TÉCNICO LISBOA

Product Development Risk: A Bayesian Network Approach

The Unbabel Case Study

Gil Afonso Correia Coelho

Dissertation to obtain the Master of Science Degree in

Industrial Engineering and Management

Supervisor: Prof. António Manuel da Nave Quintino

Examination Committee

Chairperson: Prof. Maria Margarida Martelo Catalão Lopes de Oliveira Pires Pina
Supervisor: Prof. António Manuel da Nave Quintino
Members of the Committee:
Prof. João Carlos Da Cruz Lourenço

July 2022

Acknowledgments

I would like to start by thanking my family. It is because of their tremendous support through my entire life that I have had the ambition to try and go further even when it is not easy. The reason I believe anything is possible in my personal, academic and professional life is because they have believed in me. For this I need to thank my parents, who are an example of determination and courage and whose effort has, more than opened doors for me, taught me how to open them myself. Any success I might one day have is as much theirs as is mine. The same goes for my brother who has, since I was a small child, been an example of intelligence, achievement but also kindness - characteristics which I am also happy to see in my very supportive sister-in-law.

I would like to also dedicate this thesis to my grandparents who created the opportunities so that the next generations could build better lives. A warm thank you, in particular, to my paternal grandparents posthumously who I know are incredibly proud and whose support got me to where I am today.

A kind thank you also to my aunt, who has always been supportive of my endeavours and I know is as happy as I am whenever I succeed.

I must acknowledge my dissertation supervisor Prof. António Quintino. Conducting a project of this dimension while working full-time is a hard challenge that requires many after work hours and weekends. This effort was made much easier and more enjoyable by the way the Professor shared my enthusiasm with the topic and challenged me to go further. I also need to thank for the flexibility that allowed me to conciliate work and academic life - I always felt supported and understood as I managed my responsibilities.

I want to profoundly thank Paulo Dimas who is much more than a manager, but a friend and a mentor. From the very start Paulo has had a deep focus in setting me up for success (professional, academic and personally), has believed in my capabilities, and given me opportunities to grow. Most of what I have learned about product and that is the foundation for this dissertation comes from learning directly with him. I have little doubt no other manager would provide me as much support and help me launch my professional career. I also have no doubt very few managers would have such a deep concern for my personal well-being and provided me the flexibility to conciliate work and studies that I needed. Today I am delivering a dissertation while working at a job I love because of this.

I need to thank Unbabel in general, for being an incredibly innovative company that challenges me, and Lawrence, Mariana, Phill and Adi in particular for the time they spent with me. They have enriched this dissertation with their knowledge and I own them quite a few coffees (or beers!).

I also need to acknowledge JUNITEC - which was my ”home” in Instituto Superior Técnico and that taught me much more than I could have ever imagined. It really is true that you ”get the double in return, of whatever you put in”. Many people have shaped my path there - but in particular João Santos and Sara Farias who were partners and friends.

Finally, a thank you to my friends. Our shared moments and laughs have made the past years a joy even after countless hours of study and work. I look forward to the next 50 years.

“Whether you think you can, or you think you can’t – you are right” - Henry Ford

Thanks to these people, I believe I can.

Abstract

Highly innovative technology startups and scale-ups rely on launching disruptive products. Despite facing very uncertain scenarios these companies rarely adopt “formal” and “sophisticated” risk analysis tools. There are challenges related to engineering, user experience, and business models that make it so that only a small percentage succeeds in the market [Eisenmann, 2021]. The high risk, high return nature of these ventures makes it so that any improvement in the low success rate can bring great benefits, namely to investors and to the economy.

While there is a growing number of books and articles on how to launch technological products [Cagan, 2008] [Torres, 2021] - there are few examples in the literature aligned with these new ideas. When it comes to product development risk there is a clear gap between what is being adopted by the industry and what the academia has already studied.

Bayesian Networks are a powerful technique to create visual probabilistic models which can be used for multiple applications.

The employed methodology uses Unbabel, a Lisbon-based technology scale-up, as a case stud yon how modern technology companies think about product development risk. Through expert interview 4 different bayesian networks are generated to model and predict Feasibility, Usability, Value and Viability Risk. Findings suggest that the method creates models that behave consistently under different scenarios and that are suitable for many business applications such as decision making and risk analysis. Furthermore, it is concluded that bayesian networks can formalize industry practices and concepts, bringing academia and business closer together

Keywords

Product Development Risk; Bayesian Network; Product Management; Scale-ups; Innovation

Resumo

O sucesso das *startups* e *scale-ups* tecnológicas depende do lançamento de produtos disruptivos. Apesar de estarem expostas a cenários altamente incertos estas raramente adotam métodos "formais" e "sofisticados" de análise de risco. Desafios relacionados com engenharia, modelo de negócio e experiência do utilizador causam a reduzida taxa de sucesso que existe [Eisenmann, 2021]. A natureza de "alto-risco, alto-retorno" destas iniciativas significa que qualquer melhoramento na baixa taxa de sucesso pode trazer enormes benefícios, nomeadamente a investidores e à economia.

Apesar de existir um número crescente de livros e artigos [Cagan, 2008] [Torres, 2021] em como lançar produtos tecnológicos - existem poucos exemplos na literatura alinhados com estas novas ideias. Na área de desenvolvimento de produto, existe um claro contraste entre aquilo que é adotado pela indústria e aquilo que é estudado.

Redes bayesianas são um método eficaz para criar modelos probabilisticos altamente visuais que podem ser utilizados em múltiplas aplicações.

O método de investigação utiliza a Unbabel, uma *scale-up* de tecnologia baseada em Lisboa, como caso de estudo para perceber como é que as empresas modernas de tecnologia pensam em risco de desenvolvimento de produto. Através de entrevistas a 4 especialistas diferentes redes bayesianas são criadas para modelar e prever o risco de *Feasibility, Usability, Value e Viability*. Os resultados sugerem que o método cria modelos com comportamento coerente e aplicáveis em diferentes contextos de negócio como análise de risco e tomada de decisões. Por último, conclui-se que as redes bayesianas são eficazes a formalizar conceitos da indústria, aproximando a academia e o mundo dos negócios.

Palavras Chave

Risco em desenvolvimento de produto; Redes bayesianas; Gestão de Produto; Scale-ups, Inovação

Contents

List of Figures

Acronyms

AHT	Average Handling Time
AI	Artificial Intelligence
B2B	Business to Business
B2C	Business to Customer
BN	Bayesian Network
BNs	Bayesian Networks
BPO	Business Process Outsourcing
CEO	Chief Executive Officer
COSO	Committee of Sponsoring Organizations of the Treadway Commission
CPT	Conditional Probability Table
CPTs	Conditional Probability Tables
CS	Customer Support
ERM	Enterprise Risk Management
FAQs	Frequently Asked Questions
HT	Human Translation
IPO	Initial Public Offering
ISO	International Organization of Standardization
KPI	Key Point Indicator(s)
MT	Machine Translation
NPV	Net Present Value
NPVR	Net Present Value Risk Adjusted
OKRs	Objectives and Key Results

PD	Product Design or Product Designer
PESTLE	Political, Economical, Social, Technological, Legal and Environmental
PLG	Product-led Growth
PM	Product Management or Product Manager
PMM	Product Marketing Manager
RD	Research and Development
RM	Risk Management
ROI	Return on Investment
SaaS	Software as a Service
SLAs	Service Level Agreements
SME	Small and Medium Size Enterprise
SMEs	Small and Medium Size Enterprises
SWOT	Strengths Weaknesses Opportunities Threats
TAT	Turnaround Time
UX	User Experience
VC	Venture Capital
VP	Vice President

1

Introduction

1.1 Background

Highly innovative companies, particularly startups (or scale-ups), are exposed to a high degree of uncertainty by nature. This high-risk, high-reward model has been very successfully in launching a few very large companies that provide a significant Return on Investment (ROI) to its investors but the majority of startups fail as they are exposed to a high number of risks such as having the wrong business model, being surpassed by competitors or running out of cash [Cantamessa et al., 2018].

Different definitions have been used to classify what a scale-up is, such as “companies that grow to more than $10 million by their 5th year of revenue” [van Dijk et al., 2015]. We will use this term more broadly to refer companies that are in a “growth-phase” in the Miller company life-cycle [Miller and Friesen, 1984], being more mature than a start-up but without having achieved the same scale as an enterprise.

Despite being particularly exposed to uncertainty, there is not much literature focused on risk in the particular context of scale-ups. [Villa Todeschini et al., 2017]. These companies are suddenly faced with the need to become a very different kind of organisation - in terms of structure, process and discipline - as

described by Piken in 2017 [Picken, 2017]. This environment of uncertainty could potentially benefit from enterprise-wide risk management strategies or from risk management practices focused in key areas (such as product development), which are usually studied and implemented in large firms. Unbabel is an example of such a company.

Unbabel is a translation platform founded by Vasco Pedro, João Graça, Sofia Pessanha, Bruno Silva, and Hugo Silva in 2013 and incubated by Y Combinator in late 2014.

Unbabel enables modern enterprises to serve customers in their native languages, with always-on, scalable translation across digital channels. Powered by Artificial Intelligence (AI) and refined by a global community of translators, Unbabel combines the speed and scale of machine translation with the authenticity that can come only from a native speaker.

By October 2020 it had raised a total of 92.1M$ in funding across three rounds, hiring around 200 employees spread out across offices in Lisbon, San Francisco and Pittsburgh. It is recognized as one of the fastest growing and most promising companies with Portuguese founders, considered in 2019 to be the "best scale-up" in Portugal. [BGI and EIT Digital, 2019]

1.2 Risk in the context of Unbabel

Unbabel is a cutting edge technology firm which leverages state of the art Machine Translation (MT) combined with "Human in the Loop" validation. It is exposed to a number of technology risks, combined with the typical risk factors found in a startup such as lack of funding or high employee turnover. Given the high uncertainty environment it is a particularly interesting use case to study for risk and risk management strategies. Gjerdum [Gjerdum, 2015] stated in 2015 that any successful risk management initiative "should be proportionate to the level of risk in the organisation (as related to the size, nature and complexity of the organisation)" .

Scale-ups require agility to iterate fast and find the best paths for growth but this approach comes at the cost of a lack of structured process that is found at larger companies. We propose that this positions scale-ups to particularly benefit from a Risk Management (RM) methodologies.

1.3 Purpose and Goals

The main goal of this research is to explore in general how risk management practices could be better adapted for implementation at Unbabel. In particular it will focus on a key area: product development.

It is expected that at the end there is clear deliverable that is useful to understand how Unbabel "thinks" about risk and that can potentially be used for risk management applications. A successful approach will combine methodologies from the literature with modern concepts and language from the

“scale-up” world to achieve a solution that bridges the academic and business world with regards to new product development risk - while using Unbabel as a case study.

1.4 Structure

In the Problem Statement chapter there will be a description of the research opportunity and an introduction to the company that is going to be studied. The following chapter, Literature Review, will provide a review of the key topics that are covered and provide the foundations for the research methodology. Chapter 4, Work Methodology, describes the different steps that were taken together with the subject matter experts in order to generate bayesian networks. In the Network Generation chapter each model is described in depth, surfacing the definitions of each node and important findings. The final chapter, Discussion and Conclusion, contains a description of the analysis and findings done on each of the models as well as applications and research opportunities for the work developed through this thesis.

2

Problem Statement and Introduction to Unbabel

An initial definition of the research question is crucial to provide a research focus and help make decisions regarding the research methodology, described in chapter 4. As outlined by Eisenhardt [Eisenhardt, 1989], this is particularly relevant when the Case Study is used as the Research Method - because it will aid the choice of the organisation(s) and of the data collection techniques.

2.1 Problem Statement

While there is a well established large body of literature around the topic of Risk Management, some of which is focused on new product development 3, there is a clear disconnection between methods accepted by the academia and the day-to-day of scale-up companies. Literature as been produced and widely adopted across the technology Industry as good product risk management practices, such as Cagan [Cagan, 2008] and Torres [Torres, 2021], but received little academic focus. Given the recency

of the publications and the quickly changing environment - there is an opportunity to model and study modern practices with a research focus.

From the available RM techniques - bayesian networks are a good candidate to formalize and model modern industry practices. 3 The question the author proposes is to understand if it is possible to model new product development risk decisions in the technology industry using a bayesian network approach.

Unbabel has been growing at an exponential rate since its foundation in 2013 having raised more than $90M by 2019 but it still needs to sustain a similar level of growth to achieve an Initial Public Offering (IPO) - possibly the best possible outcome for the company given the expected ROI for all stakeholders. Although it is recognized as a successful case study in transitioning from start-up to scale-up [BGI and EIT Digital, 2019], it has been highly impacted by uncertainty (firing more than 30% of the workforce for COVID-19 related reasons [Diogo Ferreira Nunes, 2020]) and is particularly exposed to technological and competitive risks. With a global workforce of around 150 employees at the time of the research, including senior executives from the technology industry, Unbabel uses modern product management practices and is a great example on how this type companies work.

2.2 Critical Risks in Product

Similarly to other startups and scale-ups, Unbabel has a product management function whose job is to "combine technology and design to solve real customer problems in a way that meets the needs of the business". This function, originally found within the marketing organizational structure, has developed in the past few decades to integrate the Research and Development departments of technology companies.

While in the past companies had very long release cycles, where it could take months or even years to go from idea until product, software has enabled companies to continuously develop and ship product. These faster cycles allow for an approach that is based on smaller releases and increments to the product, as the product managers (and product team) quickly experiment with new ideas. This means that the teams are continuously having to make decisions on where to invest limited resources (usually developer hours) so that they get the highest return on investment (customer value aligned with the needs of the business).

In this context the concept of "Product Discovery" [1] becomes very important: decision makers are always trying to "ship the right product" which means minimizing 4 key product risks:

- Feasibility - the risk of not being able to build the product/feature
- Usability Risk - the risk of the user not understanding how to use it

[1] Product organizations commonly use the term Discovery (figuring out what is the "right product" to build) to distinguish from Delivery (building the product).

- Value Risk - the risk of not providing enough user value so that he chooses to buy/use the solution
- Viability Risk - the risk of not generating revenue sufficiently higher than the costs

These 4 concepts were formalized by Marty Cagan in his popular publication “Inspired” [Cagan, 2008], which is one of the main frameworks adopted at Unbabel by the product management function.

2.3 Introduction to Unbabel

Unbabel combines state of the art AI with a global community of translators to provide a “Translation as a service” solution to modern enterprises. It combines the advantages of MT (such as speed and cost) with the expertise of native level speakers to provide a high quality translation service up to the standards of modern brands. It uses a Software as a Service (SaaS) business model - where customers acquire subscriptions that allow them to translate a given quantity of text for a specific use case and taking into consideration the specific linguistic requirements for the brand.

