Lean or Self-Inflicted Chaos

- A case study on how hyper-growth firms prioritize between flow and resource efficiency

Jadberg, Niclas & Tavassoli, Arian

Bachelor thesis in Retail Management
Stockholm School of Economics
Center for Retailing
Supervisor: Wiley Wakeman
Submitted: 18th of May 2021

Abstract

The majority of previous research within the field of *lean* views variation in a process as a phenomenon to be minimized or eliminated. Some later research suggests that variation needs to be traded off with the potential customer value it can generate. In a company that experiences hyper-growth, there is an inherent presence of variation in the demand and operational processes. This raises the question, how do hyper-growth firms handle all of this variation and what are their priorities from top management regarding flow and resource efficiency?

By conducting a multiple case study on three companies that are currently in the state of hyper-growth, the aim is to determine the priorities between flow and resource efficiency through semi-structured in-depth interviews. The findings of this paper suggest that the companies experiencing hyper-growth deliberately inflict large amounts of what the authors define as *intended variability,* an extension of unpredicted variability, to optimize their customer offer. The companies prioritize flow efficiency which in turn requires costly buffers. However, the findings imply that through the collection of large amounts of user data, the companies can use predictive analysis to make intended variability predictable. This in turn enables the companies to achieve greater resource efficiency over time while maintaining high flow efficiency.

Keywords: Flow efficiency, Resource efficiency, Variability, Hyper-growth, Buffering

Table of Contents

1. Introduction

1.1. Background

In order to achieve operational excellence and become a *lean* organization, variation in your processes must be stamped out (Womack et al, 2007). That is the account, regarding variation in operational processes, presented in the legendary book *The machine that changed the world* that introduced lean manufacturing to the western world in 1990. What does this account mean for an organization with an inherent characteristic that makes variation impossible to avoid? Companies that are experiencing extreme demand growth are victims of that one inherent characteristic. Their growth will always inflict a lot of variation on their operational processes. Variation in a process is always met with a buffer of either capacity, time or inventory (Hopp and Spearman, 2008) and the pursuit of operational excellence consists of minimizing these buffers. This raises the question of how a firm in the state of *hyper-growth* handles all of this variation. In this thesis we will study three case companies that are currently experiencing this state and investigate how they use the aforementioned buffers in their pursuit of operational excellence.

Previous research within the field of lean is vast and fragmented (Modig and Åhlström, 2013). The dominant part of this research defines lean as tools with a low level of abstraction. To exemplify, elimination of waste, zero defects, pull-scheduling and continuous improvement are four out of eight lean production principles (Åhlström, 1998) that deal with the subject at a lower level of abstraction. Our definition and the foundation which sets the direction for this paper is on a higher level of abstraction, focusing on flow and resource efficiency, two more abstract concepts which can be achieved using these eight tools. Increasing abstraction is a way to avoid coming to context specific conclusions, and instead achieve a greater level of generalizability across, in our case, the hyper-growth domain. Using this higher level of abstraction we have identified a research gap in the lean literature related to the condition or *state* of hyper-growth. Our goal is to explore what management priorities dictate the companies' strategy regarding flow and resource efficiency to promote continued growth. As it turns out, customer centricity and predictive forecasting play an integral part in determining these priorities, which gives way for an interesting contextualization of existing literature.

1.2. Purpose

To leverage the research gap defined previously, we want to explore how a subset of lean production theory interacts with rapidly growing companies in their prioritization of activities. More specifically, investigating how hyper-growth companies manage demand variability, or rapid increases in demand by managing the trade off between flow and resource efficiency to achieve a competitive customer offer. Thus, the focus of the study will be to identify management priorities regarding flow and resource efficiency that apply to hyper-growth companies, the motivation for these priorities, and how we expect them to change as the companies mature. Therefore, our research questions have been defined as follows:

Q1: How do hyper-growth companies prioritize flow and resource efficiency with rapidly increasing customer demand?

Q2: How do they use variability, buffering and flexibility to achieve this?

Q3: How are they able to prioritize the way that they do?

2. Literature Review

This section contains a summary of previous lean and growth literature. It begins by describing the origins of lean and establishes why a specific definition must be made. The controversial concept of company growth is then explained before establishing a direction on which the theoretical frame surrounding hyper-growth will be built.

2.1. Lean

The concept of *lean* originates from the Japanese post-war automotive industry and what is referred to as the Toyota Production System (TPS) (Womack et al, 2007). It has since been popularized globally for more than three decades and has seen great success. This major popularization of lean has led to definitions being extended and interpretation has made the term *lean* in itself hard to use as nothing but a label, especially for research purposes (Hopp and Spearman, 2020). This fragmentation makes it important to begin by establishing a clear definition of lean.

In the 1970's, Taiichi Ohno released the book *Toyota Production System: Beyond Large-Scale Production*. Ohno is often referred to as the father of TPS and developed the later renowned Toyota production philosophy during his 60 years with the company. In this book, Ohno explains why old concepts such as economies of scale are inferior to the productivity that is achieved through the creation of flow. He introduced concepts such as removal of waste and non value adding activities from the process as well as highlighting the importance of realizing customer orders as quickly as possible without errors. (Modig and Åhlström, 2013)

A decade and a half later, the real birth of lean production took place when James P. Womack, Daniel T. Jones and Daniel Roos released the book *The Machine that Changed the World* after years of research on the Toyota Production System. They establish the four core principles of lean being (1) teamwork, (2) communication, (3) efficient use of resources and elimination of waste and (4) continuous improvement. Later the same decade, Womack and Jones released another book that was subject to a greater focus on implementation of lean called *Lean Thinking* (Womack and Jones, 1996). This book identified five principles to implement in the pursuit of becoming lean: (1) defining value from the perspective of the end customer, (2) mapping the value stream, (3) creating flow, (4) using a pull system instead of push and (5) pursuing perfection through continuous improvement of the four previous points. After this

book, it became increasingly popular to “leanify” organizations and the concept spread quickly over the globe. In the beginning of the 2000’s more books on TPS and lean were written and a real “lean boom” in both research and practice took place (Modig and Åhlström, 2013). Lean has become fragmented and the number of variations of lean are difficult to keep track of. Adaptations such as lean startup (Blank, 2013), lean health care (Kollberg, Dahlgaard & Brehmer, 2006) or lean service (Bowen, Youngdahl, 1998), to name a few, have all attempted to apply the original lean principles to specific industries. According to (Bhamu, Singh Sangwan, 2014), the only point of consensus in lean literature is that it is unanimously accepted as being more beneficial to implement than not.

2.2. Growth Stages

Many authors have suggested that there are certain stages through which companies move when going from idea to maturity, but the concept of startup growth stages is a heavily disputed area of research. One of the most cited examples of growth stage literature is Scott and Bruce (1987) who present a five-stage model following a sequence of (1) inception, (2) survival, (3) growth, (4) expansion and (5) maturity. Each stage experiences increasing levels of complexity, structural requirements, competition, product lines and decentralization. The aptly named stage 3, “Growth” is a post-survival stage characterized by the formalization of organizational structures into functional lines and expansion of product range. The authors identify two key crises which will cause challenges as the company moves into the expansion stage one as the demand for [or management effort spent on] expansion into new markets or products and the second as the entry of larger competitors.

These types of growth models have however been criticized by some researchers, as one review by Dobbs and Hamilton (2007) points out, for being more concerned with how to internally adapt in order to continue growth rather than explaining what causes growth. Levie and Lichtenstein (2010) attempt to find a conclusive definition of growth stages by reviewing 104 separate stage models. They found that previous research in the field of growth stage modeling has not agreed upon either a number of stages, attributes nor categories of attributes. This indicates that there is no conclusive body of literature regarding this stage model theory. They find that the majority of models rely on four distinct empirical sources, some of which have clear limitations. Examples of limitations are only looking at a small subset of secondary data from homogenous companies (Greiner, 1998), omitting central parts

of quotes on which the company life-cycle theory is built (Lippitt and Schmidt, 1967). Furthermore, they find that 30% of the models link to several previous stage models who are inherently different, and that 44% of them do not have any theoretical connections to other stage models at all.

In summary, the existing literature lacks consensus on how companies move between stages. Levie and Lichtenstein proceed by proposing a model without the use of "an imminent program of development" called a 'Dynamic states model'.

Dynamic states model, Levie and Lichtenstein (2010)

	Stages of growth models	Dynamic states model
Assumption	**Organizations grow as if they were organisms**	**Each state represents management's attempts to most efficiently/effectively match internal organizing capacity with the external market/customer demand**
Propositions: what	Configuration of structural variables and management problems	Configuration of structural variables and organizational activities (aspirations)
Propositions: how	**A specific number of progressive stages**	**Any number of states**
	Sequence and order is predictable	Sequence and order may be predictable depending on context
	Incremental and punctuated transitions	Incremental and punctuated transitions, and emergence
Propositions: why	**Immanent program of development**	**Adaptive process of retaining the sustainability of a business model**
	Prefigured rules of development	Interdependent rules for development
	"Regulated" by environment	Driven by market change and opportunity creation

† Major differences shown in bold font.

The model proposed by Levie and Lichtenstein (2010), while disruptive against existing stage-model approaches, provides interesting and applicable aspects to the cases used in this report. Our paper is less concerned with the general sequential steps leading up to (and succeeding) the current state of high revenue-growth. Instead, we are interested in precisely matching the internal organizing capacity with customer demand in a 'state' (rather than stage) of growth.

