecessary to call, schedule, scan documents, and wait. In fact, in the future, when someone requests this service by clicking on a button on a website, an algorithm will be triggered in order to access the person’s data, and in infinitesimal fractions of a second, it will give the verdict (Routine 100% lean digital – Figure 18). It is important to emphasize that facial recognition technologies could be used, for example, to guarantee process reliability and prevent fraud.

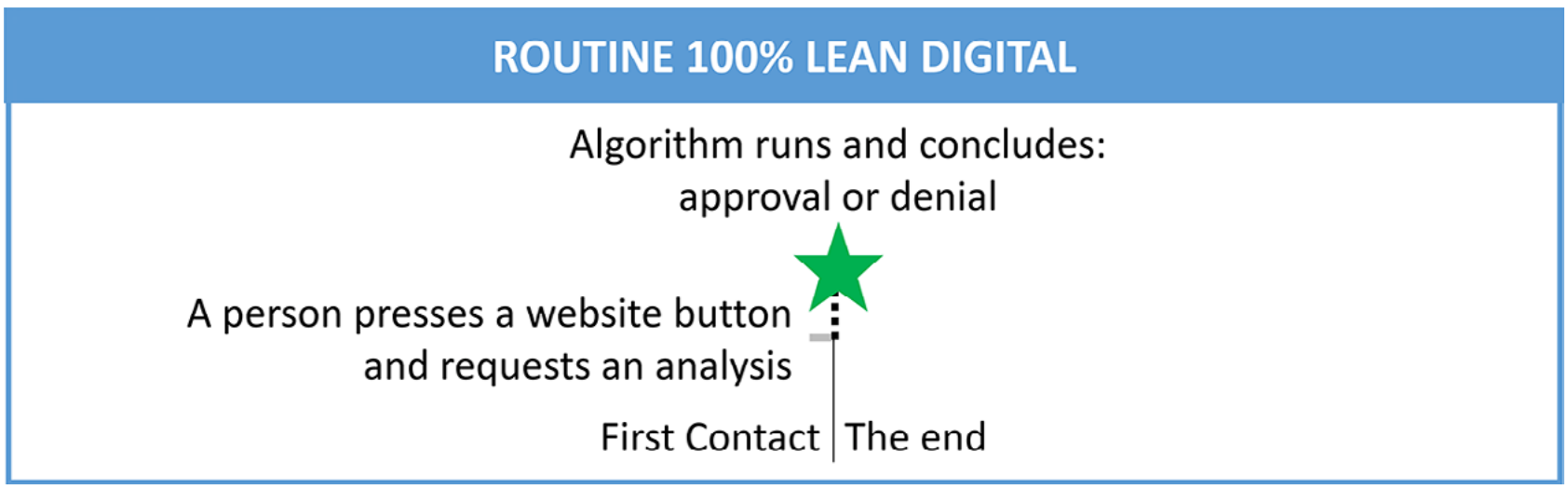

Figure 18 – Routine 100% lean digital

It is interesting to reflect that, in a process with 100% value added, the lead time can be in seconds, minutes, hours, or days. In the physical world, even if we streamlined the process as much as possible, the lead time would probably still be a few minutes. On the other hand, the digital world allows us to carry out activities that add value in fractions of a second. This is the world of possibilities and agility that opens.