It is focused in the Customer Support (CS) industry, providing translation to text based content such as emails, live chat and Frequently Asked Questions (FAQs) for global brands such as Microsoft, Logitech and Tinder [Unbabel,]. The main verticals in which Unbabel customers fall into are Technology, Gaming, Retail and Travel industries. It has also partnered up with several Business Process Outsourcing (BPO) companies such as Concentrix to provide a joint service of CS at a global scale - where the BPO provides the workforce (agents) in a given location and Unbabel translates the content to all target languages.

While the current focus is on Customer Support, Unbabel has a broader vision to “Create Universal Understanding” and aims to execute on the mission of becoming “The Worlds Translation Layer”.

2.4 Unbabel Value Proposition

By providing an “always-on” solution that scales across several digital channels and up to billions of translated words per year, Unbabel allows enterprises to centralize CS agents into specific strategic locations while serving many more geographies. This not only allows for better knowledge management - as a single agent can be trained once and support several different markets - but also for cost reduction given that agents can be hired where they are less expensive.

Unbabel uses an hybrid approach that distinguishes it from existing options to translate content:

- Machine Translation - which has a marginal cost and is almost instant. Unbabel uses custom MT engines built using historical data from the customer combined with tools such as glossaries and Translation Memories to achieve higher quality for the particular requirements of each customer.

This combination of techniques is how Unbabel is able to provide a near-instant translation while avoiding errors that a "generic" engine would make - such as translating product names or missing industry terms. Another dimension is the formality level - each brand communicates differently with the customers depending on industry and culture (a customer from Japan expects a higher degree of formality than a customer in the USA.) Thus the need for the customization of engines - through the process called Domain Adaption.

While in recent years we have seen an increasing performance to "near-human" levels, Machine Translation does not solve all the translation use cases [Lavie, 2020]. MT is still insufficient when the quality standards are high and/or when less training data is available (such as less popular languages).

- Human Translation - which has the highest level of quality but is too slow and expensive for at scale, near real-time solutions for the CS use case. Waiting time is known to be a customer satisfaction determinant [Dimensional Research, 2013] [Bielen and Demoulin, 2007] and if we consider that the total perceived waiting time for the customer will be the Average Handling Time (AHT) of the agents summed up with the Turnaround Time (TAT) for the translation having a Human Translation (HT) step can lead to poor customer experiences. Including Humans in the translation won't be viable for any chat solutions, where near-instant answers are expected and will only be viable for asynchronous communication (such as Emails) when the impact on total waiting time is low.

This hybrid approach results in different combinations of HT and domain-adapted MT depending on the speed, cost and quality requirements of each use case - Tickets (emails), Chat or FAQs.

2.5 Unbabel Product Portfolio

Given the existing trade-off between quality of translation and speed - for an email we are willing to wait for a few hours but expect perfectly translated content while for real-time chat we expect an answer within seconds but can accept a translation that is not considered to be "native-level" - Unbabel offers several solutions which are separated as "content-types". All solutions leverage domain adapted MT - the customer provides Unbabel with a large number of examples of their customer support requests which are then translated by the community. These native level translations, specific to the domain of the customer, are used to train MT engines specific to the industry or to the customer.

Depending on the use case messages are either machine translated and directly sent to the customer (Chat) or machine translated and then improved by humans (Tickets) or evaluated by several humans before sending to the customer (FAQs).

2.5.1 FAQs

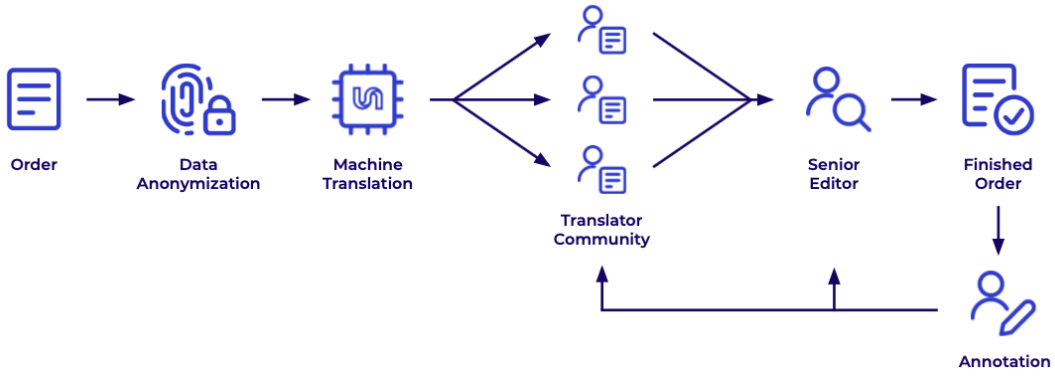

Figure 2.1: Unbabel FAQs Pipeline
Source: Unbabel

FAQs are the only long format content-type supported by Unbabel. As such, they are broken into several smaller sections that are then sent to the community for translation. A Senior Editor (editor recognized for past high quality work) is then responsible for grouping the different sections and ensuring fluency and consistency. Only the highest level of translation quality (native) is acceptable since this is a static content on the customer website.

2.5.2 Tickets

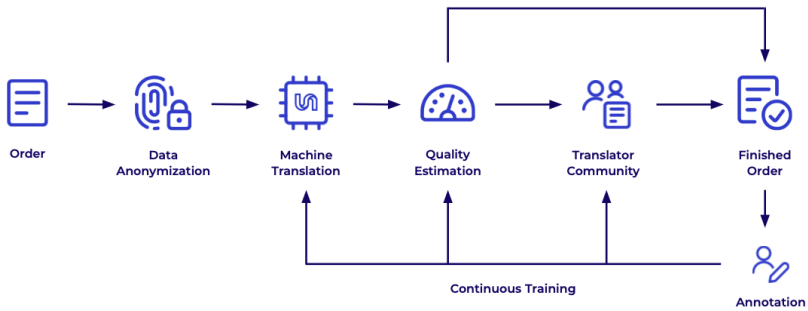

Figure 2.2: Unbabel Tickets Pipeline
Source: Unbabel

Tickets correspond to email exchanges between customers and agents. This is an asynchronous method of communication where Service Level Agreements (SLAs) usually allow from a few hours between the reception of the email and the reply. Customers expect the correct tone, fluency and grammar and as such humans revisions are applied to almost all instances after the machine translation. It should also be noted in the figure 2.2 that there is a “re-training step” which continuously “feeds” the MT engines

with new “native level” and domain adapted data to improve their performance over time.

2.5.3 Chat

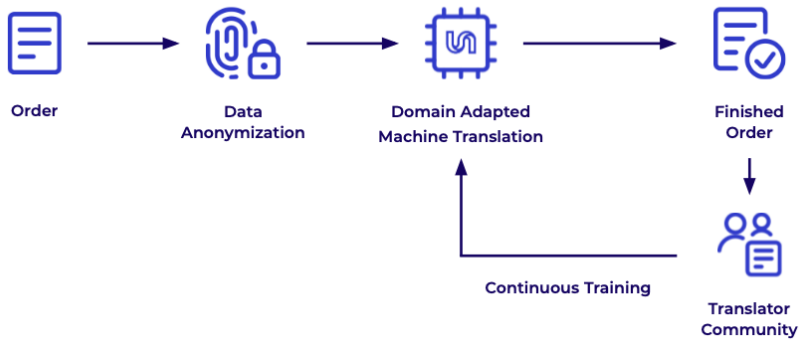

Figure 2.3: Unbabel Chat Pipeline
Source: Unbabel

Chat corresponds to real-time conversations between customers and agents. In this synchronous method of communication customers expect replies in a matter of seconds and therefore accept a certain degree of errors in fluency. This product will almost always only use MT to be able to have a TAT below a few seconds. After a given amount of data is accumulated it will be sent to the community for human annotation and revision. These improved (“native level”) translations are then used to re-train the MT engines to ensure improvement over time.

3

Literature Review

This chapter 3 presents the key concepts, definitions and state of the art that will be crucial in following chapters of the work. The goal is to approach the relevant topics under the light of the most recent literature to have a clear understanding of the state of the art and outline the knowledge gaps to which this research can contribute.

3.1 Approach and Outline

Any research needs to start with a clearly defined topic in mind - “Risk Management in the context of a cutting edge technology scale-up”.
It must be noted that research and literature review started with this broad topic but then converged into Product Development Risk as an opportunity to make a contribution was found. This is why the review starts with a broad assessment of Risk Management in this particular context.

From this topic we can define a few keywords to be used in researching literature: Risk Management; Enterprise Risk Management; Scale-up; Risk in Technology Given the very specific nature of this research problem and the lack of literature focused particularly in scale-ups, “adjacent” fields were

also explored such as risk in the context of start-ups, Small and Medium Size Enterprises (SMEs) and Research and Development (RD) projects, providing additional keywords to broaden the search.

Online tools such as Google Scholar (https://scholar.google.com), Elsevier (http://www.elsevier.com) and b-on (https://www.b-on.pt) were used to search for the keywords outlined above and papers were filtered according to their relevance to the topic, applicability and recency. To broaden the research, cited sources in the found papers where also scanned for further information, in a "cascading effect" to find more relevant information.

During this screening process papers focused on very project specific or industry specific topics were removed, particularly in the manufacturing and health sectors, and abstracts were reviewed to evaluate relevancy.

3.2 Risk Management key concepts

The International Organization of Standardization (ISO) [BSI Standards Publication, 2018] and Committee of Sponsoring Organizations of the Treadway Commission (COSO) are the most widely accepted internal standards for risk management according to several authors. [Nadali et al., 2018] [Crovini et al., 2020] These publications provide definitions and guidelines that are used across literature and frame the different steps that compose a complete RM project.

According to the ISO 31000 norm, risk is "the effect of uncertainty on objectives" [BSI Standards Publication, 2018], which encompasses both positive and negative outcomes related to uncertainty. We will follow this definition with the aim of also covering opportunities (which are the positive consequences of managing risk well in a context on uncertainty) and not just studying the events that result in a negative impact for the organisation - the Strengths Weaknesses Opportunities Threats (SWOT) Analysis technique is an example of such an approach.

Risk Management is the combination of a number of activities carried out with the goal of managing an organization with regards to risk [BSI Standards Publication, 2018]. It can serve several purposes, across all levels of a company (strategic, tactic, operational) and be used to enhance the decision making process (as a driver of strategic decisions), assure compliance and improve efficiency [Institute of Risk Management, 2010]. According to COSO a well defined Enterprise Risk Management (ERM) is also a useful tool for the management (leadership) team to be aware and make decisions knowing the risks that can have an impact on the strategy. [PwC, 2017].

The Institute of Risk Management published in 2010 [Institute of Risk Management, 2010] a guide developing and providing recommendations on the implementation of a ISO 31000 initiative. It outlines how risk management responsibilities should be spread across the organisation, from the board to the individual contributor and recommends several techniques to perform risk assessment such as:

questionnaires, audits, SWOT and Political, Economical, Social, Technological, Legal and Environmental (PESTLE) analysis. All of the above are aligned with the best practices in literature and will be useful to gather information depending on the accessibility to the company and employees.

Both the ISO and COSO framework agree that the different components of a RM process are: internal environment, objective setting, event identification, risk assessment, risk response, control activities, information and communication, and monitoring. These concepts will be explored in detail in the following chapter 4.

3.3 Enterprise Risk Management

ERM is a concept widely used across RM literature. It proposes that firms should "address all risks comprehensively and coherently, instead of managing them individually" [Bromiley et al., 2015]. It is linked to the idea of having a holistic approach to all the risks that one organization faces. [PwC, 2017] [BSI Standards Publication, 2018]. While the concept has already been explored across literature, in 2015 Bromiley [Bromiley et al., 2015] argued that there is still a gap of knowledge that would benefit from having management scholars studying the effectiveness of different practices and activities.

Grace [Grace et al., 2015] explored the effects of ERM initiatives on firm performance in 2015 by surveying a large number of firms. The author not only concluded that ERM programs had a statistically significant impact on cost and revenue efficiencies but also provided evidence on which are the tools with the highest gains. From the results of the survey, Grace argued that having a dedicated cross-functional Risk manager that reports directly to the board or CEO produces the highest increase in firm value.

3.4 Risk Management for Scale-ups and, Startups

Several definitions from different sources have been proposed to define a scale-up - we will adopt the broad definition that a scale-up is a company that is in the "growth phase" in the proposed life cycle by Millers and Friesen [Miller and Friesen, 1984]. This phase is characterized by already having some market success (many times referred as product-market fit) and being focused on quickly scaling up in terms of sales and resources. Monteiro [Monteiro, 2019], through a literature review, explains that all scale-ups are High Growth Firms but not all High Growth Firms are Scale-ups. Scale-ups are firms whose business model is inherently scalable.

Scale-ups, when compared to Enterprise companies [Institute of Risk Management, 2010] [Gjerdum, 2015] or Startups and SMEs, that have become a focus in recent years [Villa Todeschini et al., 2017] [Ferreira de Araújo Lima et al., 2020] [Crovini et al., 2020], are a much smaller subset of the economy and have enjoyed much less academic focus.

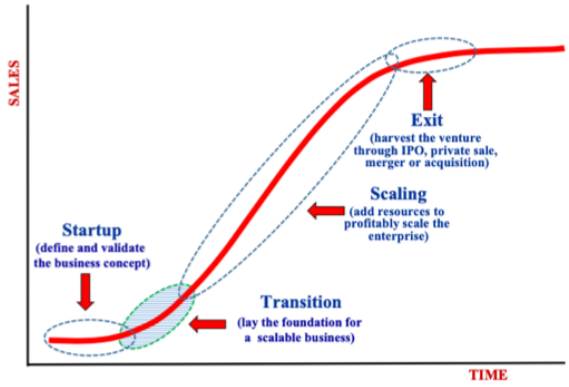

Figure 3.1: Four stages in the life cycle of an entrepreneurial firm
Source: "From startup to scalable enterprise" Picken 2017 [Picken, 2017]

They also face different challenges from a startup or enterprise company because they are in a separate phase. In a 2016 article in the Economist [Bevan et al., 2016], the authors question why "British businesses" struggle to scale-up despite many start-ups being created. This phenomena hints at the inherently different challenges that a scale-up faces when compared to a start-up. In 2017, Pickens [Picken, 2017] also agreed that scale-ups have much different needs in terms of structure and process than start-ups.

Revenue, responsibilities and expectations are much larger than on a start-up, which might only have 1-5 employees and report to a single (or no) investor. Scale-ups might already generate millions in annual review and will usually report to a board of investors on a frequent basis. Despite this loss in agility (when compared to a startup) they are still expected to grow at an exponential rate and "pivot" to quickly capture new opportunities. Process, experience and recognition are much less well established than on an enterprise company. A scale-up can be only a few years old - processes such as knowledge management, recruitment and performance review - are already crucial due to the size of the company but are new and need to adapt to the quickly changing needs of the company (going from 10, to 50, to 200 to 1000 employees in the time spawn of a few years). There is already the need to have a well implemented process for compliance and security, particularly if they operate in the Business to Business (B2B) space (which Unbabel does) and deal with large companies with strong requirements for suppliers.