3. Theoretical Framework

In this section, we will immerse the reader into the theoretical framework that will be used to analyze the collected data. It begins with motivating why our definition of lean requires a higher level of abstraction, before explaining our selected definition, the flow lens. It concludes with the definition of hyper-growth that will form the basis for the case company selection.

3.1. Levels of Abstraction in Lean

Most of the literature on lean is conceptualized at a low level of abstraction (Modig and Åhlström, 2013). An issue with applying these definitions and ways of working with lean is that they are highly context-specific, meaning that one definition of lean may not be applicable in a different context. An example brought forward by Modig and Åhlström is that the optimal process of successfully producing an apple is probably not translatable as the optimal process for producing pears. The low level of abstraction makes it hard to reach generalizability in lean research. In order to reach a higher level of generalizability in this paper, and since our case study is conducted on three companies within different industries, a definition with a low level of abstraction would not have been appropriate. Instead of looking at context-specific tools from TPS such as pull system instead of push, we will look at concepts that can be defined as outcomes or sources of these tools such as variation and buffering. Since the essence of this paper is to examine how hyper-growth companies interact with concepts of lean and these firms inherently see a lot of variation through their growth, we have chosen a definition of lean that focuses on combating just that, variation.

3.2. The Flow Lens

Hopp and Spearman (2020) interpret the lean concept through four distinct lenses, each with different levels of abstraction:: *The process lens*, focusing on waste elimination, *The flow lens*, focusing on elimination of buffers, *The network lens*, focusing on minimizing the cost of waste and *The organization lens*, focusing on organizational cultures that encourage continuous elimination of waste. For the purpose of this paper we will focus on the flow lens, or the elimination of buffers, since they as previously stated, are the side effects of variation in a process.

The flow lens takes its starting point in *Factory Physics,* a book first released by Hopp and Spearman in 1996. It is an attempt to systematize the behaviour of a manufacturing system by applying an approach similar to that of physics research with laws and corollaries. The definition of lean through the flow lens is "to minimize the cost of excess inventory, capacity or time" (Hopp and Spearman, 2020). It builds on the *Variability-Buffering Law*, one of the fundamental laws of Factory Physics that states variability in a production system will be buffered by some combination of inventory, capacity and time" (Hopp and Spearman, 2008, p. 309). Variability is defined as any deviation in either supply or demand that is not the case of absolute regularity. Furthermore, Hopp and Spearman (2008) present a corollary to the variability-buffering law with the aim of making buffering more efficient, namely the *Buffer-Flexibility Corollary*. This corollary states that "Flexibility reduces the amount of variability buffering required in a production system" (Hopp and Spearman, 2008, p. 313).

3.2.1. Variability and Buffering

Traditional lean literature regards variation and complexity in processes as the main cause of why errors occur within a process. Womack et al (2007) suggest that all forms of unpredictable "craftsmanship" should be removed from processes and that it is of no use in any type of manufacturing and an organization should, as they phrase it: "Stamp it out". Spear and Bowen (1999) further developed on this in their study of TPS where they establish that variation always brings about poorer quality and higher costs. One of the most influential implementation research on lean was *The Toyota Way* (2004) by Jeffrey Liker who establishes 14 principles that constitute the philosophy of Toyota. Through these principles, Liker also asserts that standardized tasks without variation is the key to achieving an effective production system.

A way of breaking down variation is presented by Modig and Åhlström (2013) where they ascribe the emergence of variation in a process to three different factors: resources, flow units and external factors. Variation in resources can be caused by technical breakdowns, variation in individual competencies or even the mood of the individual that performs the process on a given day. In some processes, the units that flow through it vary in complexity and time needed which will cause a variation in the amount of resources required. The last factor that causes variation is that of external roots. This is often classified as variation in demand through seasonality or societal changes for example. In this paper, the companies studied

experience all three of these factors affecting variation in their processes but the one we will focus on and the one they all share in common the most is the external factors, or more specifically, rapid demand growth.

To further elaborate on variability we look to Hopp and Spearman (2020) who divide it into two factors that affect how an organization should handle their variability, namely predictability and customer value. Predictable variability such as a planned machine maintenance can be buffered only when needed since you know when the deviation will occur. Unpredictable variability (e.g. a power outage) on the other hand needs to be buffered continuously (e.g. with a backup generator) because you never know when it will occur. The second factor of customer value relates to the variability vs. variety dilemma. Hopp and Spearman bring up the initial production of the Ford model T as an example of a clear choice of low variability instead of high variety since they only produced one model in one colour. If a car manufacturer wants to sell cars in more than one model and colour they need to let more variability into the process. In today's competitive landscape however, car manufacturers have found a worthwhile tradeoff in offering a greater variety as it increases value for the customer and as a consequence: increased demand.

The *variability-buffering law* from Hopp and Spearman (2008) explains the inevitable relationship between variability in a process and buffering. That means that any variability in either demand or supply needs to be compensated with a buffer in either *time*, *capacity* or *inventory*. To exemplify this we can think of a manufacturing company producing shoes and clothing that sees a surge in demand for shoes. In order for the demand to be met, a buffer in one of three previously mentioned types is needed. The company can have an inventory buffer which enables them to send out previously produced and stored items. The second alternative is to have a capacity buffer, meaning that the company has excess production capacity and as the demand increases they can simply increase their production. The third and final way of buffering for the variation is through time, this is most simply explained as a potential delivery delay for the customer.

3.2.2. Flexibility

The second concept from Hopp and Spearman (2008) *Factory Physics* is the buffering-flexibility corollary. This introduces *flexibility* as a tool to keep buffers as low as possible. To

go back to the example from buffering with the manufacturing company. If the same event occurs with a surge in demand for shoes we established that the company will need a buffer in either time, capacity or inventory to meet the demand. If the company would have flexible staff so the clothing producers could help the shoe producers to meet demand, flexibility in the workforce will enable the company to meet the demand with less buffering.

Table 1: Example of buffering types and flexibility as a solution, interpreted from Hopp and Spearman (2008)

Buffering type	Inventory	Capacity	Time
Descriptive	Pre-manufacture shoes to keep inventory in case of demand surge	Hire more shoemakers to be ready in case of demand surge	Make customers wait for shoe manufacturer to catch up on back-log orders
Adverse effect	Cost of inventory	Cost of excess capacity	Loss of customers/revenue
If flexibility is present	In the event of a demand surge for shoes, clothes makers and other staff can shift assignments to respond to the temporary under-capacity. After the surge, all production units return to their regular assignments, thus reducing the total risk of under-capacity in the system.		

3.2.3. Flow Efficiency and Resource Efficiency

The traditional way of looking at efficiency is through the lens of resource efficiency. Since the industrial revolution, organizations have tried to make full use of their resources by maintaining high utilization of production equipment (Hopp and Spearman, 2020). Modig and Åhlström (2013) shift focus away from the resources in a process and instead see the unit that flows through the process as the focal point. By measuring the amount of value-adding time (the time in a process where value is added to the unit) and dividing that with throughput time (how much time it takes for one unit to complete a process) they present a measurement of flow efficiency. The unit could be both a physical product that is being produced or information or even a customer flowing through a process.

Furthermore, Modig and Åhlström (2013) put variability into the perspective of flow efficiency. Flow efficiency is derived from value-added time over throughput time. As long as the denominator and numerator increase in proportion, the flow efficiency will be constant. More specifically, as long as the activities due to potential variability, add value to the same relative extent as to which it prolongs the throughput time, the flow efficiency will not be affected. The concept of flow and resource efficiency can be directly correlated to the concepts of variation and buffering presented in the flow lens. If a company is able to eliminate all buffers they are in *the perfect state* which means that both flow and resource

efficiency is high. If the buffer used to handle variation within a process is time, then flow efficiency is low but resource efficiency is usually high since they will never run out of work. If the buffer used is capacity, flow efficiency is high whilst resource efficiency is low. Finally, an inventory buffer will affect both flow and resource efficiency negatively compared to if the buffer were not necessary.

$$\frac{\text{Total value-added time}}{\text{Total throughput time}} \times 100 = \text{Flow efficiency}$$

$$\frac{\text{Total time of utilization}}{\text{Total time}} \times 100 = \text{Resource efficiency}$$

3.3. Defining Hyper-Growth Companies

Summarizing the disputed area of growth stage research, we conclude that this paper aims to bridge this gap between the internal orientation of a growth stages model, and the currently lacking explanation of why companies grow by investigating what priorities between resource and flow efficiency are made in order to promote continued growth. Using the internally oriented literature as a support for understanding these decisions, we are adapting the stage model outlined in Scott and Bruce's by removing the disputed movement between stages, and only considering an *isolated state of growth.*

3.3.1 Hyper-Growth

The three focal firms of this report exhibit similar challenges and characteristics as those growing organizations described by Scott and Bruce (1987) which was discussed in the literature review, however, the general orientation of their paper does not sufficiently emphasize the added pressure and complexity of being a *rapidly* growing company. This report explores the characteristics of companies that have experienced significant growth in a short period of time, all achieving between SEK 30-500 million in revenue during 2019. High-growth literature has, similar to the growth stage literature, no universal definition. Chan et al 2006 studied a sample of 50 companies with the prerequisite simply stated as "high business growth for three or more consecutive years, with between revenue 10-1000 million CAD ". The best attempt found at establishing a simple label is proposed by Cassia and Minola (2012) who suggest a minimum of 20 percent annual sales growth for at least four years consecutively, called *hyper-growth*. Sales growth is used as a proxy for overall company

growth in profit, employees, assets, etc. as these are found to be successors of such growth, supported by Flamholtz (2007). Meanwhile, there are documented arguments against accepting a simple standard growth term as universal. Delmar et al (2003) state that 'high-growth' is an impossible term to generalize across industries, because of the many ways in which it can actually be measured. The authors conclude that sales growth appears to be an appropriate measure only for certain kinds of firms. While acknowledging this criticism, we find that this paper constitutes a case where sales-growth is an appropriate measure for hyper-growth. The motivation for this decision is that high demand growth is a direct antecedent to high revenue growth. Since high demand growth is also the basis for high demand variability, this definition simplifies and focuses the growth term in the appropriate scope for this study.