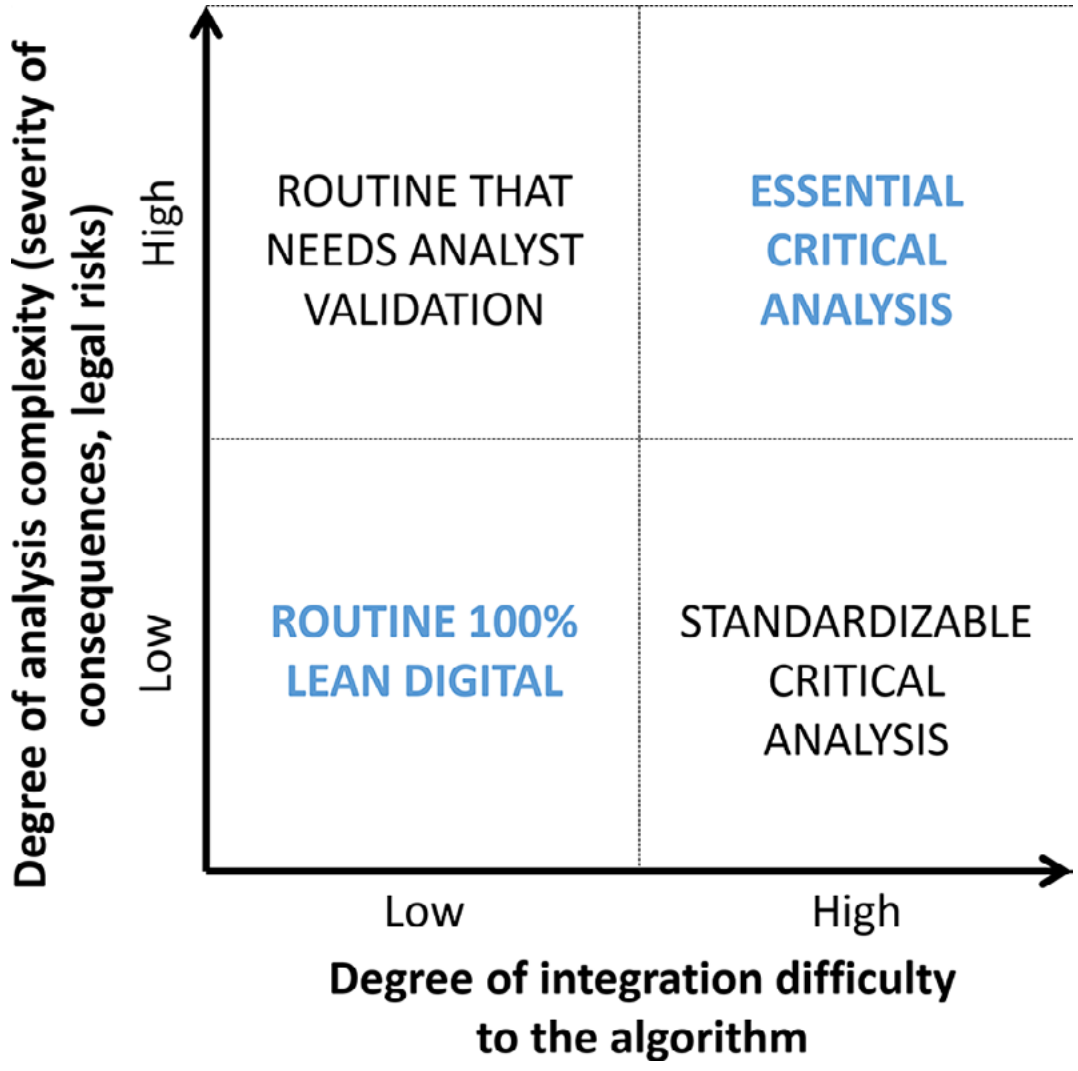

Figure 19 – Matrix to define the strategy to be adopted

Another point deserves reflection: some more complex cases will still require critical analysis by the analyst – because they have more severe consequences or because they involve greater legal risks. Thus, according to the matrix presented in Figure 19, we can evaluate the cases to be analyzed according to their level of complexity and the degree of difficulty of integration into the algorithm. Those that have milder consequences and are already foreseen in the algorithm are what we present in Figure 18 as "Routine 100% lean digital". When the risks associated with some analyses are substantial, and they are cases not foreseen by the algorithm, it is preferable that analysts remain responsible for this (Essential critical analysis – Figure 20).