As described by Monteiro [Monteiro, 2019] in 2019, High growth firms go through three main checkpoints through their trajectory: "Valley of Death", "Trigger Point" and "Turning Point". The first stage is where most startups fail and commonly happens between first receiving Venture Capital (VC) funding and generating positive cash flow. After the "Turning Point" firms then start having a linear (or exponential) and more predictable growth over time. As a final remark, the author proposes that there is a gap of knowledge about scale-up companies that should be bridged by additional vigorous investigation.

The “Trigger Point” concept used by Monteiro was introduced by Brown and Mawson [Brown and Mawson, 2013] who identified three main types: endogeneous, exogeneous and co-determined. Most modern scale-up companies, who are recipients of large investments from VC firms, have a “Trigger Point” of the latter definition (co-determined) - through a major new capital investment. After this point, firms are faced with an important transition period of high uncertainty with many “secondary” and smaller trigger points. After this transitory stage another critical moment happens when the “Turning Point” occurs, after which growth is accelerated. This moment often entails the need for more human capital, physical space and supply chain relationships as proposed by the authors. Intuitively, this notion of trigger and turning points relates to the concept of a firm chasing product-market fit and then doubling down after finding the right strategy (the moment they go from startup to scale-up).

Due to these factors scale-ups are exposed to a large portion of the same risks as startups while also facing a number of the risks that threaten enterprise companies, all without having a well established RM strategy (less than 1 out of 200 new companies will become a scale-up according to a study conducted by the consulting firm Deloitte [van Dijk et al., 2015]).

In 2017 Picken [Picken, 2017] identified the essential tasks to be undertaken by any startup to scale successfully. The author argues that the growth stage is potentially the most critical to the success of the firm. The research concludes that there are 8 main “hurdles” to reach success of which we would like to highlight the last one: “Managing Risks and Vulnerabilities”. Picken pinpoints that rapid growing entrepreneurial firms are particularly vulnerable and exposed to an environment of high uncertainty. The author goes as far as differentiating some of the risk categories that a scale-up faces - technical, market, competitive and executions - and recommends that firms should recognize and proactively manage their risks.

Nadali [Nadali et al., 2018] published in 2018 “A Conceptual Framework of Risk Identification For Scale-up Companies in Transition Period” and proposed dividing companies in the entrepreneurial environment between startups, scale-ups and unicorns. The target of the research was the the process of identifying risk in the transitory state between startup and scale-up. The author acknowledges, similarly to what is done in this thesis, that the ISO 31000 [BSI Standards Publication, 2018] main and supporting documents provides the most relevant international standards for Risk Management but that companies in growth phase face “special risks” and highlight the need to have specific classification system for this type of companies. After a review of the relevant literature the author proposes evaluating the 8 best practices to scale-up a company (using the results from Picken’s research) [Picken, 2017] under the 6 areas with the most potential to disrupt the financial goals of a company to compose a table which can be used for a structured risk identification process adapted for scale-ups. This is a relevant advancement in adapting the ISO 31000 framework to the growth stage reality but only addresses risk identification. It is proposed that further research should focus on the Risk Analysis, Evaluation and Treatment and

even on case studies. This thesis is well positioned to provide contributions aligned with the gaps left by Nadali.

Also in 2018 Pukala et al. [Pukala et al., 2018] studied 25 Polish startups with the help of 10 experts to try and understand how startups handle risk. Through a survey it was concluded that 96% of respondents either had none or only partial risk management procedures in place (while the number is likely lower for scale-ups - the order of magnitude is illustrative). The authors then proposed that insurance is an optimal instrument for financing startups business risks, taking into account that most start-ups fail to actually transfer risk (remaining self-insured).

In 2002 Davis [Davis, 2002] proposed a new method to merge the widely accepted method to make investment decisions (Net Present Value (NPV)) with the subjective but undeniably relevant risks associated with launching new products and ventures. The Net Present Value Risk Adjusted (NPVR) framework relies on using experience and judgement to subjectively assess risks. New products, platforms or ventures are evaluated with regards to Market, Technical and User risk using simple and intuitive weighted risk scorecard to produce a NPVR value. This development can be incredibly useful in the context of scale-ups because it allows for a quantitative and well structured approach when even available resources are low and uncertainty is high - by leveraging the knowledge and experience of the people within the company.

Teberga et al. [Pedro Marins Freire Teberga, 2018] further explored the concept of NPVR [Davis, 2002] in the context of a startup. The authors developed a case study around identifying, analysing and treating risks for a Brazilian crowdfunding company. At the time of the study, the chosen company, had operations involving around 38M Brazilian reais ($6M) which positions it close to the definition of a scale-up. The authors concluded that there were very little procedures in place to mitigate risks and that the framework provided by Davis was effective to identify, analyse and treat risks in the context of a startup. Additionally it was proposed that additional case studies are developed in order to understand if results could be generalized.

Given the recency of the topic and lack of established literature focused on growth stage companies, it is also insightful to look into adjacent fields such as Risk Management for SMEs (most scale-ups are still an Small and Medium Size Enterprise (SME)).

Crovini [Crovini et al., 2020] provided in 2020 a literature review of risk management practices in the specific context of SMEs (scale-ups are still an SME). The author concludes that although managers many times lack knowledge and understanding of RM practices, the main problem lies in getting SMEs to actually adopt specific procedures and processes to manage risk. To achieve this, a higher engagement between researchers and practitioners is encouraged.

Before Crovini, Verbano and Venturini [Verbano and Venturini, 2013] had already identified Risk in SMEs as a subject of growing interest but highlight that while having a growing body of knowledge it still

is a “spot” subject. The authors conclude that only 6 % of papers highlight the Risk Treatment process - which is a clear gap in the literature.

3.5 Risk Management in Technology

Technology Change Management is the process through which technology is identified, monitored, evaluated and implemented across an organisation, system or product. It is one of the key activities of any highly technological company and relates to several of the risks it faces. Mosier [Mosier et al., 2000] proposed in 2000 that processes for technology change and risk management could interact with each other to create synergies and improve results. The author argued that each step of the RM process could benefit directly from the Technology Change Management process through activities such as identifying unproven technologies, ranking risks by assessing the critical characteristics of the involved technologies and providing information on emerging technology.

Connecting technology change management to RM is aligned with the vision of having “knowledge silos” - a concept seen in a 2013 Deloitte report [Deloitte, 2013]. These “silos” would imply having specialized knowledge of the key technologies that are used, so that risk could be managed effectively. Under the characteristics of a “Risk Intelligent”, Deloitte advises having this specialized knowledge of risks - but then bridging the gap between the different “knowledge silos”, cultivating a “risk aware” culture across the organisation and pursue risk as a means to achieve a reward, and not just as a value protection mechanism.

Technology firms also have a different stance when it comes to managing risk: Kim et al. [Kim and Vonortas, 2014] studied young firms across Europe to understand how they managed risk during the “formative years”. The authors concluded that firms in knowledge-intensive sectors were much more likely to engage in strategic alliances and risk mitigation strategies. Kim confirms this kind of firms use internal risk mitigation strategies to manage the unsystemic risk, primarily related to technology and operations.

When looking at the importance of RD in the context of High Growth Firms, Holzl [Hölzl, 2009] concluded that the economic environment plays a major role. Firms located in countries working at the frontier of technology benefit much more from highly investing in RD while in “catch-up” countries the impact is lower. One possible explanation for this effect is that catch-up countries don’t have available the same human, technological and financial resources to drive disruptive innovation and highlights the importance of seeking “innovation clusters”. This effect can also be related to the “Silicon Valley” phenomena seen in the past decades, where a very large number of high growth firms all locate their offices in the same area.

In 2018 Souza et al. [de Souza Savian et al., 2018] performed a literature review of ERM in the context

of Technology based companies. After selecting 18 articles from a universe of 2323 search results, the authors concluded that literature has a knowledge gap in the study of specific ERM methodologies applied to technology firms.

Besides the gaps in the literature, small and medium technology firms also fail to apply the existing RM tools and best practices. When surveying over 200 high technology firms in Korea, Ryu [Ryu et al., 2016] observed that companies with less than 300 hundred employees hardly used any RM tools.

There is a clear gap not only in literature, but particularly in knowledge application in "real enterprise context".

3.6 Bayesian Networks

A Bayesian Network (BN) is a probabilistic graphical model that is composed of nodes (corresponding to random variables) and arcs (expressing conditional dependence). Bayesian networks use Bayesian inference to compute probability and are particularly effective to represent problems in domains where expert knowledge is probabilistic and easily modelled into a network structure. Because of their characteristics Bayesian Networks (BNs) are a viable method to perform risk analysis. They are particularly suited to deal with very innovative projects because they do not necessarily rely on a detailed data base - as they can calculate probability by taking advantage of expert opinion.

Another advantage of bayesian networks is that they are flexible in accepting inputs on the state of any node (variable) while producing a new outcome. In other words, if we correctly model the risk factors of a product and now know the state of any variable - the network will update its output. Finally, updating the conditional probabilities of a BN is a simple task (given that the correct software is available) - so they can be continuously improved as new data is available.

Because of their characteristics BNs have been applied to different use cases. Examples from the literature include applications in medicine (diagnosis and prediction) [Yet et al., 2014], operational risk in financial institutions [Neil et al., 2009] and risk assessment on the feasibility of startup projects [Akhavan et al., 2021].

Bayesian Networks have been successfully applied to model and analyse the risk of launching new products [Cooper, 2000] [Nadkarni and Shenoy, 2001]. Some research has even been done to advance the conditional probability generation [Chin et al., 2009] in this concept. While these developments advance the use of BNs for product application they were not focused on highly technological software products and use concepts that are now outdated in the industry. None of the examples use a framework that is broad enough to cover the risks across the different product related areas seen at Unbabel - engineering, design and revenue. We propose expanding on this research by combining BNs for product development risk with the modern concepts outlined in 2.2.

A key step in using bayesian networks is generating the Conditional Probability tables that inform its behaviour. There are a number of different elicitation methods that have been explored for different applications. Advanced methods are able to find inconsistent judgements which can be reviewed with the experts as seen in the research by Monti [Monti and Carenini, 2000] for example. The author modified the original elicitation method used in the AHP process [Saaty, 1987] so that it could be used in the context of probabilities. Another example from which inspiration can be taken is the MACBETH software from Bana e Costa. et al [e Costa et al., 2012] - it helps structure the problem hierarchical; elicit judgments in an intuitive way (through pairwise comparisons); rank each factor; detect inconsistent judgments and choose between different options.

Overall, there a few different methods that can be explored for probability elicitation.

4

Work Methodology

This chapter provides a review of the methodologies that will be used to frame the research question, data collection and data analysis.

4.1 Qualitative Stage

As outlined in the Literature Review 3 Bayesian Networks are a powerful tool to model risk under uncertainty conditions. They are particularly useful when there is limited data but expert knowledge is readily available. One of the characteristics of these networks is that their generation always require two stages: qualitative and quantitative. These stages can be seen across the many applications of bayesian networks in the literature: such as risk assessment in startups [Akhavan et al., 2021], new product development [Chin et al., 2009], Investment [Neil et al., 2009], and Clinical Decisions [Yet et al., 2014].

The qualitative stage is required to create the directed acyclic graph that contains the nodes and the arrows representing the risk factors and the relationships between them.

A number of authors (such as Myers [Myers, 1997], Creswell&Creswell [Creswell and Creswell,], and Yin [Yin, 1981]) have written about the basic characteristics of Qualitative Research that make it suitable to answer specific research questions. One of the most important characteristics, relevant to this project, is what Creswell&Creswell call “Participants’ Meaning”. Keeping “Participants’ Meaning” implies that the researcher keeps the focus on learning the meaning that the participants (not the academic community) hold on the problem or issue. This characteristic is what will allow the researcher to keep the language of the organisation in the resulting network.

4.1.1 Research Method - Case Study

As stated by Yin [Yin, 1981], the case study is the best strategy when a contemporary phenomena is to be evaluated in a “real-life” context. The author also argues that it is the best choice to study knowledge utilization - which is particularly relevant in this case since there is a gap between actual new product development risk management practices in the industry and literature proposed practices.

It must be acknowledged that case study research does not use sampling research. The author does not propose that through a single case study (or even a several) it is possible to find statistically significant proof to generalize from individual to population. Instead, as proposed by Yin [Yin and SAGE., 2003] in his publication on Case Study Research, the attempt is to generalize from case to theory. The goal is to better understand on how modern digital scale-ups think about and manage new product development risk using Unbabel as an example.

4.1.2 Data Collection Technique - Interviews

The data collection technique is tied to the chosen Research Method and to the intended output. In this case there is a very clear deliverable to be created which is a Bayesian Network that aggregates the expert knowledge across the “product-related” Unbabel functions.

The interview is the best method to extract the expert knowledge because it is efficient at generating knowledge transfer and allows the researcher to enquire about the thoughts and reasoning of the interviewees. The researcher is also able to conduct the conversation towards the research topic and provide examples from the state of the art but it gives the interviewee the opportunity to use the right abstraction level for him.

One of the key goals is to create a network that is easily used and explained by people in the organisation - so it must use the language of the organisation and of its people. As Myers mentions [Myers, 2009] a well run interview will allow the interviewee to use its own language - rather than one imposed by the interviewer. This is the exact goal - to use the language of the experts from Unbabel - so that the outcome mimics the terms and abstractions that are used on the day-to-day of the business. This will facilitate the adoption of the tool as it is custom "fit for purpose".

After adopting interviews as the data collection technique then the it is necessary to understand if they should be:

- Structured - All questions are pre-formulated, the interview adheres to a strict order and schedule.
- Semi-Structured - Some questions are pre-formulated but there is no strict adherence. It is possible that new questions arise during the conversation.
- Unstructured - Very little or no pre-formulated questions. There is no set time limit.

These are the 3 three basic types used by Myers [Myers, 2009] to categorize interviews.

Semi-structured interviews were selected for this research due to balance they strike between both extreme methods. This thesis is looking to formalize concepts in the specific bayesian network format and thus some structure is required to guide the interviewees. But it also attempts to keep the language of the organization and surface interesting concepts presented by the interviewees.