We conclude with a definitive definition of hyper-growth firms as having a minimum 20 percent annual sales growth for the last four years. And in order to exclude excessively small companies we have also looked for growth in absolute sales to be significant with a minimum of SEK 10 million in turnover per year.

4. Method

In this section, we will motivate why a qualitative case study method has been used, and why a grounded theory approach was important for the theory generation in this paper. Next, we will give an account of the data collection by presenting the case selection, interview procedure, interview objects and how the research questions evolved throughout our process. Finally, we will discuss the methodological limitations and critically evaluate the data collection.

4.1. Choice of Method and Research Approach

Based on the research question, we have conducted this study with a qualitative method. A qualitative method is one that aims to create an understanding of the relationship between theory and research, where emphasis is placed on theory generation and where the context in which the phenomena is interpreted matters (Bryman and Bell, 2013). As described by the research question, this study aims to create an understanding of the prioritization between flow and resource efficient operations in the context of hyper-growth companies that experience rapid increase in demand, or demand variability. Due to the explorative nature of this research question, a quantitative method was dismissed as there was no clear hypothesis that the study could be built upon.

4.1.1. Case Study Approach

This qualitative study aims to understand the priorities of management in a context of leading a hyper-growth firm. Therefore, it was of specific interest to investigate how, in reality, these kinds of firms make priorities. The case study is one of the most used and powerful research methods within the field of operations management (Voss et al, 2002). Yin (2014) broadly defines a case study as:

> An empirical inquiry that investigates a contemporary phenomenon in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident. (p. 16)

The aim of this study fits into this description, as we are interested in investigating a real-world case of flow and resource efficiency priorities, and assume that the context of a

hyper-growth firm has implications on the result compared to if it were any other type of firm. The author proposes that the case study method is especially interesting if the formulated research question intends to find *how* or *why* a phenomenon occurs, which is congruent with the formulation of our research question. Because we are interested in understanding the research question in the specific context of a *hyper-growth firm*, we argued that observing a single hyper-growth firm would not allow for sufficient generalization regarding this category. Therefore, we decided on a multiple case study that would allow us to include multiple companies and then draw a single set of "cross-case" conclusions based on the similarities between them (Yin, 2014).

4.1.2. Grounded Theory

In addition to a case study approach we adopted the principles of a grounded approach to develop the theory surrounding the study. The grounded theory method is appropriate when trying to understand a phenomenon by studying how individuals (in our case, hyper-growth firms) interpret reality (Suddaby, 2006). The two key concepts that Glaser and Strauss (1967) described as the key components of grounded theory research are *constant comparison* and *theoretical sampling*. This was the cornerstone of our approach, as we systematically compared the data collected in our interviews with the existing literature and did analysis continuously as part of the sequential step in which we added interview objects. As conceptual frames of our findings took form throughout the process we adjusted the interview script (Appendix) to collect the necessary data in order to approach theoretical saturation in the subject. Additions to the interview guide were made in an effort to produce more comprehensive and perhaps conflicting accounts on what priorities are made by management at the different case companies.

4.2. Data Collection

4.2.1. Selection of Cases

The selection of cases in this study follows a literal replication logic where the chosen replications have altered conditions that we consider unimportant to the aim of this study (Yin, 2014). As was discussed in the literature review, there is no scientific consensus regarding the sequential steps through which start-ups grow, or the *"*Prefigured rules of development regulated by the environment*"* (Levie and Lichtenstein 2010). This implies that

selecting cases based on company age, employee growth or other numerical indicators are subordinate to the chosen criteria of *high demand growth* which is the most easily discernible cause for demand variability and thus congruent with the theoretical underpinnings of this study.

The cases and interview objects were added in sequential steps according to the theoretical findings from analyzing each individual case on a continuous basis. The primary and overarching characteristics were that the case company should be in a state of hyper-growth as defined by Cassia and Minola (2012) with a minimum 20 percent annual sales growth in the last four years. Secondly, we determined that the selected cases should operate in an environment where the operational processes, as in needing to move products from one place to another, are of critical importance to the success of the company. The motivation for this decision was that the lean body of literature was conceived from and pertains primarily to production companies. While a plethora of emerging research has attempted to apply lean principles on service, healthcare and banking, the core literature is most applicable to companies with a product flow. In addition to these criteria, we wanted the companies to be in different sectors, in order to promote generalizability and not limit the study to an overly specific group of e.g. “hyper-growth personal mobility companies” which would not sufficiently emphasize the state of hyper growth as the distinct common characteristic.

4.2.2. Interview Approach

In total, we conducted seven semi-structured interviews following the best-practices highlighted by Bluhm et al (2010), who concluded that semi-structured interviews should be held with open-ended questions and clarifying follow-up questions to topics related to the study. Furthermore, Bluhm et al suggest that interviews should be held with employees in executive power over the decisions regarding the topic relevant to the case. Lastly, they establish that interviews should be held with more than one informant per case in order to ensure data quality. Interviewing *managers* can give birth to especially interesting insight, but their status and power often make them difficult to get access to (Bryman and Bell, 2013). Thankfully, this was not an issue for us as we were fortunate to have access to middle or top management in all three cases in this study. Each manager interviewed had significant influence or executive power over strategic decisions in order to reflect the true management intentions regarding the interview topics for each case company.

Table 2: Interview objects of this study, ordered by company.

Label	Position	Company	Company age (years)	Tenure (years)	Interview length
Manager 1	Chief Operating Officer	Reviver	7	6	44 min
Manager 2	Head of Business Development	Reviver		2	46 min
Manager 3	Process Engineer	Reviver		5	31 min
Manager 4	Product Manager	P-mob	3	3	70 min
Manager 5	Fleet Optimization Manager	P-mob		3	52 min
Manager 6	Head of Logistics	SPS	5	3	45 min
Manager 7	Business Developer	SPS		2	39 min
				Total	327 min

An interview guide was established before each interview, starting with open and broad questions such as “What do you do at Reviver/P-mob/SPS” and gradually narrowing down with e.g. “What problems do you face in normal operations” and finally explicitly asking “What sacrifices have you made to flow efficiency in order to achieve greater resource utilization”. Some follow-up questions were pre-determined while others happened organically as the conversation took form. A conscious effort was made to not force any explicit answers to our research question in the early-to-middle stages of the interview, so as not to invite stylized answers towards the end. There were two revisions made to the interview guide. The first revision included “what considerations do you make when entering a new market” with the purpose of understanding the company’s approach to a nascent state of operation. The second revision included the explicitly stated flow and resource efficiency questions which were intended to give the interview object an attempt to identify these phenomena within their companies themselves. (Appendix)

4.2.3 Data analysis

Following a grounded theory approach, the data analysis occurred in parallel with the data collection and literature review. The first interview acted as a pilot and quite substantially changed the direction of this paper. It was here that we realized that the lean concept as a whole was too broad for the purpose of this study. After the pilot interview we returned to the literature in order to find the appropriate definition of lean, a process from which the focus on Hopp & Spearman’s flow lens emerged. In all the interviews, the answers were recorded,

transcribed, thematized, revised and analyzed for connections and contradictions. The recording was conducted with a conscious effort to not collect any personal data. Directly after each interview, a meeting was held between the authors to capture spontaneous reflections which were documented. Within two to three days, each interview was transcribed by one of the authors and proof-read by the other. A conscious effort was made to alternate the transcription such that both authors transcribed one interview from each case company. The transcribing author was tasked with identifying themes from the interview. The main themes identified are reflected in the structure (i.e. headings) in the empirics and analysis section of this paper. After the thematization of each interview, the findings were analyzed and discussed in relation to all the previous interviews, notes and existing literature during what can be described as a self-led seminar. This was usually the step in the analysis where we realized if a pivot was necessary, both in terms of interview script (as described previously) and supporting literature, in order to get the data necessary to continue our analysis. After each interview, our understanding of the subject and common characteristics of the case companies grew as we approached saturation.