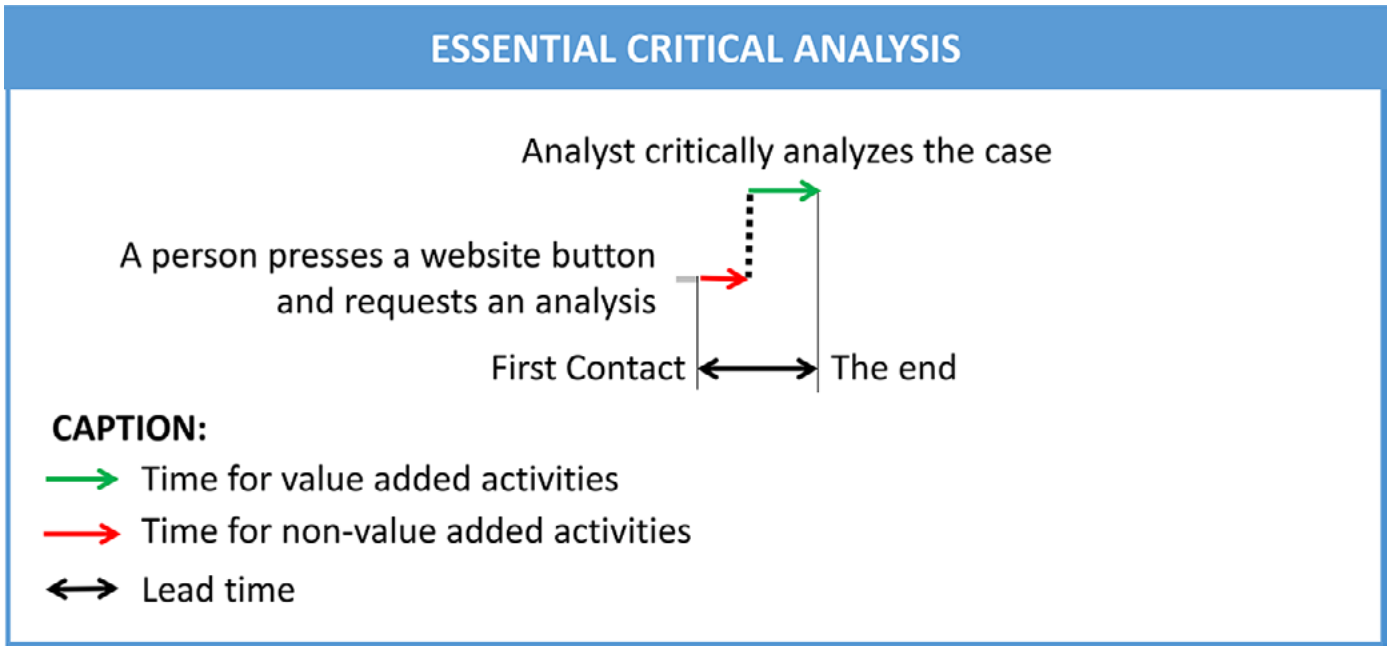

Figure 20 – Essential critical analysis

Eventually, there may be cases that the algorithm could analyze, but as they involve more severe consequences, they could be quickly validated by the analyst to mitigate unnecessary issues (Routine that needs analyst validation).

More complex cases may also arise, but the algorithm could learn these (Standardizable critical analysis). Thus, the first time they appeared, they would indeed have to be performed by an analyst, but the next time, they could be integrated into the existing algorithm, making it increasingly robust. That is, standardizable critical analysis tends to become a routine 100% lean digital. After all, artificial intelligence can be used in this continuous improvement process. Thus, one case can serve as a learning experience for other cases that may arise in the future.

In conclusion, the evolution of work accompanies the possibilities that are created with the evolution of Information Technology. Thus, there are certainly several questions regarding the psychology and sociology of work in the new digital world. Although this book does not develop the subject, it is worth emphasizing that this is a very important reflection to be made so that the transition occurs in the most harmonious way possible.

Another point of attention, verified from this example, is the increasing importance of an education focused on critical thinking to the detriment of the ability to repeat and copy. After all, we cannot control the future, but the more we try to predict and adapt to it, the greater the chances of being successful.

Now, this journey of transformation is over. I hope that the concepts of what it means to be lean and what it means to be lean digital are now clearer.

FINALLY, I CHALLENGE THE READER TO BREAK THIS PRODUCTIVITY PARADIGM: FIRST INTERNALLY; AND, ONCE THIS CHANGE OCCURS WITHIN US, MAY IT OVERFLOW TO TRANSFORM PEOPLE, PROCESSES, AND ORGANIZATIONS IN AN INCREASINGLY DIGITAL WORLD.

REFERENCES

Goldratt, E. M., & Cox, J. (2016). **The goal: A process of ongoing improvement.** Routledge.

Lira, E. G. (2020). **Estudo de tempos e movimentos: Uma abordagem lean para aumentar a eficiência de processos físicos e digitais.** Belo Horizonte: Ed. Da Autora.

Marchwinski, C., & Shook, J. (Eds.). (2003). **Lean lexicon: a graphical glossary for lean thinkers.** Lean Enterprise Institute.

Modig, N., & Åhlström, P. (2012). **This is lean: Resolving the efficiency paradox.** Stockholm: Rheologica.

Ohno, T. (1988). **Toyota Production System: Beyond large-scale production.** New York: Productivity Press.

Rother, M., & Harris, R. (2001). **Creating continuous flow: An action guide for managers, engineers & production associates.** Lean Enterprise Institute.

Rother, M., & Shook, J. (2003). **Learning to see: Value stream mapping to add value and eliminate muda.** Lean enterprise institute.