The following context and questions were consistent between interviews:

- Context - *"The goal of this session is to generate a network that will model new product development risk decisions at Unbabel. We will be using the key 4 principles of a successful product presented by Marty Cagan in Inspired [Cagan, 2008] - Value, Feasibility, Viability and Usability.*
  You are supposed to mention the concepts you use on your every-day at Unbabel to make formal and informal decisions. We will be trying to model your thought process on the introduction of a new product to the portfolio.
  We are going to be focused on the principle that is the closest to your role - and build a network up from its node"
- All interviewees were then presented with the example of a New Product Development Risk Bayesian Network from Chin, K et al. [Chin et al., 2009]. Which is presented in figure 4.1.

- Question - “Please identify all the risk factors that are related to your function and that you take into account when evaluating new product development risk in your organisation.” At this stage no priority is given to building causal relationships or arriving at a specific number of nodes. The aim is to build a collectively exhaustive list of risk factors - this is, to make sure all relevant risks are identified.

- After the interviewee is satisfied with the list of risk factors they are instructed to - with the support from the researcher - start building the connections between the nodes. The guidelines provided by Nadkarni et Shenoys [Nadkarni and Shenoy, 2001] are used as reference:

  – Direct Causality between Variables

  – Conditional Independence every time there are missing connections

  – Providing a brief rational behind each relationship

- Finally, the resulting network is reviewed by going through (in conversational format) each relationship. Through this process the nodes are refined, merged and gaps are identified to arrive at a final version.

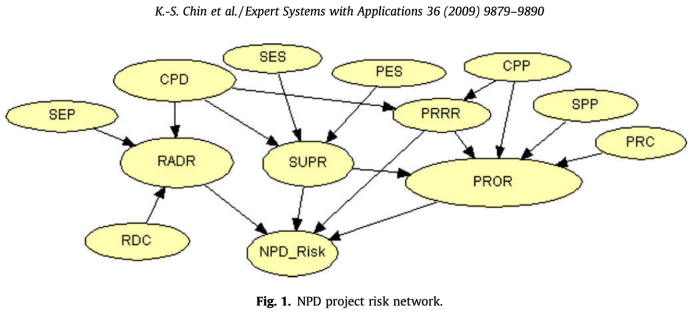

Figure 4.1: New Product Development Project Risk Network
Source: “Assessing new product development project risk by Bayesian network with a systematic probability generation methodology” Chin, K et al. [Chin et al., 2009]

All interviews were conducted online through Zoom (virtual meeting) and using Miro `https://miro.com/` as a visual aid tool. It allowed the interviewer to pre-prepare a board with materials and share the screen while making questions; the network was built as the conversation happened so that both could visually review the concepts.

The pre-prepared board also contained several other examples of bayesian networks or risk factors lists should the interviewee need. In the end - they were only used upon request - to avoid biases or replications.

4.1.2.A Selection of Interviewees

As stated in the literature review chapter 3, Cagan, M [Cagan, 2008] that the goal of product discovery is to address four critical risks (Value, Usability, Feasibility, Viability).

While in the typical modern technology company the Product Manager is responsible for preventing and mitigating all of the four - there is clear role-based specialization happening within the "Product Trio" - (Product Manager, Product Designer and Tech Lead).

As the main carrier of holistic knowledge over the customer needs, business and market - the Product Manager is best fit to lead all efforts to minimize value risk. A Senior Product Managers (or Vice President (VP) of Product) will be interviewed to build the network up from the Value node.

Feasibility and Usability Risk require much more technical and specialized knowledge to be evaluated. Assessing if the current technology allows for the creation of a given feature relies on knowledge over not only the state of the art but also the capabilities of each engineer in the company - the Tech Leads (as a senior leader of engineers) are best positioned to build the network up from the Feasibility node.

The Product Designer, as an expert in interfaces and User Experience, will have the experience to evaluate how difficult it would be to make the product easy to use and understand. It is then best equipped to evaluate Usability risk.

Minimizing Viability risk is within the scope of the responsibilities of the Product Management or Product Manager (PM) but is also aided by the deep Market knowledge that the Product Marketing Manager brings. As such, a Product Marketing Manager (PMM) will be interviewed in the context of the Viability Node.

4.2 Quantitative Stage

4.2.1 Conditional Probability Generation

Quantification of probabilities is required in order to generate the Conditional Probability Table (CPT) that inform the bayesian network. After an acylic graph has been drawn with inputs from the expert each parent-child node relationship must be mapped through conditional probabilities. This process is done with the expert by presenting scenarios for each possible parent state combination. This is, if a node has two parent nodes (A B), each with two possible states (A1,A2 and B1, B2) - the expert will provided a judgement for each possible combination (A1,B1 / A2,B1 / A1,B2 / A2,B2). There are multiple methods for expert probability elicitation as described by Monti and Catenini [Monti and Carenini, 2000] on their research around dealing with inconsistent judgements. Taking into consideration the number of CPT that must be generated and the limited nature of the expert's time - a direct elicitation method was used.

This is, the expert was asked to directly produce the likelihood (in %) of a given state given the state of the parent nodes.

While this method has many deficiencies (it doesn't assess the expert confidence on the assessment, it can lead to inconsistent judgments across the network, might be harder for the expert to understand) it has the great benefit of being much faster to collect. Under the best possible scenario a single question is needed per scenario.

If we compare against other methods that would require multiple questions per scenario or more complex formulations - the time to collect the conditional probabilities could easily increase by several orders of magnitude - which is not practical when time with the expert is limited. In future research, when expert opinion is widely available or the scope is very limited, it would be very interesting to see more sophisticated elicitation methods with the goal of reducing inconsistency in the judgements.

4.3 Considerations on Heuristics and Biases

Throughout this project expert elicitation will be one of the main methods to gather data. The challenge is that, as mentioned by Bana e Costa et al. in Decision Support Models lectures notes [e Costa et al., 2021], humans (even experts!) use primitive cognitive technique to make assessments.

It is then important to identify the main biases that will have an impact on this project:

- Availability Bias - The estimation of a probability will be impacted by how easily the expert can retrieve similar events. This bias will have a very significant impact given that the expert will be trying to recall from memory and will undoubtedly put more emphasis on more recent events. If a given project was highly affected due to a non-performing team it is likely that the expert will be biased towards the relevant of the team in the success of the project.

- Representativeness Bias - The estimation of a probability is made based on similarities within a group. It can bias the expert into not taking into account sample size (where there multiple projects with those given characteristics or only a few? Where they enough to make a judgment?).

- Anchoring Bias - The estimation of a probability is highly impacted (anchored) on the first assessment that is made. Providing a random number as an example during the interview can influence the interviewee to provide judgments within the same order of magnitude - even when there is no logical reason to do so. This is a very significant bias particularly related to how the researcher conducts the interview.

- Motivational Bias - Happens when there is a (usually hidden) incentive for the expert to provide a given assessment. It happens when the individual doing the estimation believes that the expressed judgements may affect them personally.

- Confirmation Bias - Illustrates a tendency from people to favor information that already confirms their beliefs. While it should not have an impact on the expert (his beliefs and hypotheses are exactly was should be surfaced) it can affect the interviewer if he already as made a given assessment and doesn't take into consideration how questions are made

While these biases can't be fully avoided, there are mitigation strategies that can be put in place to reduce their impact on the expert. Some examples that were put into practice were: asking the expert to recall past examples; challenging the first or any inconsistent assumption; avoiding mentioning any number as example (preventing anchoring); setting a clear academic (and not business) goal

5

Network Generation

5.1 Process

Model development is an iterative process that should be repeated several times before a valid and useful BN is established. The process to define the objectives, context and conceptual model is described in Chapter 4 and the iterative process to parameterise and evaluate the models will be described in this chapter.

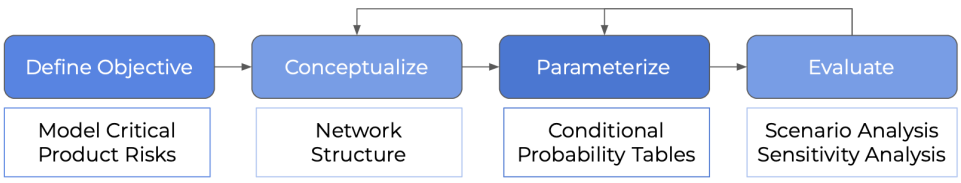

Figure 5.1: Major steps to develop a Bayesian Network

Versions will be used to better distinguish iterations of the same network. The X.Y format will be used, where increments to X and Y represent major and minor changes, respectively.

After a first round of interviews covering all of the major product risk areas - Value, Viability, Feasibility and Usability - it was decided to start the research focusing on one single network and product risk area. The purpose behind this decision was to first refine the process of building and iterating the network so that following projects would be conducted in a more efficient way taking previous learning into account.

The first product risk area to be tackled was feasibility. This is because the expert was readily available and because the concepts are more intuitive than in other areas. It is generally more intuitive to understand the causality between technical challenge and the feasibility of a project than to understand how "customer willingness to change" will impact the value of a product.

5.1.1 Sensitivity Analysis on Netica

Netica provides the ability to perform a sensitivity analysis on all or a few select nodes. This report provides information on which nodes will be the most important to crystallize the most probable states for the rest of the network. In a classification task these metrics are relevant to understand the most valuable questions to ask (i.e having limited time or resources which exams will provide a diagnosis with the highest confidence?). In the case of our network - they are useful to assess which are the most relevant risk factors. *What are the most important factors for the success of a given product development project and by how much? Should we invest in a better team? Or in having more available resources?* - these are relevant questions we can answer by using the reports Netica provides. Sensitivity Analysis is also a crucial tool for the evaluation step of building a BN - it will quantify the beliefs of the network and, when presenting unexpected values, will help raise the attention to inconsistencies.

Norsys recommends that Mutual Information is the single number which best describes the degree of sensitivity of one node to another when nodes have non continuous states. For this reason, it will be used for sensitivity analysis purposes.

Before defining Mutual Info it is useful to revisit the definition of Entropy (**H**) in the context of a random variable. The concept was first introduced in the field of Information Theory by C. E. Shannon in 1948 [Shannon, 1948]. This field studies the quantification, storage and communication of digital information and has many applications such as data compression or channel coding. While it is not the study subject of this project - its definitions help provide intuition on the meaning of the measures.

Entropy is the average level of uncertainty inherent to the variable's possible outcomes. Putting it in another way - the entropy of a random variable is the average level of information that comes from the different possible outcomes. Formally, we can quantify it in the field of Information Theory through equation 5.1:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i) \tag{5.1}$$

Where X is a discrete random variable and x_1,...,x_n are possible outcomes with given probability P(x_n). Different logs can be used to measure entropy but using 2 is a good option because it provides an easier interpretation. When using base-2 the results will come in bits. If we consider that a fair coin toss can generate 2 possible states with equal probability (heads or crown) - when tossing 1 coin we will have 1 bit of information (equation 5.2). If tossing 2 coins we will have 2 bits of information (and 4 possible outcomes).

$$-[0.5 * log_2(0.5) + [0.5 * log_2(0.5)] = 1 \tag{5.2}$$

First given by Shannon Weaver [Shannon, 1948] and based on Norsys's explanation of the software at https://www.norsys.com/WebHelp/NETICA/X_Sensitivity_Equations.htm we have about the Mutual Information Score:

- Mutual Info (**I**) - defined as the expected reduction in entropy of node Q due to findings at node F. Another way to think about it is as the reduction in uncertainty about one random variable given knowledge of another - it quantifies the "amount of information" obtained about a random variable from observing another random variable.

  "Given that I now know the Team is strong - how has this knowledge affected my perception of the likelihood of having a low feasibility risk?"

  This is particularly useful in the uncertain scenarios where BN are used because it can tell us what is the parent node which, when known, will most impact our belief about the state of the child node. In the end - the Mutual Information score can be used to quickly understand what is the most important factor for success. If Team has a higher Mutual Information score than Timeframe, then I would rather have a Strong Team on a Short Timeframe rather than a Weak Team on a Long Timeframe. This is true because I know that a change in the state of the Team node will be a better predictor of changes in the state of the Feasibility Risk node.

Defining **H** as Entropy and **I** as Mutual Information we can calculate:

$$I(Q,F) = H(Q) - H(Q|F) \tag{5.3}$$

$$I(Q,F) = \sum_{Q,F} P_{QF}(Q,F) \log \frac{P_{QF}(Q,F)}{P_Q(Q)P_F(F)} = E_{QF} \log \frac{P_{PQF}}{P_Q P_F} \tag{5.4}$$

Where Q is the query variable and F the varying variable with $P_{QF}(Q,F)$ joint probability distribution. $P_Q(Q)P_F(F)$ are the marginal probability distributions and E_{QF} is the expected value over the distribution of Q and F.

The results from equation (5.4) will come in bits. The results are best used for comparison purposes - if the Mutual Info score of Team is higher than the Mutual Info score of Technical Challenge - then the

expert belief is that the strength of the team has a higher impact on the likelihood the product being feasible.

5.2 Feasibility Network

"Can we build it?" - this is the question Marty Cagan [Cagan, 2008] asks when evaluating Feasibility risk for a product. He goes on to elaborate in his book, Inspired, that engineers have to consider a number of questions when trying to come up with an answer:

- Do we know how to build this?
- Do we have the skills on the team to build this?
- Do we have enough time to build this?
- Do we need architectural changes to build this?
- Do we have on hand all the components we need to build this?
- Do we understand the dependencies involved in building this?
- Will the performance be acceptable?
- Will it scale to the levels we need?
- Do we have the infrastructure necessary to test and run this?
- Can we afford the cost to provision this?

It is not worth investing resources into starting to build a product that will not be completed. While in some scenarios it is instantly obvious that something is not achievable because the technology does not exist or the team does not have the capability in most cases expert opinion is required to predict an uncertain outcome.

This prediction from the expert will be useful to negotiate with leadership and decide if a project should be pursued or not. The process below was followed with the support from a Senior Engineering Manager at Unbabel. This is the person that is responsible for evaluating feasibility risk for all products supported by his team.

Table 5.1 helps navigate the iterative process that was used to generate the Feasibility Network.

Table 5.1: Feasibility Risk Network Versions

Version	Expert Interview	Summary
1.0	First	First network structure without CPT
1.1	No	Minor changes to improve readability
2.0	Second	First network with CPT
2.1	Third	Team Node relevance is decreased

5.2.1 Version 1.0

The first version of the network was generated through a brainstorming session using a digital white-board tool and following the process described in 4.