4.3. Research Question Development

As the study progressed, the research questions became more focused and relevant based on our increasing understanding of the topic and case companies. The initial question of "how can lean operations be implemented in high-growth startup" was changed as the realization occurred that the concept of lean is much too broad to be captured within the scope of this (or arguably any) single study. The first major revision was defining the lean concept in terms of the prioritizations between flow and resource efficiency, a subset of the lean literature. By revisiting the literature underpinning the theoretical frame in between interviews, we identified the source of variation in relation to the *flow lens* described by Hopp and Spearman (2020) we thus landed on the final research question: *How do hyper-growth companies prioritize flow and resource efficiency with rapidly increasing customer demand?*. In order to achieve a greater depth in the answer to the main question, two follow up questions were posed as well: *How are they using variability, buffering and flexibility to achieve this?; Why do they prioritize the way they do?*

4.4. Data Reliability and Methodological Limitations

The quality of research design for multiple-case studies can be evaluated by its construct validity, external validity and reliability. A valid criticism would be that we have not fully triangulated the collected evidence in a systematic way. While we have strived for construct validity by using several interview objects in each case company, the other sources of information have been personal experiences with the products offered by the case companies. This allowed us to confirm the different kinds of customer offers claimed by the companies, but could arguably have been deeper using external sources. External validity has been sought by not considering the cases as samples representing the whole population of hyper-growth companies, but rather giving enough breadth in types of case companies to make relevant analytic generalizations about the concepts of flow-efficiency in any hyper-growth context. Reliability has been attempted by being transparent and documenting the number of time spent and content planned in the interviews. We assess that any researcher that applies the same theoretical framework to our three focal companies will reach similar, or at least not contradictory conclusions.

Bryman and Bell (2013) discuss the practical limitations of grounded theory in many research projects as one common reason for why the full extent of the method is often not followed., Facing similar limitations, we categorized the material in parallel with the transcription of the data but did not compile it into one comprehensive document which constitutes different levels of *coding*. Given the scope and time constraint of this thesis, we did not reach complete *theoretical saturation* in our data collection either.

It is also worth noting that P-mob has only existed for three years as of writing this paper, which does not fit within our defined case selection criteria (four). We argue, however, that this is not reason enough to dismiss the case, because the expected revenue growth for the coming year supersedes the 20% amount by quite a margin. As this is the fastest growing company out of the three selected, we were adamant about still including it in our paper. We decided against changing the selection criteria based on the case companies when existing research says otherwise, since it is not sound research practice to do so.

Finally, due to the physical meeting restrictions caused by the Covid-19 pandemic, 6 of the 7 interviews were conducted remotely by digital video conferences. We would have prefered to

have more meetings in person, as these are likely to yield more casual conversations and passive insights that could be useful (note that the one physical interview was the longest at 70 minutes). However, we also appreciate the fact that the general population's increase in comfort with having remote meetings positively affected the level and number of interviews we were able to achieve. Three of our interviews would likely not have been possible due to geographical limitations, who instead were able to squeeze us in between work meetings.

5. Empirical Data

In this section, we will present the three case companies separately, by briefly explaining each of their operations, flow and resource efficiency priorities and variability in their processes.

5.1. Overview of Companies in the Study

Table 3: Overview of companies in the study

	Reviver	P-mob	SPS
Descriptive	Second hand e-commerce	Personal mobility company	Last-mile logistics platform
Revenue 2019, MSEK	100-300	300-500	10-50
AAGR	144%	3271%	957%
Employees 2019	226	409	10
Most important KPIs	minutes per task, time to dispatch	service availability, service reliability	stops per minute, Co2 emission per delivery

Note: Numbers retrieved from allabolag.se (2021-05-15), KPIs retrieved from case interviews.

5.2. Case Company 1 - Reviver

Reviver is a secondhand e-commerce company that was founded in 2014. They sell their products through their own platform as well as third party secondhand marketplaces. The products sold by the company through these platforms are from customers who send in items they no longer want or need, this means that they have customers on both ends of the spectrum, both their sellers (products in) and their buyers (products out).

Reviver has grown exponentially since its inception with annual growth rate ranging from 50% to 400% from year to year. Today, Reviver is active in seven markets and is rapidly expanding geographically with another 20 market entries planned for the upcoming year. The company has around 400 employees where 350 of them are active in the operative function of the company and the rest are active within overhead functions.

5.2.1. Reviver's Operations

Revivers business as a second hand e-commerce involves unique and very complex operations. They accept all items that the sellers can fit into the designated bag that Reviver sends home to the selling customer on demand. Once the bag is filled by the customer, it is picked up and taken to one of the production sites that Reviver operates. Upon arrival at the

production site the bag gets placed in a backlog of bags, waiting to be processed. At the production site the bag's content is sorted into different categories and described, photographed and packed before it is stored on the shelf as inventory. After the item is packed it is also posted for sale on their marketplace.

The process is labor intensive and even though automation has been introduced in certain areas, the vast majority of time spent on an item is still connected to human manual labor. Employees in the operative functions are often specialized in one department of the production process.

5.2.2. Flow Efficiency and Resource Efficiency Within Reviver

In Reviver's operations we define value adding time as the time spent on an item that generates value to it meaning time spent on describing, photographing and packing an item. The throughput time is the time elapsed from the first step in the process when a bag is ordered for pick-up until the items are packed and stored on shelf in their warehouse.

In its early days, Reviver had a major focus on moving the items quickly through the process in order to keep customers happy, but the process was still new and underdeveloped. Nevertheless, as soon as a bag came in from a seller it was instantly processed and non-value-adding time within the process was minimal. In this nascent stage Reviver had a clear focus on flow efficiency.

When Reviver had established a more stable presence in the market and initial capital was starting to run short they shifted their focus into becoming more resource efficient. At this stage, Reviver had started to grow rapidly and the amount of bags waiting to be processed in their backlog had grown. This in combination with their shortage of funds drove Reviver to shift priorities in an attempt to utilize all of their resources to the fullest extent possible.

> When we split up the production it was not optimal from a flow perspective because there will be more people touching the items but we need to do it in order to make it work from a financial standpoint. - (COO, Reviver)

They divided the process up into more parts in order to standardize work tasks at a specific station, they specialized photo stations so that one station only photographed one category of items. Through this they eliminated downtime as they no longer needed to adapt the working stations when items of another category arrived. This shift resulted in a higher resource efficiency aligned with their temporary goal until funds were stabilized.

As the company's exponential growth started to result in major volume increases they kept on focusing on their resource efficiency. They started to build large buffers between all steps in the process and their backlog increased rapidly. The buffers acted as a safety net in order to always be able to have full utilization of their resources and never run out of work in progress at any work station. To continue to grow the operations, most of the initiatives led to a further departmentalization of the production. This in its turn resulted in the flow efficiency decreasing drastically.

At this point in time, which is around the fall of 2020, a selling customer could need to wait up to 17 weeks in order to get their bag processed. During the past year, as Reviver turned their focus from their domestic market to a rapid international expansion, the focus began to shift towards flow efficiency as well. They needed to offer a better experience to all their new potential customers outside of their domestic market. During our interviews we asked how they would prioritize between flow and resource efficiency going forward.

> Today it is more about flow efficiency. In the beginning when we did not have as much money we needed to utilize every resource to its max. We use an airplane model with over staffing so we count on people being sick in order to not have too many stations empty. We still try to utilize the resources as much as possible but we would choose better flow over utilizing a resource to its maximum today. - (COO, Reviver)

5.2.3. Variation Within Reviver

Reviver is a company where variation is a major part of their day to day operations and variability is built into the business model itself. As described earlier, the bags in which the selling customers send in items they want to sell, they are free to pack whatever they want. This means essentially that Reviver does not know what inputs to their processes they will

get. The process for describing, photographing, packing and storing a large down parka jacket is vastly different from that of doing the same process with a video game.

This is one of the reasons why buffers have become such a vital part of their process, since they need a safety margin to cope with the potential variation in inflow of products. But as they are growing, the amount of data they have increases and with that they are able to turn previously unpredictable variability to more and more predictable variability.

> [Revier] has existed for 6 years, and we have learned to expect what to get in a bag. We know it's 90% clothes in the average bag. We know that a certain share will be sellable and another share not worth trying to sell. So through historical data, we can build capacity for different types of products and functions that we need to handle the amount that has to be sorted, photographed, packed and placed in the stock, and those that need to be donated. - (Expansion Operations Manager, Reviver)

Another way that Reviver encounters variation is through demand variability. There are many reasons for demand variation, it can be built into events that occur regularly on a weekly, monthly or yearly basis. These events can be weekends, where people tend to shop less, end-of-month pay-day when people tend to shop a lot more and major holidays such as Christmas creates even greater demand spikes. Even though these peaks in demand are subject to predictable variation, Reviver lacks the organizational flexibility to meet them. As described earlier the process is very labour intensive which makes it hard to meet these spikes with flexibility, therefore the result is often that time buffers (e.g. delivery delays) are used to meet demand, and remedy the variability.

> But we don't really have any flexibility in the short run to put in more people, because we have a very labour intensive process, and we don't have enough bodies to meet the demand and bring that up or down, because it takes time to hire people, and it's legally complex to get a staffing deal that can bring up and down people [labour capacity]. - (Expansion Operations Manager, Reviver)

The third type of variation that is constantly present within their operations is their growth. It is to some extent predictable in mature markets but as Reviver is aggressively expanding internationally, it is hard to predict the effects of it in a reliable way. Up until the fall of 2020,

Reviver had a steady growth with a potential customer base (number of inhabitants in their active markets) of 10 million people. In April 2020, this number had increased to 120 million people. This makes it hard to predict how much and when demand will increase. To cope with this current and future variability, Reviver has begun investing heavily in operations capacity internationally and currently holds excess capacity to meet the growth of their international markets.