With the help of the interviewer the expert started by framing the scenario in his mind. He considered that he had to evaluate the feasibility of a new product of product feature in similar context to Unbabel. In this scenario the expert the expert started by sharing the risk factors that instantly came to mind:

- Specifications Clarity - very broad specifications such as the “Uber for dogs” represent a much higher risk than well defined specifications such as “this behavior is expected when this action happens and this different behavior is expected when another action happens”.
- Team - after instantly identifying it has one of the most important components the expert autonomously started to refer the different components that impact how a team performs (seniority, culture, focus...)
- Time Frame - this factor came to discussion after commenting that team members spend their limited time into a number of different smaller tasks
- Technical Challenge, Complexity of the design and Similarity with existing products - complexity, challenge and similarity were instantly mentioned as inter-connected given an engineering task will be highly impacted by the resources (such as code) that can be re-used from previous work
- Available Technology and Available Resources - the expert remembered a specific product feature in which he had made the decision to acquire a library of front-end assets to significantly reduce development time and risk

As the ideation happened most of the relationships between factors was organically called out by the expert and noted by the interviewer. After the expert was satisfied with how comprehensive the network was the interviewer went through each node and connection out loud. This process helped identify and correct inconsistencies and was repeated until both expert and interviewer had a common understanding of meaning. The end result can be seen in figure 5.2.

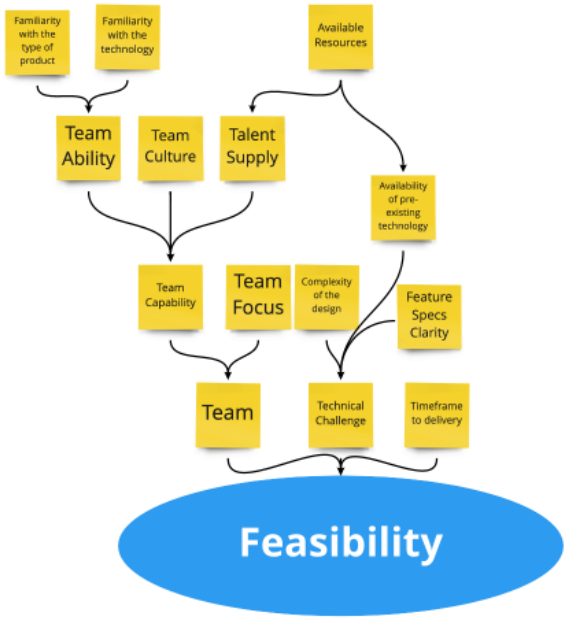

Figure 5.2: Feasibility Network Version 1.0
Source: First Interview with the Expert - Screenshot from Miro

5.2.2 Version 1.1

This version was defined by minor changes done by the researcher to improve the readability of the network and input it into Netica. In order to decrease the amount of time required from following sessions a first iteration of the Conditional Probabilities table was also input without expert help.

Some nodes were removed and others renamed in order to make the network easier to understand. These changes can be seen in version 2.0.

5.2.3 Version 2.0

This iteration of the network was created during a second 1 hour session with the expert. The goal was to input conditional probability tables directly from expert estimation. In order to simplify the estimation process all nodes were taken as if only having two possible states, each representing the worst or best possible state for the given factor.

We can see in figure 5.4 the CPT specifically for the Feasibility Risk node.

A direct estimation method was used - where the expert provided his best assessment on the probability of a given state for the child node when give the states of the parent nodes. In this first iteration the expert first saw the CPT already completed with an example provided by the researcher - this is to be avoided in future iterations due to Anchoring Bias.

Questions were placed in the following format: *"Taking your past experience at Unbabel into account,*

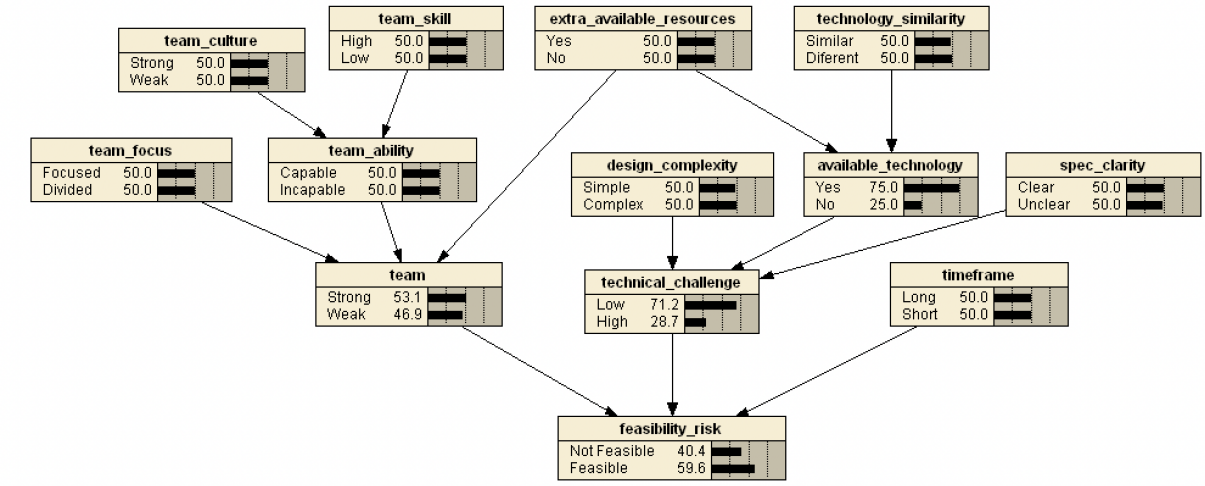

Figure 5.3: Feasibility Network Version 2.0
Source: Second Interview with the Expert - Screenshot from Netica

team	technical_challenge	timeframe	Not Feasible	Feasible
Strong	Low	Long	5	95
Strong	Low	Short	20	80
Strong	High	Long	10	90
Strong	High	Short	40	60
Weak	Low	Long	50	50
Weak	Low	Short	70	30
Weak	High	Long	80	20
Weak	High	Short	95	5

Figure 5.4: Feasibility Network Version 2.0
Source: Second Interview with the Expert - Screenshot from Netica

when presented with a Highly Skilled and Focused Team what is the likelihood they will perform as a Highly Capable Team?; *Will they be highly Capable 100% of the time? 80%?* - this cue also contributed to Anchoring Bias and will be avoided in later iterations.

This process was repeated for all scenarios for all CPT that compose the network.

The first reaction from the expert was that under no scenario he would consider a likelihood of 100% for a given state. This was a good hint that the expert displayed good probabilistic thinking and understood that given human factors were involved - all scenarios were probabilistic and not deterministic.

When making more surprising (or less intuitive) judgements the expert was questioned to better elaborate on the reasoning behind the assessment. For example, under the scenario of a team with a Strong Culture but Low Skill, the expert assessed that he would consider it a Capable team more than 50% of the time - after being questioned he explained that the lack of technical skill could many times be overcome by a strong culture of help and passion for the product.

After having collected all the probabilities from the expert the next step was to evaluate the resulting

network against edge cases to understand if it would behave logically. Two tools provided by Netica were used: “Case findings” - through which one or more nodes can be set to a specific probability (e.g: Team is set to 100% strong). And “Sensitivity to Findings” through which Netica computes the relative weight of each parent node on the child node.

The first important fact to notice is that under full middle-term uncertain conditions (all parentless nodes set to 50% on each state) the probability of the product being feasible is almost 60% (5.3). This reveals a natural tendency for optimism coming from the expert and is not necessarily inconsistent. When confronted with this fact the expert was not surprised and justified that his judgement is based on past experience at the company - where the majority of the challenges are feasible and present a relatively low technical challenge. This is also consistent with the default state for the Technical Challenge node which has a 71.2% probability of being low.

Another surprising result came from the analysis of the Feasibility Risk node present in figure 5.5, where Team was an order of magnitude more relevant than both timeframe and technical challenge. While it is not surprising that the expert considers the team the most important factor - it is unlikely that even a strong team can achieve success when faced with an incredibly complex challenge to be solved within a very short time period. Closer inspection of the CPT 5.4 reveals that the scenario where a Strong Team is faced with a High Technical Challenge and a Short Time Frame is set to 60% probability of being Feasible. At the same time all scenarios under a Weak team have 50% or less probability of being feasible. This result will be improved on for version 2.1.

Sensitivity of 'feasibility_risk' to a finding at another node:

Node	Mutual Info	Percent	Variance of Beliefs
team	0.21248	21.8	0.0674491
timeframe	0.02696	2.77	0.0089347
technical_challenge	0.02507	2.58	0.0084665

Figure 5.5: Feasibility Risk Sensitivity to Findings on V 2.0
Source: Second Interview with the Expert - Screenshot from Netica

5.2.4 Version 2.1

To improve on the shortcoming of Version 2.0 a third session was run with the expert. The focus was on the Feasibility Risk CPT. When confronted with the sensitivity to finds of the node - while not surprised about having team as the most important node - the expert agreed the order of magnitude between nodes was not accurate. With this in mind, the Feasibility Risk CPT was reviewed to keep Team as the

most important factor while reflecting the reality that no team (no matter how good) can perform well under any circumstances.

One factor that was likely to cause this bias towards the importance of the team is the lack of a very clear definition of what a “weak team” is. In this instance the subjectivity of the definitions is likely affecting the consistency of the judgements. In order to improve upon this limitation the expert was reminded of the definitions and presented with example scenarios - which in past interviews helped arrive at judgments.

- “Weak” vs “Strong” Team - As we can observe by the structure of the network a weak team is one that does not already have the required technical skills (or a fast and easy way to acquire them) and lacks focus (because they are working in multiple projects at once, for example), and. One example scenario to consider is a team composed of junior developers only or of engineers that come from a different project and have never worked with the required technology stack. This does not mean the team cannot learn and evolve through the project. A strong team would have the opposite characteristics - one can consider a team led by a strong Engineering Lead at Unbabel and that is composed of experienced engineers working with a single objective in mind.

- “High” vs “Low” Technical Challenge - A high technical challenge arises mainly from a complex design. This complexity is usually generated due to a product that requires multiple layers of technology to interact with each other. As an example it can be compared that a new dashboard for the “reporting product” - the Unbabel Portal - using data that is already available in the back-end is not very challenging because it only requires changes across one dimension - the front-end. On the opposite end - the launch of a fully new translation system (to request, review and get translations) is very complex because it requires multiple layers of complex applications - from the interface until the translation engine.

- “Short” vs “Long” Timeframe - The timeframe is always related to the product development being evaluated. When doing an estimation of the effort (in development hours) if there is a big margin between the required hours and the available ones we would consider that the timeframe is long. This margin can be invested in learning and simplifying the project (which minimizes the impact of a weak team and high technical challenge). As an example we can take the agile methodology (adopted in a modified version across Unbabel) effort estimations. After a few sprints of work a team learns what is their expected delivery rate in story points - if the project requires, for example, less than 50% of the average delivery rate - the timeframe is long. If the project is closer to 100% of the average delivery rate then there is no margin for unexpected events and the timeframe is short.

One particular important point that arose from the discussion was that the expert was considering

that there was some level of flexibility within the specifications. From his experience he knew that it is possible to negotiate the features with the Leadership or Product Manager. This fact justifies the strong bias towards the importance of the team - given technical complexity or timeframe can be highly adjusted by changing the specifications of the solution - while the team is immutable.

While it is accurate that all projects are negotiated - adjustments to the specifications would actually generate new input values to the network. So the expert was asked to consider that negotiation was not possible - to consider the specifications immutable.

Once again, specific examples of teams and projects at Unbabel were discussed to help arrive at judgements.

The reviewed table can be seen in figure 5.6

team	technical_challenge	timeframe	Not Feasible	Feasible
Strong	Low	Long	5	95
Strong	Low	Short	20	80
Strong	High	Long	10	90
Strong	High	Short	50	50
Weak	Low	Long	30	70
Weak	Low	Short	50	50
Weak	High	Long	50	50
Weak	High	Short	90	10

Figure 5.6: CPT for the Feasibility Risk Node for Version 2.1
Source: Third Interview with the Expert - Screenshot from Netica

And the resulting sensitivity to findings is in figure 5.7

Node	Mutual Info	Percent	Variance of Beliefs
team	0.08267	9.13	0.0245002
timeframe	0.04786	5.28	0.0142126
technical_challenge	0.03728	4.11	0.0116610

Figure 5.7: Feasibility Risk Sensitivity to Findings on Version 2.1
Source: Third Interview with the Expert - Screenshot from Netica

These results are now much more consistent with intuition. Team is clearest the most important factor (the expert stated that upfront) - but it is only around twice more relevant to Feasibility Risk than timeframe. Timeframe is still the second most important factor - by a higher margin than in version 2.0. This is expected given that time is more easily “interchangeable” in technical projects - it can be invested in learning new methods, solving complex issues or fixing “bugs”. Even a project that requires a low technical challenge is subject to unexpected problems - all of which require time to solve.

In summary - a strong team given enough time will be able to overcome most challenges.

An unexpected byproduct of this revision was an even stronger “optimism” to the network. While keeping all first level parent nodes at 50/50 a project was expected to be feasible within the given time with a probability of 67.9%. This result clearly comes from a much more optimistic view over “weak teams”. For example, while in version 2.0 a Weak Team under a High Technical Challenge and with a Long Time Frame is expected to fail 80% of the times - in version 2.1 they are already expected to succeed half of the times.

This drastic change also reveals some inconsistency from the expert - and maybe some bias introduced by pointing out how much worse weak teams performed than strong teams.

5.2.5 Considerations on Optimism

In order to match this optimism against the historical values found at Unbabel - the Objectives and Key Results (OKRs) completion ratio was evaluated. Given that these metrics convey how well Unbabel delivered against the objectives that are set - they should have a correlation with the product developments features being completed in time. From the available data we have that the overall (across the whole company) completion ratio was 70.26% for Q2 2021 and 61.68% for Q3 2021.

There is some noise in the data given the completion ratio is impacted by how accurately the employees report their progress but the optimism from the Feasibility Expert is exactly aligned with the historical values seen across the company - which is a strong indicator of the validity of the accuracy of his judgments. Even for Q3 2021 which has a slightly lower completion ratio - even zooming in on the RD teams closer to the expert - the average completion ratio was 73%.

The OKRs completion ratio is not a perfect metric to compare against but given that feasibility risk is not formally evaluated - this is the closest proxy. This correlation between the optimism of the Tech Lead under a “fully uncertain” scenario and the historical values of Unbabel might be explained by availability bias - the expert is judging based on the results seen in previous quarters.

5.3 Usability Network

“Can the user figure out how to use it?” - as outlined by Cagan [Cagan, 2008] many modern digital products have complex workflows and it is up to the product designers to ensure that interactions with the product make sense to the user and allow them to full advantage of its potential without creating confusion. When assessing if a user will figure out how to use the product, Designers have to ask multiple questions such as:

- Do we need to teach the user about the domain?
- Does the user know how to navigate the functionalities?

- What is the right balance between configurability and simplicity?
- Will the user give up if he doesn't quickly find value?

Usability is often overlooked when compared to Value or Feasibility due to its more subjective nature. It is undeniable that if a given product can't be built or provides no value it will face failure - while it can be argued that a less than ideal User Experience does not necessarily imply a product will fail. While this is true, the success of many digital products over their competition, not due to better functionality or engineering, but due to easier and more pleasant user experiences - illustrates the importance of Usability as a critical success factor for a product. The growth of Usability related roles in large technology companies is proof that they have realized this. A single search on `carrers.google.com` for "Design" in the United States returns almost 500 open jobs with different titles such as: Product Designer, UX Researcher, Interaction Designer, Visual Designer, Device Interaction Designer.

Product-based companies, such as Unbabel, rely on good User Experience (UX) to be able to scale. If the user is not able to "self-serve" on using the product - an Unbabel employee will have to be involved in helping set up or use the product. This "service" approach does not scale to the ambitions of technology scale-ups such as Unbabel that rely on having revenue scale much faster than costs (in a services based company costs will be mostly proportional to revenue). The process below was conducted with a Product Design or Product Designer (PD) at Unbabel. The PD is responsible for assessing, designing and testing all Usability related matters of the product.

Table 5.2 helps navigate the iterative process that was used to generate the Usability Network.

Table 5.2: Usability Risk Network Versions

Version	Expert Interview	Summary
1.0	First	First network structure without CPT
1.1	No	Structured changed after reviewing recording
2.0	Second	First network with CPT

5.3.1 Version 1.0

The brainstorming process to generate the acyclic graph (still without the CPT) was similar to the one described in 4 and for the Usability Network example. The resulting structure can be seen in figure 5.8.

The Usability expert started the discussion around risk by talking about the user. She started by mentioning scenarios where a product is designed with a young user in mind but then the actual user is an elder person. From this example we can identify two major risks: the poor identification of the user (which is surfaced as Discovery Risk) and the different usability needs depending on the persona (which is surfaced as User Risk).

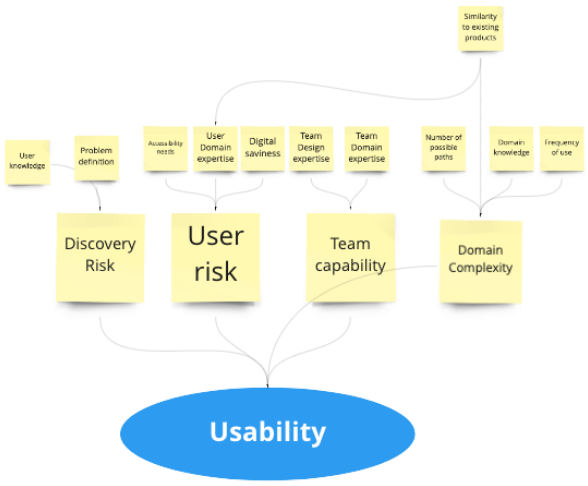

Figure 5.8: Usability Network Version 1.0
Source: First Interview with the Expert - Screenshot from Miro

"Discovery" is a term commonly used in product organizations that was popularized by Cagan [Cagan, 2008]. It is used to distinguish the the work done to figure out what is the *right product* from the work that is done to *deliver* the product. The delivery discovery dichotomy is used as a framework to explain everything that goes into creating a successful product.

From the main factors that affect user risk one that was commonly found at Unbabel was "User Domain Expertise". While the company provides translations as a service for its customers - the users and buyers of the product are usually in the Customer Support industry and do not have a deep knowledge over linguistics or translation. This often impacts the design decisions and increases the likelihood of the product being less usable. As a concrete example - Multidimensional Quality Metrics (MQM) is a translation industry standard to evaluate quality. This data point could be easily surfaced to domain experts who are familiar with the metric - but Customer Support personas require different approaches to quality.

Given we are studying digital products - the "digital savviness" is also highly relevant to assessing the user. If the user is not even familiar with the platform or device that is being used (such as mobile, a computer or a website) - it will be more complex to make it simple to use. Because there will not be any learned behaviors to work it. Accessibility is critical because if it's necessary to design for someone who is blind or can't use his hands - there are many more constraints.

Similarly to what happened in the Feasiblity Network - Team surfaced as one of the most important factors. The expert explained how it was both important to have capable designers (that understand the rules of UX) and that understand the domain in which they are solving a problem. If the designer doesn't understand well the domain - it will be harder for him to "translate" the information for the end user.

From the previous nodes - Domain Complexity surfaced naturally as a major risk. One point of

discussion was if the Domain Complexity only affected User Domain Expertise and Team Domain Expertise, but the interviewee explained that it also impacts usability directly. Designing for a hospital (let's picture a machine that supports surgeons) is more complex even if the designer and user are experts.

Another interesting node that requires further explanation is Number of Possible Paths. When calling an Uber for most trips there are only a few possible options to choose from:

- The pick-up location
- The drop-of location
- The ride type (Regular, Premium, Low-Cost)

A complex B2B product such as Salesforce or SAP is capable of producing hundreds of different outcomes (from creating reports to editing or linking entities to setting up APIs). Having many possible configurations and options (paths) increases the risk that the product will not be usable without frustration. Frustration then leads to evaluating the frequency of use. There is a much higher tolerance for poor UX when a product is only used occasionally. If I use an app every day I am more likely to search for new options as soon as I feel frustrated with the usability than if I only use it once a quarter.

The similarity to existing products is important across the board. If there are other products (from the company or not) that try to achieve similar outcomes then it is both possible to leverage learned behaviors - the user already has an expectation when a given button is clicked - and the domain might have already been broken down in a structured way.

5.3.2 Version 1.1

This version was created after a review of the recording from the first interview and to arrive at the first model on Netica. It was done asynchronously and approved by the expert at the beginning of the second interview.

The main change is that Domain Complexity not only affects Usability directly but it also influences the likelihood of the team and user performing as experts in the domain. The more complex a domain is the less likely the user and team will already be experts.

Domain Knowledge was also changed to "Domain Nature" to better express the idea that is being conveyed. Knowledge is already included in the team and user nodes. What this node is expressing is the nature of the domain: *Is it a new topic unfamiliar to most people?* Such as Artificial Intelligence or Virtual Reality? Or is it a domain that is present in our everyday lives? Shopping for groceries, for example, includes many possible paths because there a multitude of products, way of paying, sizes and flavours - but it is a very familiar domain to everyone.

The similarity to existing products was removed due to a limitation on the software license (Netica only supports up to 15 nodes on regular licenses). This node was chosen because it doesn't reduce the explainability of the network and is easily included in the evaluation of the domain complexity.

The resulting network is equal to the structure found in 5.9 but still without conditional probabilities.

5.3.3 Version 2.0

As learned during the construction of the Feasibility network it is important to define very well the states before inquiring the expert around the conditional probabilities. As such before eliciting probabilities for each child node, the expert was asked to define the extreme states under each parent and child node. The interviewer avoided intervention as much as possible to avoid creating any biases (as defined in chapter 4). Bellow are the definitions provided by the interviewee before the probability elicitation:

- Discovery Risk - We are facing a high discovery risk when there is both very little understanding of the problem that we are trying to solve for the user and a high uncertainty with regards to whom will be using the product. A low discovery risk happens when the team has had the time to research and conduct experiments resulting in strong validated hypothesis on both areas.
  - User Knowledge - Knowing your user very well is not defined on demographics but rather on the persona and objectives each type of user is trying to achieve. There is clear data and understanding around the key characteristics of the user (as can be seen in the "User Risk" node) and about the functional, emotional and social jobs they are trying to accomplish. (the "jobs" framework was popularized by Christensen in his book on innovation [Christensen et al., 2016])
  - Problem Definition - A poorly defined problem is either based on a wrong assumption or has no underlying assumption at all. A concrete example from Unbabel was used - the company wanted to motivate customers to submit quality complaints directly through the product (instead of using email) in order to store it in a structured and useful format. The starting premise was that allowing customers to complain about specific text segments would increase their adoption of the complain feature in the product. This assumption was also based on the internal needs of the company given it is useful to have quality reported at a segment level (for AI purposes). After research was conducted - through interviews and usability tests - the team realized the problem was actually that customers reported many problems at the same time. Given that the complain feature in the product required many clicks for each complain - this deterred the users from using it more. The starting premise was based on a wrong assumption and unless discovery work was ran - it would have resulted in a product with low usability.

- User Risk - This risk defines how complex it will be to get the user to the point where he can take full value out of the product. The higher the “starting ability” of the user the easier it will be to design for him because there is less need to educate and adapt to user specific needs and a higher focus on the specific problem to be solved.
  - Accessibility Needs - this characteristics is associated with any disabilities that might prevent the user from taking full advantage of the product. The most extreme scenarios (such as being blind or not being capable of using a keyboard) were not considered given they are rare and usually mean the user has some sort of helping device. Considering the nature of most products at Unbabel a high accessibility need was considered to be someone with a high level of color blindness - given that the analysis of the graphs and data points in the product could be highly hindered by this disability. No accessibility needs means the user has 20/20 vision and no other disabilities.
  - Digital Savviness - The extreme scenario of “digital inexperience” is the person who conducts no operations through any digital product. They conduct all of their tasks (going to the bank, shopping, reading the news) in person and on-site and will never or rarely use digital products. The opposite, a “digital native” is a user who conducts most of his daily operations through the internet - he uses his mobile phone and apps to achieve most of his goals.
  - User Domain Expertise - A domain expert is someone who is a qualified and experienced professional in the area. Considering Unbabel - this would either be a Linguist (with a degree) or a Customer Support professional with many years of experience. A domain beginner would be someone who has little or no knowledge over translation, customer support and AI.
- Team Ability - Defines how well the team is equipped to design the solution and will affect how capable they are in minimizing all the other design risks. A very strong team is able to mitigate discovery risk by running proper research, knows how to cater to different user needs and quickly finds solutions to simplify a complex domain.
  - Team Design Expertise - An expert team will have formal design education, several years of experience in the area and has mastered several areas of design (research, interfaces, experience). It is fully independent and can minimize all the risks that affect design.
  - Team Domain Expertise - This definition will closely align with the “User Domain Expertise”. A team that has domain expertise will have a deep knowledge over the industry and problem that is being solved. At Unbabel this would mean that (even if formal education is missing) the designer has worked closely with translation, AI and customer support products for many years and knows the space inside out.

- Domain Complexity - This is best defined through an example: Neurosurgery is a highly complex domain while transportation from one place to another is a simple domain. The first domain has a myriad of different possible scenarios and requires a very specific knowledge that few people in the world have. Transportation - which is currently being solved by Uber - is a task that most people in the world have to accomplish on a frequent basis and doesn't involve many degrees of possibilities.
  - Number of Possible Paths - This node was defined in the previous sub-chapter 5.3.1. The extreme example of a low possible number of paths is shopping online (you define a quantity and add to the cart) while the other extreme scenario is the complex product B2B Salesforce - which has hundreds of configurations and can produce many different outputs.
  - Frequency of Use - A product that is rarely used will be useful once a month or every quarter while a frequently used product is part of the daily life of the user.
  - Domain Familiarity - A well established domain relates to a trivial daily task (shopping for groceries) while a new domain is largely unknown to most people such as AI.

After all scenarios and definitions were clear and agreed between the researcher and expert then the probabilities were directly elicited. The interviewer aimed to avoid as much anchoring, confirmation and availability bias as possible. The resulting network can be seen in figure 5.9.

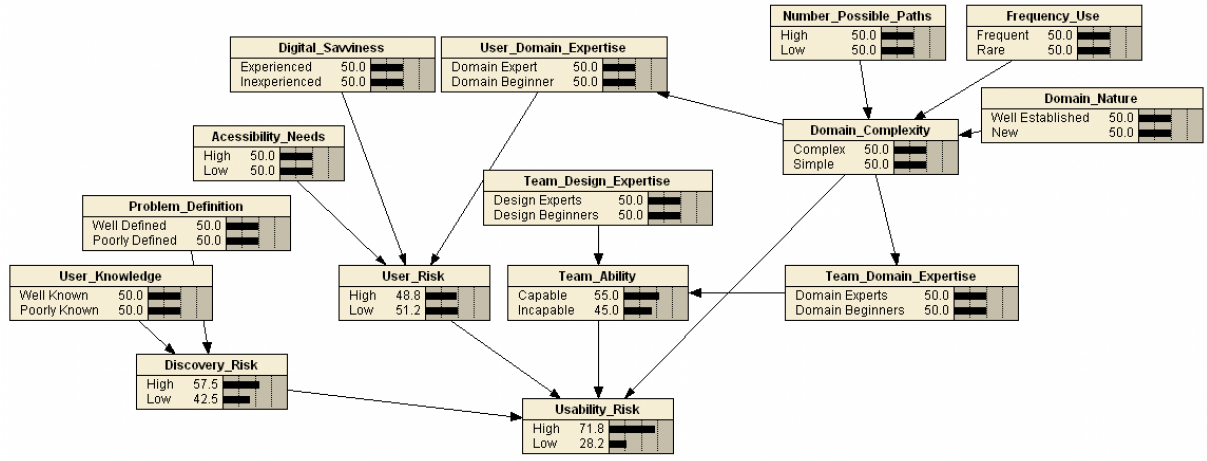

Figure 5.9: Usability Network Version 2.0
Source: Second Interview with the Expert - Screenshot from Netica

The first interesting point to note is how different the optimism is between the feasibility expert and the usability expert. While the engineering leads considers that by default 60% of the products will have a low feasibility risk - the designer considers that more than 70% of the time the usability risk will be high. There are a few possible explanations for this discrepancies between the experts:

- Lack of correlation between the risks - Feasibility and Usability are different and not necessarily correlated. It is very likely that the products that this team owns don't normally represent a very high feasibility risk while representing a huge design challenge. This is a good hypothesis given both experts work on the "Unbabel Portal" - a highly visual interface with many different graphs and data points. Most of the technical work conducted by the team is related to front-end work and doesn't deal with the most complex technical areas of Unbabel (AI and architecture) but the product does have the most complex and user facing interfaces of the company.
- Biases - it is possible that one (or both) of the experts are highly impacted by different biases and that lead them to having very different judgments. It could be, for example, that most recent projects where highly feasible but very challenging in terms of design - which combined with availability bias could lead to these results.
- Nature of the experts - It could also be related to the characteristics of the experts being inquired. Some people are naturally more optimistic than others - and their judgments need to be adjusted and normalized.

The correlation between the optimism of the expert and the historical Unbabel results observed in the chapter on Optimism is not observed for the Usability Risk. This is not unexpected given that OKRs will mostly evaluate the delivery of new features (the implementation of the technology) rather than adoption from users or increases in usability.