5.3. Company 2 - P-mob

P-mob is a personal mobility company that was founded in 2018. They have since seen one of the most explosive growth journeys in the european start-up sphere with an annual growth rate of 3400% and a yearly revenue in the hundred millions (SEK). They have expanded across more than 50 cities across Europe and are expected to continue on a similar trajectory.

5.3.1. P-mob's Operations

As a personal mobility company, P-mob operations evolve around availability to the customer. In essence, if P-mob vehicles are not around when a potential customer wants to get somewhere, that customer is lost. Therefore their day-to-day operations are about deploying and moving vehicles to the optimal locations.

To execute the work they have third party logistics providers (3PL) that get tasks assigned for a specific area within a city. P-mob is responsible for the analysis and decision making that generates the tasks for the 3PLs. One of the main tasks is making sure that all vehicles are charged. To do this they operate charging facilities where batteries are charged and once a vehicle is low on battery, a task is assigned to swap the battery on that particular vehicle. Secondly, P-mob makes decisions on where vehicles should be placed in the beginning of each day to maximize availability. They also analyse daily operations and make rebalancing tasks in order to restore optimal positioning for availability throughout the day. The 3PL picks vehicles up and distributes them over the cities.

Another flow of the operations is that for the repair of broken vehicles. When a vehicle breaks down, a task is assigned to pick it up. Once the 3PL has picked up the vehicle they bring it to the closest warehouse for repair. At the warehouses the vehicle is fixed by P-mob engineers

and redeployed by the 3PL as fast as possible to ensure availability of the vehicles in the streets.

5.3.2. Flow Efficiency and Resource Efficiency Within P-mob

When looking at P-mob, we define throughput time as the time it takes from when a user opens the application and starts looking for a vehicle until their ride has ended. The value adding time is the one spent travelling with the actual vehicle. These two components are what makes up flow efficiency at P-mob.

One of the key objectives within P-mob is to ensure availability. In the context of flow efficiency this means that once a customer wants to move with the help of P-mob, they need to be as close as possible to maximize the flow efficiency. In the beginning this prioritization was more extreme, or at least the means to reach it. One manager told us:

> The whole thing about our first six months we burned insane amounts of cash. We threw out money in order to make sure the [vehicles] were available. [Back] then we used to [employ] with gig workers and the prices [wages] were fluctuating. If there were few people wanting to collect [vehicles] then we increased the price [compensation]. That could become 2x, 3x compared to base price. - (Fleet Optimization Manager, P-mob)

In the beginning P-mob did sacrifice huge amounts of resources in order to obtain a good flow efficiency. There were a lot of underlying incentives for this and the main focus from the start was to grow, fast. The investors wanted to see P-mob grow at a rapid pace and maintain a stable relative demand for their product. During this time, all decisions with regards to placement, replacement and general routines around executing operational tasks were subject to the local operations manager. This was not necessarily a bad thing but as with any execution of a process that is reliant on humans, it tends to produce variation. In this case, not all operations managers had the same competencies and the flow and resource efficiency could vary across markets.

There was also a lot of gut feeling involved in the forecasts of how many vehicles they would need at a certain place at a certain time and as flow efficiency was the number one priority,

resource efficiency was often sacrificed. However, this gut feeling did not always produce the most reliable result and from time to time they were operating with low efficiency from both parameters.

As P-mob grew at an incredible pace and proved that they can sustain a good demand and utilization of their service despite this expansion, the external demands and internal priorities shifted. The Fleet Optimization Manager told us “We went from 0 to 1 million rides in 6 months. That was a check to show that we could do that”. The next task for P-mob was to show that they actually could make money from their operations.

Despite the new priority, availability remained one of the most important measurements within P-mob. Since they no longer could operate with a low resource efficiency but could not compromise on their availability, or flow efficiency. Instead they had to innovate in order to stay flow efficient and increase resource efficiency. The greatest example was the implementation of swappable batteries for the vehicles, this change meant that they could swap batteries out on the streets instead of bringing the vehicles to a warehouse for charging. This was a huge leap for reducing downtime that went down from 11 hours to 30 seconds, or a cost reduction for charging of 50%.

5.3.3. Variation Within P-mob

Since P-mob is a personal mobility company there is a lot of variation built into the operations by nature. Every ride a customer takes with one of P-mobs vehicles is unique and unpredictable variation appears in where customers park the vehicle, what happens during the ride, how long the ride is etcetera.

The second source of variation for P-mob is that of variation in demand. Their product is clearly primed towards the warmer parts of the year and is barely used during the coldest winter months. Roughly 80% of all their usage takes place between April and October. There is also a lot of demand variation looking at a shorter time horizon such as variation throughout the week and even during the day. There are clear spikes in demand during mornings and afternoons during weekdays and evenings during the weekend for example. “How many resources do we need at what time? We can not really handle the peaks and that is not quite optimal. But it is not a catastrophe either”, the Fleet Optimization Manager told

us. P-mob's capacity is enough to handle roughly 80% of peak demand, a number they are striving to increase.

From the beginning there was a lot of manual gut feeling involved in creating the decisions on capacity and balancing. This is a third source of variation that makes it difficult to cope with other variation in an effective way. Today P-mob has centralized these tasks with a fleet optimization team located at their HQ that have developed more sophisticated ways of making the unpredictable demand more predictable without sacrificing the freedom of the customer to park and ride where they want in the city. As the usage of P-mob has grown rapidly, data from all the rides creates an opportunity to leverage that into predicting user behaviour and optimizing their fleet in terms of positioning, timing and down time.

> When we created the FO [fleet optimization] team we became more data driven. That is their legacy to go from shooting from the hip to becoming data driven and building ML models to predict demand better than if every respective person sits and does simple analysis. - (Fleet Optimization Manager, P-mob)

This leverage of data has helped P-mob to predict many sources of variation. Even though every ride is still unique, user behaviour is becoming more and more predictable with the help of data. The variation in demand is becoming more predictable on the smallest level possible which enables P-mob to optimize their fleet not only on a daily but hourly basis.

The launch of their swappable batteries also gave them tools to cope with variation through flexibility. As the process of charging a vehicle became so much simpler, the circumstances such as where and when became a lot more flexible. This gave them a tool to drastically lower the buffers needed in time and capacity to meet consumer demand.

Another way P-mob has worked with reducing variation, or making it more predictable, is by priming consumers to use their vehicles in a more predictable way. This has been done by initiating specific parking zones where if the vehicle is parked there, the ride gets discounted. They have also launched a monthly pass which enables the customer to ride as much as she wants during a month. This makes customers use P-mob as part of their daily routine in commuting which makes the use more consistent. P-mob's Fleet Optimization Manager said "Since we launched the monthly pass the continuity increases as well since someone might

use it to commute to work [etcetera] so there is variation but it is starting to get easier and easier to see the trends.", a tactic that has proven itself successful.

5.4. Company 3 - SPS, Swedish Postal Service

SPS is a technology platform offering last-mile product delivery solutions for e-commerce retailers. By subcontracting haulier companies under their own brand, SPS' platform bridges supply and demand for home-delivered products within 2-24 hours depending on the end-customers location. The company was founded in 2016 and has an average annual growth rate of 957%. Today, the revenue is in the magnitude of 10-50 MSEK.

5.4.1. SPS' Operations

SPS' operations are centered around the technology platform that matches customer (retailers) delivery demand with haulier supply. The product flow is very labour intensive but the majority of the process is outsourced to partner companies who operate on behalf of SPS. The product flow starts with an order booking from an e-commerce retailer located anywhere on the domestic market. SPS' contracted haulier picks up the product using heavy trucks during the first half of the day and delivers it to one of several terminals, the biggest one being located in the country's capital. Here, the inbound packages from all e-commerce retailers are sorted into different cages organized by the city of last-mile delivery. This separation of packages triggers two different product flows, one same-day delivery (SDD) and one next-day (NDD) delivery. The same-day delivery flow is prioritized first, where packages are put in smaller trucks for outbound delivery to the end-customer during the evening. After a certain cut-off time, the same-day delivery flow is paused and a fleet of smaller delivery vehicles leave the terminal to begin the SDD distribution. After this cut-off time, the NDD flow commences with a similar procedure stretching throughout the late evening. Packages are sorted according to the end-customer's city of residence and transported overnight for evening delivery the next day.

The value created by SPS lies primarily within their technology platform. When the booking of packages occurs, the software generates optimized delivery routes automatically and instantly on a daily basis. Once the packages arrive at SPS' sorting terminal, the sorting procedure is already predetermined according to this route optimization. Thus, the final step, or 'stint', of going from SPS' terminal to end-customer is decided already before the first stint

when the orders are picked up from the e-commerce retailers. This rapid information flow enables greater efficiency of executing the product flow, but has some tangible drawbacks when variations occur, which will be discussed later.

Managers at SPS explain a few interesting efforts to cope with the high demand growth for their services. Most of the physical labour is outsourced to haulier companies which allows for flexibility in their capacity to a certain extent.

> Our contracts [with haulier companies] are written so that we have one part guaranteed and one part flexible. This flexible part exists there to have a daily buffer available for the volume fluctuations. - (Business Developer, SPS)

This is a material difference compared to other logistics competitors who buy their own fleet of trucks and hire drivers, with the biggest regional competitor having over 1600 trucks in their arsenal. While this outsourcing strategy offers a more cost efficient solution, the flexible component can only supply so much capacity before it takes a toll on the customer experience, as the Business Developer at SPS explains: “we are very adaptable to [demand] peaks, [...] we have a foundational volume forecast that can comfortably go up by 10-20%, but if it goes up 50%, then we’re pretty … screwed”.

Another effort SPS has deployed is resorting to external terminal ‘hubs’ for loading the inbound packages to outbound delivery trucks. These hubs are facilities required during only about one hour each day, borrowed from external companies, and are used as a capacity smoothing tool when the amount of inbound packages start creating bottlenecks in the sorting process within their sorting terminals.

> ...what we see in peaks is that there is a bottleneck when everyone has to sort at 16.30-17, and we don’t have enough room for that. So these hubs located around [the capital] where we at 16.30 take the goods and the deliverers go there to load the trucks instead. - (Business Developer, SPS)

5.4.2. Flow Efficiency and Resource Efficiency Within SPS

To understand the flow efficiency for SPS, we define throughput time as starting when an e-commerce customer completes a purchase and ending when that customer's products have been delivered to their door. The value-added time is defined as when a package is being transported to be delivered as soon as the customer wants it (usually, but not always as soon as possible).

SPS as a company has created a new type of value network that is less dependent on fixed assets than what has traditionally been the case for logistics companies offering last-mile delivery. Being a technology platform at the core, SPS' solution has since their inception created a relatively flow efficient ecosystem, with few fixed resources that need increased utilization for SPS, but increased value added time for SPS and their haulier companies respectively. The aforementioned fixed plus variable capacity of drivers and trucks in combination with the routing optimization ensures that very few trucks or drivers stay idle during any given day. Manager 2 explains: "We don't own the trucks, so they can do other missions during the day for other businesses." This has further implications on the flow efficiency of SPS. The technology platform simplifies the processes and allows for any operating unit with the correct tools to conduct the work. In other words, if for example "Haulier Company 1" has a machine failure during the morning, then "Haulier Company 2" can effortlessly replace them and continue the value adding process. This implies that the variability is remedied with flexibility in the production units.

Their operational coping efforts mentioned in the previous section also have implications on flow and resource efficiency. The external hubs as an effort to handle demand spikes is in fact a tradeoff between the two. While utilization remains high throughout the described process, the true value added time suffers. The lack of capacity, or terminal space, creates a bottleneck that results in secondary needs by requiring packages to travel to the central terminal only to be relocated to the external hubs for dispatch. As explained by Modig and Åhlström (2013), the law of bottlenecks states that throughput time increases as bottlenecks occur. When SPS have needed to take a stance on how to address this process bottlenecks, they explain that they prioritized short term resource efficiency.

> We had a challenging period at the end of Q4 2020, where we had too little room in our car-park outside the facility to load all the cars, and we had an opportunity to move into a new facility which would cost a lot, but we would have been able to meet demand better. But we said that this is not the right thing to do because it will only be 'right' for one quarter, and if we accept the offer it will be a 4 year commitment. - (Head of Logistics, SPS)

5.4.3. Variation Within SPS

The firm's demand is closely related to the market growth of e-commerce and its spikes which can cause both predictable and unpredictable variability. The biggest occurrence of predicted variability occurs during Q4 each year. The company copes with it by synchronizing their resource investments to match the expected seasonal spike.

> ...looking at the coming year, we know that the absolute peak will be in November/December. And starting in January it will drop quite a bit… But we also know that in April we will have achieved a similar demand as in previous year's Q4 [due to growth], so Q4 sets an important bar for the future… So, we make capacity increases in Q3 to match the peak in november/december (Q4), and then 'grow into it' by april. But we want to grow enough to be right on the line [of normal resource utilization] by the late spring. - (Head of Logistics, SPS)

This implies that they leverage expected seasonal demand growth in order to reduce the time at which they operate in under-capacity. Since capacity increments often need to be large to warrant an increase, this approach to adapting to variability attempts to optimize resource utilization. This data-driven decision implies that SPS can reduce the unpredictable variability of long-term demand and seasonality by becoming "smarter" the longer they operate in the market. When asked explicitly about the role that data has played in their long term ability to better predict the incoming orders, the Business Development Manager replied " absolutely [emphasis], it has made a change".

Unpredicted variability does, however, still occur in many different forms. One example caused by their retail partners was presented by the head of logistics:

> ...when a retailer puts up a massive promotion without telling us, we get a big deviation from the volume forecasts we have predicted... and that doesn't take us down every time but it's obvious that when we are working with tight schedules and we don't get any warning, then the risk for late deliveries grows. - (Head of Logistics, SPS)

Similar to the other firms, SPS engages in strategically self-inflicted, or intended, variability to improve their customer offer. One of the biggest tradeoffs between flow and resource efficiency is the flexible times at which end-customers can set their desired delivery. The standard offer is set based on order time, i.e. if ordered in the morning it will be delivered in the evening, and if ordered in the evening, delivered the next day. Consumers can however delay the delivery to several days in the future to better fit their personal schedules. This may sound like a simple request, but in practice this implies another split in production flows to accommodate. Not seldom, packages are sent out to the customer even if the customer re-books it, and it will instead count as a failed delivery. This is because the sorting and automated routing has already occurred when the customer changes their delivery time, which means that the process of taking a single package out and changing the route is more burdensome than simply delivering it twice.

> For example, offering every customer every day to change the delivery day. It's great for the customer offer, but terrible in a flow sense of the word. It means that we need to sort things several times, we can't sort it to disappear but have to reserve it for tomorrow. For some retailers we have niched offers (e.g. quicker deliveries, specific time deliveries) that absolutely don't fit well with the capacity we have, and the result is that we need to re-produce a certain product several times because it doesn't fit well with how the flow is set up. - (Head of Logistics, SPS)

Here it is worth mentioning that the Head of Logistics' reference to *flow* is distinct from the flow efficiency referred to in this report. As was described in the introductory phase of SPS' empirical presentation, the value added time occurs when the package is being processed to be delivered as soon as the customer wants it. Thus on this occasion, the package sitting idle in SPS' terminal is regarded as value-adding time for the customer who benefits from what is effectively temporary storage of their package.

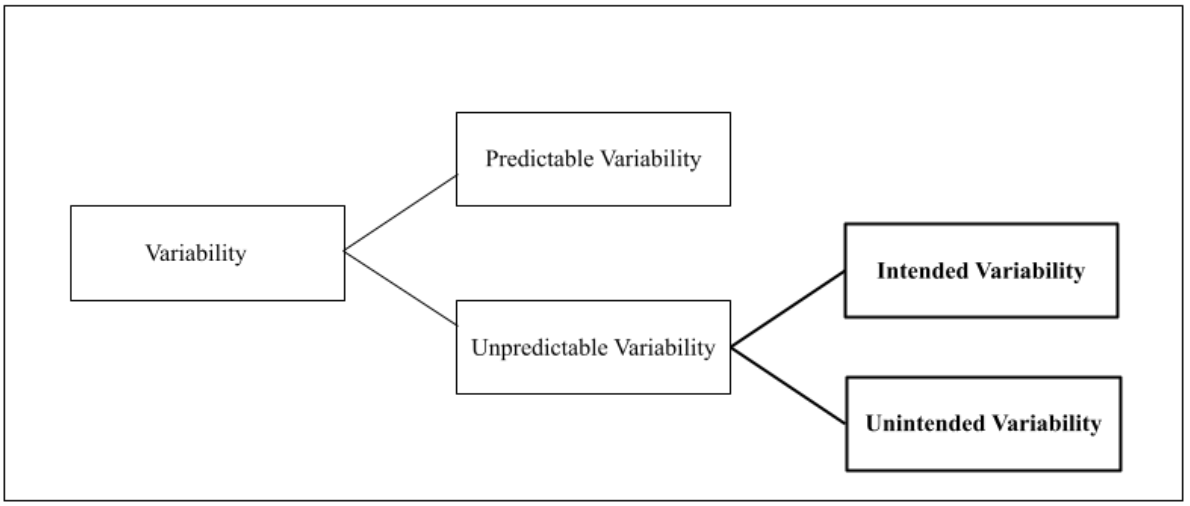
6. Analysis

In this section we will synthesize the empirical data by presenting a summary of our theoretical development: the two kinds of unpredictable variability. Next we will elaborate on each type, while providing some concrete examples from our case companies and motivate why and how variability can be a good thing. Finally, we will summarize our findings and give an answer to our research questions.

6.1. Our Finding: Two Kinds of Unpredictable Variability

Most literature regarding lean and flow efficiency makes compelling arguments for why variability should be eliminated from a process in order to achieve high flow efficiency. Hopp and Spearman (2020), however, theorize that behind every business decision, a clear cost-benefit analysis should pave the way and that this is where the variability vs. variety dilemma appears. In addition to this, they highlight that unpredictable variability, in the cases where it is essential to customer value, might be desirable from a business perspective. This should be contrasted to variability that does not add value to the end customer which should always be minimized. Combining the empirical findings in this paper with the theorization by Hopp and Spearman, we have categorized these phenomena as *intended* and *unintended* variability. The purpose is to concretize the distinction for practitioners, and provide a deeper analysis regarding the ways that hyper-growth companies prioritize the different efficiency measures to promote rapid growth, and ultimately, a long-term achievement of both.