Sensitivity of 'Usability_Risk' to a finding at another node:

Node	Mutual Info	Percent	Variance of Beliefs
Discovery_Risk	0.11997	14	0.0331496
Team_Ability	0.04317	5.03	0.0116665
Domain_Complexity	0.01369	1.6	0.0038177
User_Risk	0.01154	1.34	0.0032135

Figure 5.10: Usability Risk Sensitivity to Findings on Version 2.0
Source: Second Interview with the Expert - Screenshot from Netica

In figure 5.10 we can see the Sensitivity to Findings for the Usability Risk node. Discovery Risk is clearly the most important factor for the expert. This finding is consistent with the intuition from the expert who said (paraphrasing) : "In any scenario where the Discovery Risk is high it will always be very likely (above 80%) that there is a high Usability Risk. Because if you don't have a strong understanding of what you're trying to solve - you'll most likely not arrive at a usable product". Throughout the interview process the expert put an emphasis on how important discovery is to minimize the risk of a product that

fails. This observation is aligned with both Cagan [Cagan, 2008] and Torres [Torres, 2021] who state that discovery is key to building the right product. It is also consistent with field observations at Unbabel where the Discovery process is mostly lead by the Product Designer and Product Manager - who deeply value a sound understanding over the problem and user - while the Tech Lead will usually lead discovery.

From Discovery Risk we find that the Problem Definition is much more important than User Knowledge 5.11. While the order of magnitude is unexpected (almost 10 times more important) this is also consistent with the shift towards the "jobs to be done" framework [Christensen et al., 2016] where demographics and user profiles become less relevant to building a product when compared with understanding over the "job to be done" or, in other words, having a well defined problem.

Sensitivity of 'Discovery_Risk' to a finding at another node:

Node	Mutual Info	Percent	Variance of Beliefs
Problem_Definition	0.28765	29.2	0.0900000
User_Knowledge	0.04668	4.74	0.0156250

Figure 5.11: Discovery Risk Sensitivity to Findings on Version 2.0
Source: Second Interview with the Expert - Screenshot from Netica

Team is the second most relevant factor to minimizing Usability Risk. Further analysis of the sensitivity to findings of the Team Ability Node show that the Team Design Expertise has a Mutual Info Score 4 times higher than Team Domain Expertise 5.12. While it can be explained that an expert design can quickly acquire domain knowledge or can compensate through research skills - it is also likely that there is some confirmation or motivational bias from the expert. Given that an expert designer is being interviewed - it is normal that they might value their own skills (design) over domain knowledge skills.

Sensitivity of 'Team_Ability' to a finding at another node:

Node	Mutual Info	Percent	Variance of Beliefs
Team_Design_Expertise	0.12010	12.1	0.0400000
Team_Domain_Expertise	0.02935	2.96	0.0100000

Figure 5.12: Team Ability Sensitivity to Findings on Version 2.0
Source: Second Interview with the Expert - Screenshot from Netica

User Risk and Domain Complexity appear as much less relevant than both Discovery and Team. This is probably explained by the nature of Unbabel's products - given the limited number of users in a very specific context (Unbabel is usually bought by experienced executives and used by Customer

Support agents and mid level management) it is unlikely that products were affected in the past by the nature of the user. While Unbabel uses very advanced technology (particularly AI) it is mostly handled by Unbabel employees and not surfaced to customers - which greatly simplifies the domain in which users need to work. Thus, it is possible that Representativeness Bias played a part in the judgments.

From Domain Complexity it is particularly interesting to note how little impact Frequency of Use has when compared to Number of Possible Paths (with the highest Mutual Info score by a large margin) and Domain Nature.

While during the first interview the Expert stated that the frequency of use was highly relevant because it could impact the level of frustration the user was willing to accept - when eliciting probabilities the expert realized that they would provide very similar judgments. The designer explained that while we are much more willing to spend a few minutes looking for a functionality in a product we only rarely use - in a product we frequently use the complexity itself is reduced because he can quickly learn and improve over time. This was a very interesting moment because while the expert recognized it as a very important factor - the elicitation method led him to the realization that it ended up having little impact on the Usability Risk. In the end the CPT for Frequent and Rare Frequency of use was the same 5.13.

Frequency_Use	Number_Possible_Paths	Domain_Nature	Complex	Simple
Frequent	High	Well Established	70	30
Frequent	High	New	90	10
Frequent	Low	Well Established	5	95
Frequent	Low	New	65	35
Rare	High	Well Established	70	30
Rare	High	New	90	10
Rare	Low	Well Established	5	95
Rare	Low	New	65	35

Figure 5.13: Domain Complexity Conditional Probability Table
Source: Second Interview with the Expert - Screenshot from Netica

5.4 Value Network

Before diving into the Network it is important to define that Value within the Product industry is centered around the user. It is not the financial value that the company is able to generate but rather the benefit that a customer gets by using a product to satisfy their needs, minus the costs (or effort). A similar definition can be found across different modern technology and product thought leaders [Palan, 2021] [Employee, 2021] [Autor, 2022].

The business (financial) value that a company is able to extract from users is further explored under the definition of Viability. 5.5

Cagan, in his foundational book “Inspired” [Cagan, 2008] goes further by asking *“Will the customer buy, or choose to use, this product?”* - in order to evaluate if a product has value. The author also

argues, in a later blog post, that Value is one of (if not the) most important risks and it should be addressed earlier. [Cagan, 2017]

While not sufficient, providing value to users is a required condition for the success of any product. A confusing interface, a slow loading time or an expensive subscription can be accepted by the user - but if the product or feature is not solving a real problem or "doing a job" for the customer, then it will not succeed. It is also not only about solving a (real) problem for the customer in an effective, but also doing it (much) better than the competitors.

A valuable product will solve a (real) problem better than the competitors and drive many customers to use it

This is why the product team will spend most of their time minimizing the risk of prioritising products and features that deliver little to no value.

The Product Manager, who is the main (but not sole) owner of the Product Value is trying to get multiple answers:

- Is there a real problem (pain) to solve?
- How much do customers actually care?
- Are they willing to pay or spend their time to solve the problem? (How much?)
- How many people actually go through this? (How big is demand?)
- How are customers currently solving this problem? (What is the competition level?)

Value is, at the same time, most likely the easiest risk to understand and recognize but one of the hardest to measure. There are very specific scenarios in which one could precisely quantify the value a product brings: when talking about advertising technology - we could use the revenue generated as a key indicator. But this this is not the reality for most products. Many products either don't directly affect revenue/costs and/or also have other unquantifiable effects.

A good framework to understand how a product can impact the user, and that was briefly discussed in the Usability section of the current chapter, is the "Jobs to be Done" framework. Christensen [Christensen et al., 2016] argues that "people buy products and services to get a job done" and identifies three main types of jobs:

- The core functional job - this is the main process the user is trying to get complete. It is the reason the market exists. Google Maps delivers the job of deciding the best trajectory from one place to another in the least amount of time possible.
- Related jobs - these are additional jobs the user might be trying to get done before, during or after the execution of the core job. Finding a parking spot is a related job to finding the best trajectory.

- Emotional and Social jobs - are related to the way the user is perceived or feels when completing the core job. Overcoming the anxiety of arriving late or getting stuck in traffic is an emotional job achieved through Google Maps.

As a Product Manager working on Google Maps - should I prioritise launching “dark mode” (a feature where the screen adjusts the color scheme to be easier on the eyes) or a feature to share your location with friends? Neither feature contributes directly to the core functional job of getting to the location in the least amount of time but both can contribute to the value the product provides and to the decision of the user on whether to use the competition or not.

Moreover, we have to consider that for each user segment each job will play more or less significant roles. And each user segment could be larger or smaller, more or less profitable... From research we understand that in other applications 60% of the users use “dark mode” on a daily basis while only 5% have installed any sort of application to share their location. If we evaluate by the size of the user base then we would choose “dark mode”, but what if this feature is only a “nice to have” that doesn’t justify switching products from the competition - while the 5% of users would definitely move to Google Maps and become loyal customers to be able to share their location?

Even if we know we are solving a real problem. Do we have a big enough gap over the competition to justify a change? Depending on the product there is always a level of friction to switching between solutions. Marty Cagan states it clearly when he says *“just because someone can use our product doesn’t mean that they will choose to use our product”*. [Cagan, 2008]

Reliability, UX, or the business model can be adjusted individually - code can be refactored, new interfaces deployed and price changed without inter-dependencies. Pivoting the value of a product might imply changes to all of the previous mentioned - one of the reasons why it is so important to “get it right”.

Table 5.3 helps navigate the iterative process that was used to generate the Value Network.

Table 5.3: Value Risk Network Versions

Version	Expert Interview	Summary
1.0	First	First network structure without CPT
1.1	Second	Structured reviewed with support from recording and expert
2.0	Second	First network with CPT
2.1	Third	Reference Customers identified as a “symptom”

5.4.1 Version 1.0

The process to generate the network structure was similar to the one followed for the previous networks. For this instance, the interviewee was a senior executive with more than 30 years of experience in building technology products.

The interview started with an interesting reaction from the expert: when looking at the bayesian network from Chin et al. [Chin et al., 2009] described in 4, he commented that these were “old-school” concepts that were not used anymore in the industry. This observation further validates the hypothesis that previous Bayesian Network approaches to mapping Product Development Risk were interesting but the used concepts are now outdated. There is an opportunity to review and update the terms to better reflect how modern technology companies build product - and the purpose of this thesis is to make a contribution.

The expert then started to outline how he thinks about risk while the interviewer noted down the key ideas using the virtual whiteboard tool. The resulting (still unstructured) concepts can be seen in figure 5.14.

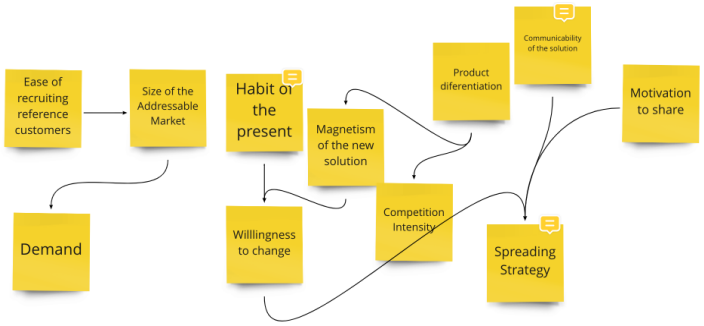

Figure 5.14: Brainstorming of value risk factors
Source: First Interview with the Expert - Screenshot from Miro

5.4.1.A Demand

The executive explained how important discovery practices **??** are to assess the value of a given product in an efficient way. He explained how the ease of recruiting reference customers (a common practice in product discovery to collect early customers to test and discuss products and features) is a good indicator for the demand of a given product.

5.4.1.B Spreading Strategy

Quoting his experience working with VC firms the interviewee mentioned how crucial the spreading strategy is. It includes the way through which new customers will get to know, experience and buy the product. Customers will only adopt a product if they understand the value proposition and know it exists. The concept of spreading strategy could be explained by relating it to Product-led Growth (PLG) - which is a common term in the technology industry. According to the Product Led Growth Collective [Collective, 2022], a community started around the topic, PLG is a business methodology in which user acquisition, expansion, conversion, and retention are all driven primarily by the product. This strategy comes in

contrast with a Sales or Marketing led organisation, where money is invested in marketing campaigns and hiring Sales representatives. A product with an effective spreading strategy will more easily pursue a PLG strategy.

5.4.1.C Willingness to Change

The "Jobs to be done framework" [Christensen et al., 2016] was then used as a reference to explain that there are multiple forces driving the decision from the user (this is not the same concept as the "jobs" mentioned before, but is part of the same framework.) There are "status-quo" forces pushing the customer to keep his current solution and "innovation" forces moving the customer into the new product. To illustrate this concept the expert showed a visual representation of "the forces" (image 5.15).

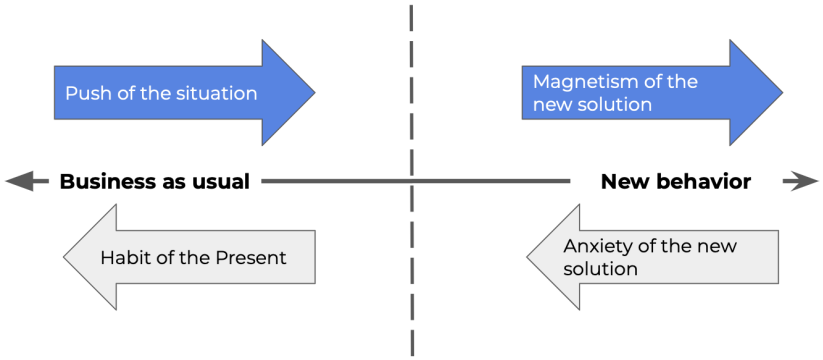

Figure 5.15: Forces pushing the user into or away from adopting a new solution. A valuable product will have stronger forces pushing the user into adoption (new behavior) than pulling into keeping the current solution

Source: Based on `jobstobedone.org` [Spiek, 2012]

The magnetism of the new solution represents how well the product solves the problem while the habit of the present will be stronger the less the user is willing to change the status-quo.

5.4.1.D Competition

Previous products from the company were discussed to highlight that it is not enough to "match" the qualities of the existing solutions (feature parity is not enough). For a user to be willing to change it is required that the problem is solved "10x" better when compared to the competitors. This led to another key factor which is the competition level.

5.4.1.E Communicability of the solution

When making the decision to acquire a product the user always needs to be convinced but many times there are other stakeholders. Under a B2B setting a prospective user might have to convince his Man-

ager (who is the actual buyer). This acquisition process will be as easy as it is to communicate the effectiveness of the product in solving the needs of the user and the buyer. If an executive is evaluated under a specific Key Point Indicator(s) (KPI) - a product that has a clear demonstrable impact will be easily adopted.

5.4.1.F Motivation to Share

One of the main mechanisms for a product to spread is to have current users actively inviting new users to try it out. This behavior will only happen if there is a motivation for the user to share: a reward, which can be of two types:

- Explicit Reward - Happens when there is an "prize" for the user who brings in additional users. This is mainly found in Business to Customer (B2C) products - the paying system Paypal (`https://www.paypal.com/hk/webapps/mpp/invite`) ran a campaign that awarded $10 to customers if new users signed-up using their link.
- Implicit Reward - If a product makes the user be perceived under a "better light" it has more chances of spreading. There is an intrinsic motivation to share.

  If we consider B2B products: revenue teams use the term "Customer Champion" to identify someone in the customer organisation who is both happy to use and share his experience with the product. A VP of Customer Support might be perceived as a better manager if he is using Unbabel to serve customers across more than 20 languages under a single English-speaking team instead of using agent teams scattered across the world. There is a clear connection to the Social/Emotional jobs discussed earlier in the chapter.