Graphic 1: The tree of intended/unintended variability, Jadberg and Tavassoli (2021)



6.1.1. Intended Variability

We define intended variability as the kind of variability that an organization has actively chosen as a business decision to provide a more attractive offering towards their customers. This phenomenon is seen many times within our case companies, manager 6 at SPS provided a clear example of this:

> For example, offering every customer every day to change the delivery day. It's great for the customer offer, but terrible in a flow sense of the word. It means that we need to sort things several times, we can't sort it to disappear but have to reserve it for tomorrow. - (Head of Logistics, SPS)

In the case of Reviver it becomes evident that they also have accepted a lot of unpredictable variability in order to keep their customer offer as attractive as possible. For example they do not regulate what kind of items customers can send to them to sell. Many competitors have specialized in e.g. women and children's clothing to lower the amount of variation.

> It's a fairly unique logistical problem, where it's all about infinite amounts of variation, and how we can standardize all of this variation. We know roughly how much will come in, but we never know what will come in and how we can standardize all of this variation. - (Expansion Operations Manager, Reviver)

P-mob is also the case of a company that has a lot of intended variation in their operations. They could for example have been a lot more strict when it comes to when you can ride or where you can park. Instead they focus on the customer value and allow for all the variation that it brings to their operations.

6.1.2. Unintended Variability

This is the version of variability that organizations should strive to minimize. Unintended variability can be caused by e.g. machine failure or poor communication. Within all our case companies, processes include a lot of human labour which introduces natural unintended variation to a process since all humans are unique individuals. This makes it even harder to combat variation and puts higher pressure on standardized processes (Modig and Åhlström, 2013). P-mob experienced this with their local operation teams making individual and manual

judgement-based decisions about capacity and balancing which caused a lot of variation in performance in the different markets. Unintended variability is also apparent within Reviver for the same reason as in P-mob, that being the human element in the process. Since the majority of all tasks within the production process at Reviver is based on manual labour, individual skill and performance will have a major impact on the performance of the company as a whole. Lastly, at SPS we can find unpredictable variability of a different kind. As discussed in the empirical data section, their e-commerce partners initiate campaigns without communicating to SPS which sometimes increases the demand to the extent at which their capacity buffers are not enough. Even though this is predictable variability for E-commerce, it becomes unpredictable and unintended variability for SPS who may need to buffer in time which is the last kind of buffer they want to use according to their KPIs.

6.2. Variability as a Competitive Advantage

Looking at the business models and operations of all our case companies they include many complex challenges. Reviver is an e-commerce company with millions of unique items and every new item needs a unique advert. SPS is a logistical company that can offer 2-hour deliveries with flexible re-booking. And P-mob is a personal mobility company with vehicles you can travel with and park in almost any part of a city. It is clear that all our case companies have grown their business with customer centricity at the heart of their business model and value offers. Looking at their choice of establishing intended variability that inherently damages their operational efficiency, it appears that these companies have used intended variability as a competitive advantage.

All of our case companies operate in highly competitive markets that are rapidly growing where it is crucial to gain market share in order to stay competitive. This is congruent with the challenges outlined by Scott and Bruce 1987 and as part of our defined state of growth. The similar pattern we see in all our companies is that their unique selling point is their superior customer offer. As discussed by Hopp and Spearman (2020), there is always a tradeoff from a business standpoint when it comes to customer value and operational efficiency regarding the variability vs variety dilemma. It becomes clear that our case companies have chosen to prioritize variety and a great customer offer. This intended variability acts as one of the enablers for the staggering growth that these companies experience.

Despite all of this variation in the companies' processes, the flow efficiency has still been an important factor since it is closely connected to the customer experience. What we find when analyzing the data collected is that it is expensive for the companies to uphold a high level of flow efficiency when the intended variation is high. This is because it requires a lot of buffers in the form of capacity since they can not predict what or where the demand will come from. From our interviews it became apparent that these companies were all well funded by external investors which enabled them to maintain such a strong focus on customer experience. There was a period in the case of Reviver where they had to shift focus to being resource efficient in a "scramble for cash". Today however, Reviver has secured funds and once again focus has shifted towards flow efficiency. This implies that major external investments are necessary to be able to accept high levels of intended variability to achieve high levels of growth.

A second occurence of prioritizing resource efficiency was found in SPS who decided against extending their warehouse at a time of capacity shortage. The explanation was the overwhelmingly large tradeoff it would have implied, renting a whole terminal that would only be used for a few months when the lease commitment was 4 years long. This example brings some nuance into the analysis, but is regarded as less significant to the findings of the study. The conclusion from this exception is that long-term strategic fit supersedes very short-term flow efficiency. Conversely, the flow efficiency would have been reduced due to "human error" for the startup period of the new facility, meaning that the tradeoff was never really that relevant. Today, SPS has secured a long term lease that will fit them for what they estimate is 10 years, meaning that the bottle-neck challenges mentioned in the empirics section will be reduced significantly, thus increasing flow efficiency for the foreseeable future.

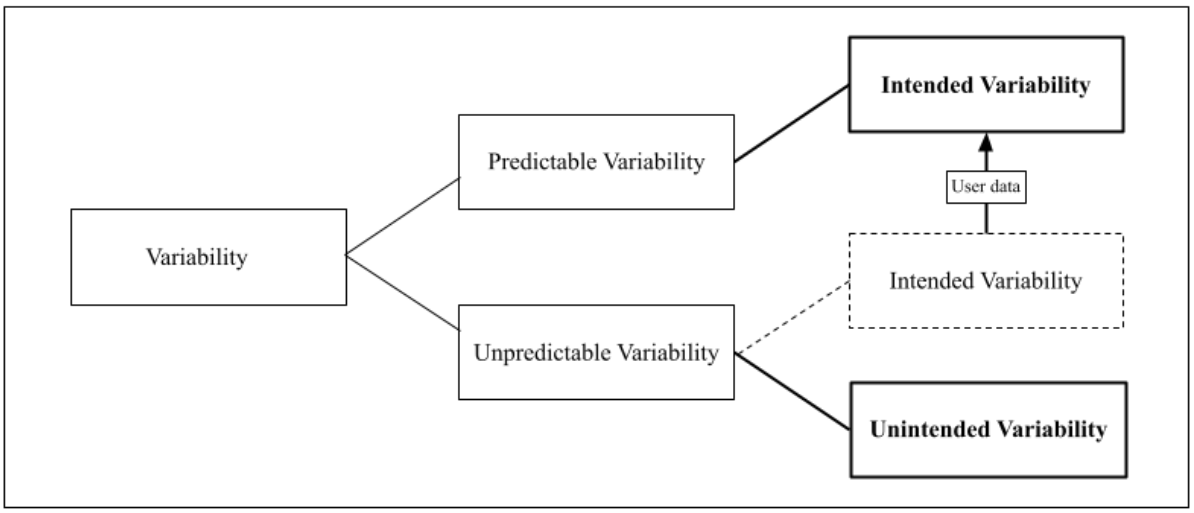
6.3. Data: The Key to Predicting Variability

As the case companies studied in this paper have scaled fast and quickly gained a large customer base, they have in the process gathered a lot of user data. What we see in common between the cases is their ability to leverage this data at an advanced level, which has unlocked major potential in improving the flow and resource efficiency within the companies despite large volumes of intended variability.

As mentioned previously, the case companies have difficulty increasing profitability because of their operations being costly due to the high levels of intended variability built into the customer offering. The empirical data shows that because of user data all the case companies are on the route of turning the previously unpredictable intended variability into predictable variability. As unpredictable variability becomes predictable, an organization enables a shift in buffering decisions. Instead of continuously having to keep costly buffers to ensure a good flow efficiency, they can buffer only where and when it is needed.

Our findings suggest that this is the main reason why the case companies choose and are able to operate with such high levels of variability initially. Since the customer experience is crucial to gain market share and achieve hyper-growth, it is a prerequisite to succeed in these highly competitive landscapes. However, thanks to predictive analysis of user data the operational efficiency will increase drastically as they are able to predict the intended variability built into their business model.

Graphic 2: The post-data tree of variability, Jadberg and Tavassoli (2021)



6.4. Summary of Conclusions

To answer the research questions in a clear way we will summarize our conclusions from this paper.

Q1: How do hyper-growth companies prioritize flow and resource efficiency with rapidly increasing customer demand?

It is clear that the case companies studied view customer centricity as an important factor to enable growth. The findings in this paper suggest that the priorities between flow and resource efficiency may shift depending on the financial state of the company. As long as the companies are well funded, the customer value and thus flow efficiency is the priority.

Q2: How are they using variability, buffering and flexibility to achieve this?

Variability plays a big role in all of our case companies' operations. That is because of the intended variability they bring on themselves in order to create a better customer offer that enables them to grow rapidly. They meet this variation by buffering capacity which inherently lowers their resource efficiency. This finding further strengthens the claim that flow efficiency is their priority as mentioned in Q1, and broadens the view of existing literature on how variability can be subject to a cost-benefit analysis in a business decision. Within the companies studied in this paper, intended variability is a key part of their core strategy.

Q3:How are they able to prioritize the way that they do?

The findings in this paper suggest that the reason the companies make these priorities is connected to predictive analysis of user data. The rapid growth and thus rapid collection of user data is the reason why they can inflict large amounts of intended variability on their operations. With the predictive analysis of the user data, they are able to turn the unpredictable intended variability into predictive variability (see graphic 2). This enables them to turn a rapidly growing business with high flow efficiency but low resource efficiency into a potentially large business with both high flow and resource efficiency, or what is referred to as the perfect state. It is yet to be seen if it will be realized, but the findings in this paper suggests that they are well underway and that the strategy is working. An implication on this can be found in the case of P-mob where a shift towards increasing resource efficiency and moving towards the perfect state has already been initiated.

7. Discussion

The claim that customer centricity is critical for rapid demand growth in a competitive market can be regarded as trivial. Previous literature has established that flow efficiency is a central aspect of achieving alignment with customer needs. Thus, the actual findings in this paper can be critically viewed as more of a contextualization of existing literature. The case companies have proven more homogenous than previously expected, which unintentionally strengthened the replication logic of the case selection. Their similarities are being in a state of hyper growth and having received external investments to achieve said growth. The differences however are many, including: industry, business model, revenue model, employee count and so on. Thus, the broad conclusions can be generalized beyond our specific case companies given that the two conditions of hyper-growth state and external investment are fulfilled.

7.1. Contributions

Comparing the results in this study with previous research, we find coherence with findings by Hopp & Spearman. These findings give a more nuanced account of variation compared to early lean literature, which in our conception is still the dominant perspective of how practitioners and academics interpret lean. As mentioned in the theory section there is a wide spread of literature on how variation in processes is nothing but subject to elimination in order to achieve operational efficiency. This paper presents empirical observations supporting the point that variability can act as a direct competitive advantage for hyper-growth companies which stands in contradiction to earlier research on lean but builds on the more modern research by Hopp and Spearman (2020).

This paper makes two main contributions to the previous literature. Firstly, the paper shows how priorities between flow and resource efficiency is done within the previously unexplored context of hyper-growth. Secondly, it brings empirical evidence and extensions to the previously explored variability vs variety dilemma, presenting variability as a competitive advantage. This nuance may not have been observed if it were not for the case study method and the selection of cases being limited to hyper-growth.

7.2. Limitations

A valid criticism of the findings in this report is that we could have done a better job at finding the actual values of flow efficiency at the three firms, or at least getting a representative figure for what these may be. These could have been used to validate the management intentions described in the interviews, or benchmarked against the flow efficiency values of other non-hyper-growth firms in equivalent industries to confirm whether the focal firms of this study are relatively more flow efficient.

By using several interview objects describing the same case companies, we have attempted to exclude the possibility for blatant misrepresentations of the truth. Each interview object was also informed that their companies would be anonymized in our reports, meaning that there would be no commercial benefit to misrepresenting facts other than to impress the interviewers. However, since the empirical data that underpins the conclusions was collected from semi-structured interviews, the conclusions can and should be treated with some skepticism. If triangulation had been properly conducted, the reliability of the results would have increased.

7.3. Future research

For future research, this paper raises the question of whether or not hyper-growth firms reach the perfect state through the strategy of transforming intended unpredictable variability into predictable variability. Thus, it would be interesting to follow up this study with a similar methodological approach focusing on matured previous hyper-growth companies. Such a study could focus on proving or disproving the implications seen in the summary of conclusions, that hyper-growth companies are able to maintain a high flow efficiency whilst improving their resource efficiency and thereby their profitability over time. Another interesting research topic would be to investigate and compare the flow-efficiency of these relatively young companies to their more mature equivalents, and see whether the customer offer has different limitations based on the growth of the firm.

8. References

Ahlstrom, P. 1998, "Sequences in the implementation of lean production", *European management journal,* vol. 16, no. 3, pp. 327-334.

Allabolag.se. 2021. allabolag.se - Företagsinformation om alla Sveriges bolag. [online] Available at: <https://www.allabolag.se/> [Accessed 15 may 2021].

Blank, S. 2013, "Why the lean start-up changes everything", *Harvard business review,* vol. 91, no. 5.

Bluhm, D.J., Harman, W., Lee, T.W. & Mitchell, T.R. 2011, "Qualitative Research in Management: A Decade of Progress", *Journal of management studies,* vol. 48, no. 8, pp. 1866-1891.

Bowen, D.E. & Youngdahl, W.E. 1998, ""Lean" service: In defense of a production-line approach", *International Journal of Service Industry Management,* vol. 9, no. 3, pp. 207-225.

Bryman, A. & Bell, E. 2013, *Företagsekonomiska forskningsmetoder,* 2nd edn, Liber, Stockholm.

Cassia, L. & Minola, T. 2012, "Hyper-growth of SMEs: Toward a reconciliation of entrepreneurial orientation and strategic resources", *International journal of entrepreneurial behaviour & research,* vol. 18, no. 2, pp. 179-197.

Danese, P., Manfè, V. & Romano, P. 2018, "A Systematic Literature Review on Recent Lean Research: State-of-the-art and Future Directions", *International Journal of Management Reviews,* vol. 20, no. 2, pp. 579-605.

Delmar, F., Davidsson, P. & Gartner, W.B. 2003, "Arriving at the high-growth firm", *Journal of Business Venturing*, vol. 18, no. 2, pp. 189-216.

Dobbs, M. & Hamilton, R.T. 2007, "Small business growth: recent evidence and new directions", *International journal of entrepreneurial behaviour & research,* vol. 13, no. 5, pp. 296-322.

Flamholtz, E.G., Randle, Y. & Mozilo, A.R. 2007, *Growing Pains: Transitioning from an Entrepreneurship to a Professionally Managed Firm,* John Wiley & Sons, Incorporated, Somerset.

Greiner, L.E. 1998, "Evolution and revolution as organizations grow. 1972", *Harvard business review; Harv Bus Rev,* vol. 76, no. 3, pp. 55-68.

Hopp, W.J. & Spearman, M.S. 2020, "The lenses of lean: Visioning the science and practice of efficiency", *Journal of Operations Management,* .

Hopp, W.J. & Spearman, M.L. 2011, *Factory physics,* Waveland Press.

Kollberg, B., Dahlgaard, J.J. & Brehmer, P.-. 2006, "Measuring lean initiatives in health care services: Issues and findings", *International Journal of Productivity and Performance Management,* vol. 56, no. 1, pp. 7-24.

Levie, J. & Lichtenstein, B.B. 2010, "A Terminal Assessment of Stages Theory: Introducing a Dynamic States Approach to Entrepreneurship", *Entrepreneurship theory and practice,* vol. 34, no. 2, pp. 317-350.

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

Lippitt, G.L. & Schmidt, W.H. 1967, "Crises in a developing organization", *Harvard business review,* vol. 45, no. 6, pp. 102.

Loubser, J.J. 1968, "The Discovery of Grounded Theory: Strategies for Qualitative Research.Barney G. Glaser , Anselm L. Strauss", *The American journal of sociology,* vol. 73, no. 6, pp. 773-774.

Modig, N. & Åhlström, P. 2012, *Detta är lean : lösningen på effektivitetsparadoxen,* 2nd edn, SSE Institute for Research, Stockholm.

Scott, M. & Bruce, R. 1987, "Five stages of growth in small business", *Long range planning,* vol. 20, no. 3, pp. 45-52.

Spear, S. & Bowen, H.K. 1999, "Decoding the DNA of the Toyota Production System", *Harvard business review,* vol. 77, no. 5, pp. 96.

Suddaby, R. 2006, "From the editors: What grounded theory is not", *Academy of Management journal,* vol. 49, no. 4, pp. 633-642.

Voss, C., Tsikriktsis, N. & Frohlich, M. 2002, "Case research in operations management", *International journal of operations & production management,* vol. 22, no. 2, pp. 195-219.

Womack, J.P. & Jones, D.T. 1996, *Lean thinking : banish waste and create wealth in your corporation,* Simon & Schuster : Free Press, New York.

Womack, J.P., Jones, D.T. & Roos, D. 2007, *The machine that changed the world : [how lean production revolutionized the global car wars],* Simon & Schuster, London.

Yin, R.K. 2014, *Case study research : design and methods,* 5th edn, SAGE, London.

9. Appendix

Interview Guide

How long have you worked at __?
What is your formal role?
What level of management are you (how many layers?)

Scope
Understanding the priorities within the organization with regard to flow and resource efficiency, as well as to what extent buffers are considered in the process

Questions
Understanding the interview object
- What do you do at ___ ?
  - Can you describe your function (team)

The objectives, flow and variation
- Can you describe the flow within operations in order to get your product to your customers?
- What common objectives do teams work towards (that are critical for ____)
  - Walk us through some performance metrics that you work with on the day-to-day

- What are the problems you face in your normal operations?
  - How does demand variation affect your business ?
    - - and performance metrics?
    - Is your demand: stable/unstable growth/seasonal
- How do you cope with demand variation (Buffers: Time/Capacity/Inventory/Flexibility)

How they build a better business
- What changes have____ made within operations since its inception?
  - What did you want to achieve with these changes?
  - How did you do it?
  - What results did it have?

Later addition 1
- What considerations do you make when entering a new market?
  - When entering a new market, what priorities do you set? (with regard to resource/flow efficiency)

Later addition 2
- What sacrifices have you made in flow efficiency (customer centricity) in order to achieve greater resource utilization?
- What sacrifices have you made to resource efficiency in order to be more customer centric (or flow-efficient)