Revisiting figure 5.14 we can find an intuitive definition. A product will have a higher chance of being adopted if it solves a real problem better (Willingness to change), if that value is demonstrable (communicability), and if the user has (explicit or implicit) motivations to share

$$Chance of spreading \propto Willingness to Change + Communicability of the solution + Motivation to Share$$

Instagram `https://www.instagram.com` and Facebook `https://www.facebook.com` are a few of the best known examples of products in which the motivation to share and communicability played a key role. Facebook rolled-out progressively first in prestigious universities within the United States, later to Europe and only then to the broader audience. [Phillips, 2007] Users both felt part of an exclusive group (implicit motivation to share) and got network rewards from recruiting friends.

To note that the spreading strategy is relevant both for Value and Viability 5.5. For Value because it increases the odds of adoption from additional users and Viability because it reduces the costs of acquiring new customers (which is relevant to the business viability of the product).

After structuring all the factors that were generated during the brainstorming a first version of the network was built. 5.16. Note that a few concepts were renamed by the researcher in this final stage: Habit of the Present to Barrier to Change, Communicability to Demonstrability (which was segmented into buyer and user), Spreading Strategy to Growth Strategy.

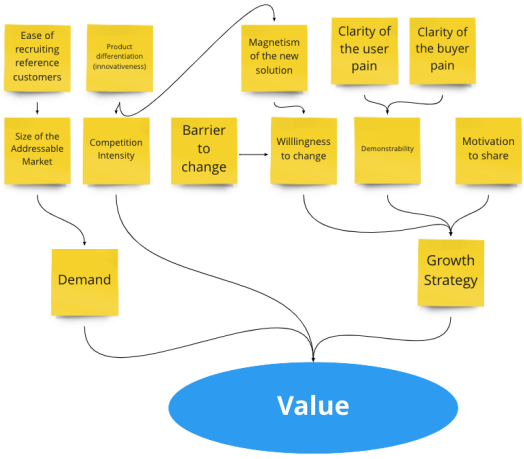

Figure 5.16: Value Network Version 1.0
Source: First Interview with the Expert - Screenshot from Miro

5.4.2 Version 1.1

The generative process for version 1.0 was more complex and less structured than for the Feasibility 5.2 and Usability Network 5.3. As a consequence, the resulting network was harder to read and interpret by someone outside the industry.

In order to improve on this situation the recording of the first interview was reviewed and the second interview started with a revision of the network structure with support from the expert. The resulting structure can be seen in figure 5.17.

The main changes from Version 1.0 to Version 1.1 are the following:

- Reference Customers - the ease of recruiting reference customers is a good predictor not only of the existence of demand but also of willingness to change from users. Being a reference customer requires (at least) a time investment - which users are only willing to do if they expect to have a meaningful problem solved in a better way.
- Demonstrability - while important it is actually not independent of the other factors (it is a sub-section of magnetism, network effects and motivation to share). So it was removed.

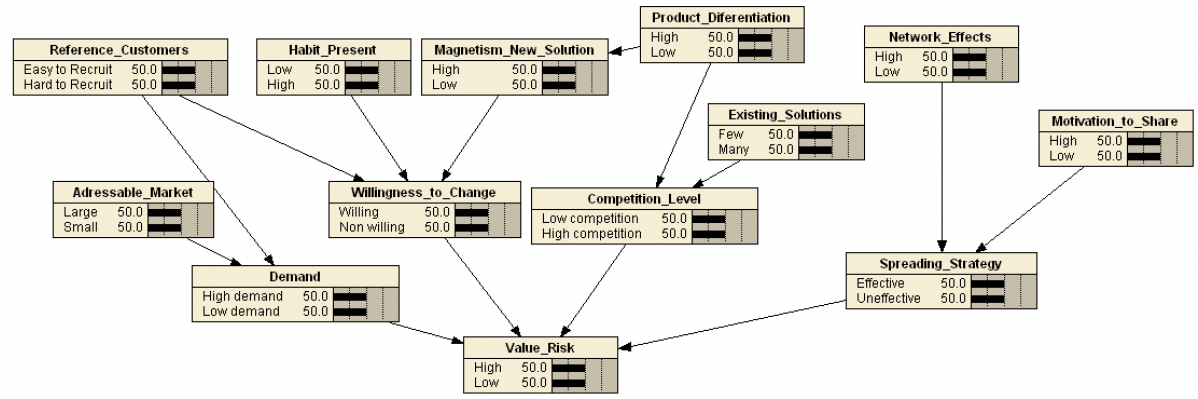

Figure 5.17: Value Network Version 1.1
Source: Second Interview with the Expert - Screenshot from Netica

- Willingness to change - was decoupled from Growth Strategy and connected directly to the Value Risk node. This better reflects the difference between a product that solves a real problem well (causes a high willingness to change in users) and a product that "organically" spreads through a large number of users.
- Network effects - this node captures the advantage that some products have whereas value grows proportionally (or exponentially) with the size of the network of users. When this phenomenon occurs - the product has a much higher chance of spreading because users will be drawn to the product that others people use causing a compounded effect. Social Networks and Marketplaces are the best examples of these type of products.
- Existing Solutions - the competition level is going to be the determined by the number of competitor products minus how different and disruptive the new product is (product differentiation).

Some of the renaming of nodes that was done in figure 5.17 was reverted as the expert preferred the initial terms.

5.4.3 Version 2.0

Similarly to what was done (successfully) to build the CPT of the Feasibility Network the expert was asked to define all the scenarios before providing estimations. This exercise provides a much clearer understanding of what the expert considers to be a situation where Demand is "high" versus "low" or what is a product that has "high magnetism".

Bellow are the definitions (and examples) for all nodes:

- Reference Customers - A reference customer is an early "real customer" (with no hidden motivations) that is willing to run a product in production and both make some sort of investment

(monetary or in time) and willing to tell others how much they love the product. Note that within all the networks this is the only “symptom” node. A Reference Customer is also usually eager to provide feedback that will improve the product. This definition and concept is further explored in Version 2.1 5.4.4.

- Demand - A high demand implied that there is a high number of customers that have a specific problem and where the pain is strong enough for them to be willing to invest money to solve it. The absolute definition of what “high” is will depend on the market being studied.
- Willingness to Change - is the amount of friction when changing from one of achieving an outcome (old solution) to a new way (new solution). It is best described at the beginning of this chapter. An example of a high willingness to change would be a user who has no fidelity to the current solution and is amazed by the capabilities of the new product. An example of a low willingness to change would be the traditional process to change banks - people trust a specific organisation, there is a high amount of paper work to change and there is no high differentiation (and thus no High magnetism) on the product offering between banks
  - Habit of the Present - Continuing the banking example: if a user is already accustomed to visiting a specific office, with a specific account manager and uses a specific app (because he has been with the bank for 20 years) there will be a strong habit of the present preventing change
  - Magnetism of the New Solution - is proportional to the “excitement” that the new solution triggers in the user. Spotify provided a completely different and innovative way to access 1000’s of musics when compared to locally stored files - and users were eager to try it out.
- Competition Level - Is a measure of how much a product can stand out amongst the competition. it will be proportional to the amount of existing solutions in the market (a crowded space increases competition level) and inversely proportional to the delta the product can create through product differentiation. Space Tourism is (for now) an industry with few existing solutions while second-hand clothing marketplaces are a crowed space with low differentiation.
- Spreading Strategy - Is comprehensively defined in the beginning of this chapter. An example of a product with a very effective spreading strategy is Linkedin `https://www.linkedin.com` because the value of a social/professional network grows with the size of the network and because there is an implicit motivation to share in displaying professional achievements to peers. Many B2B products could be used as examples for ineffective spreading strategies because customers rarely have motivation to share the solution they use and the value is usually decoupled from the size of the network.

It should be noted that the probability elicitation for the Value Risk node was particularly challenging. As the number of possible states increases with the number of parent nodes the exercise becomes more and more abstract for the expert. The CPT for Value Risk contains 16 rows - which opens up at a high risk of inconsistent judgements caused by confusion. This should be addressed in future work - and other elicitation methods, already used in well established methods such as Macbeth [e Costa et al., 2012] or AHP [Saaty, 1987] are a potential mitigation strategy.

After the expert and researcher agreed and clarified all definitions the CPT were built through direct elicitation. The resulting network can be seen in figure 5.18

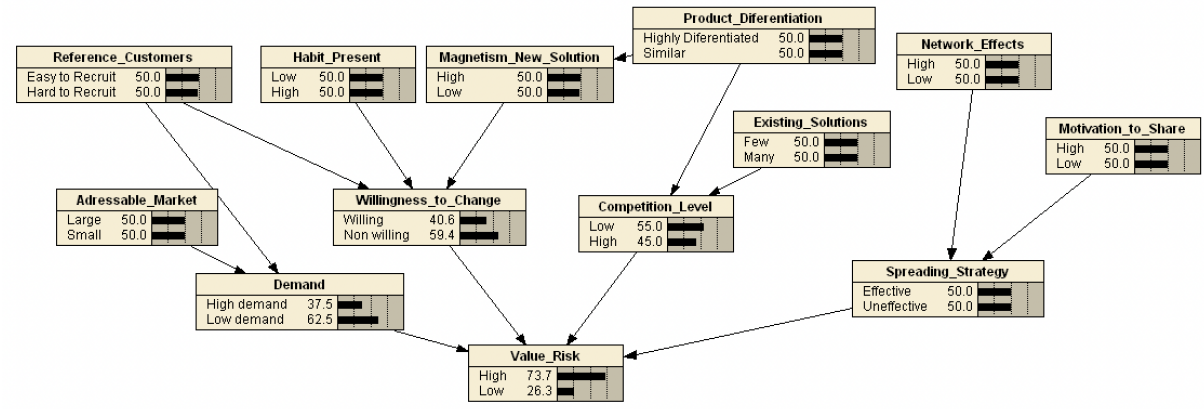

Figure 5.18: Value Network Version 2.0
Source: Second Interview with the Expert - Screenshot from Netica

Similarly to what happened with the Usability Network 5.3, under a fully uncertain scenario Value Risk sits at almost 73% chance of being high. This pattern across usability and value could be related to how much they rely on discovery vs delivery 2.2. While the technology that Unbabel builds is within a well defined and known scope - the disruptive nature of most of the products means that there is high risk of "not building the right thing" (failing at discovery - highly associated with value and usability) even if they manage to "build it right". (getting delivery correct - highly associated with feasibility)

From the possible explanations for this discrepancy outlined in the previous sub-chapter 5.3 the Nature of the Expert is highly relevant. The interviewee is an experienced executive in the "innovation space" (disruptive rather than iterative improvements) - where a higher failure rate is expected. One of the teams led by this professional has the motto of: *"If less than 50% of the products fail, we are not being ambitious enough"* - which is aligned with a more "pessimistic" view of the rate of success for products.

When analysing the Sensitivity to Findings for the Value Risk and parent nodes 5.19 Demand comes out as the most important factor by a large margin. This is consistent with intuition that the starting point for any valuable product is to solve a real problem - which is the first question outlined at the beginning of this chapter. Willingness to change is the second most important factor which is also expected given

Sensitivity of 'Value_Risk' to a finding at another node:

Node	Mutual Info	Percent	Variance of Beliefs
Demand	0.19474	23.4	0.0520645
Willingness_to_Change	0.07385	8.88	0.0199816
Competition_Level	0.00635	0.764	0.0016916
Spreading_Strategy	0.00183	0.221	0.0004929

Figure 5.19: Value Risk Sensitivity to Findings on Version 2.0
Source: Second Interview with the Expert - Screenshot from Netica

that it represents the “force” that is driving the user into a new behavior.

While the Usability and Value expert took different approaches to structure their networks - they’ve converged into having discovery as the most important risk factor. The Usability expert mentioned discovery explicitly while the Value expert focused on Demand and Willingness to Change - which, together, are the closest to the definition of discovery within the value network.

Competition Level and Spreading Strategy have a surprisingly low impact on the Value Risk. Even under the scenario where we are 100% confident that Competition Level is High and the Spreading Strategy is effective, the probability of the Value Risk being high only moves to 79.4% (around 5%).

Isolating the impact that of Habit of the Present and Magnetism of the New Solution have on the willingness to change we get to the following Mutual Info scores 5.20

Sensitivity of 'Willingness_to_Change' to a finding at another node:

Node	Mutual Info	Percent	Variance of Beliefs
Habit_Present	0.03407	3.5	0.0112891
Magnetism_New_Solution	0.02647	2.72	0.0087891

Figure 5.20: Willingness to Change Sensitivity to Findings on Habit of the Present and Magnetism of the new solution
Source: Second Interview with the Expert - Screenshot from Netica

As expected, they are within the same order of magnitude but the Habit of the Present is slightly a better predictor of Willingness to Change than Magnetism of the new solution. This leads to an interesting conclusion that the success in making a user move from one product to another is better dictated by how strong their current habit is than by the magnetism of the new solution. Even under a smaller quality delta between new and old solution - it is a better investment to launch a product in a category where the current habit is not as strong.

Looking at Competition Level, the Mutual Information score for Product Differentiation (0.282) is one

order of magnitude higher than Existing Solutions (0.029). Looking at the CPT we can see that the expert considers that even under a "crowded scenario" where there are many solutions, competition level will be low with 80% confidence if the Product Differentiation is high (see figure 5.21)

Product_Diferentiation	Existing_Solutions	Low	High
Highly Diferentiated	Few	90	10
Highly Diferentiated	Many	80	20
Similar	Few	40	60
Similar	Many	10	90

Figure 5.21: CPT for the Competition Level Node for Version 2.0
Source: Third Interview with the Expert - Screenshot from Netica

5.4.4 Version 2.1

A deeper analysis of the logic behind the network reveals that the Reference Customers node has a different nature from the others. The existence of Reference Customer does not cause neither Demand nor Willingness to Change - it is actually the reverse. If there is a high demand (large addressable market) and customers are willing to change into new solutions it will be easy to recruit early users to test and provide feedback on the product (Reference Customers). As such, while all other nodes are connected as causal factors, Reference Customer is a "symptom", a signal, that updates our confidence on the state of Demand and Willingness to Change.

To reflect this reality, the direction of the relationship (represented through the arrows) was reversed and the CPT were updated with support from the expert. These changes can be seen in figure 5.22. The color of the node was also changed to better reflect its different nature.

This type of node did not appear in the other networks - which can be explained by the elicitation method and wording. As referenced in chapter 4 the wording of the questions when building the network was around risk factors and causality rather than around any data point that could update our confidence in the risk level. The potential of including "symptoms" in future iteration is very interesting and the implications are further discussed in chapter 6.

The review of the CPT caused a slight change in the "optimism of the network" which moved from 73.7% to 68.9% mainly caused by the change in the state probability of the Demand Node. In version 2.0 Demand was High with 37.5% confidence due to the CPT including Reference Customers as a Causal Factor. This new iteration is more accurate and the slight change in "optimism" of the network does not change the conclusions draw for the previous version.

Addressable Market was removed as a node to simplify the structure and given that (as a sole parent node) it is redundant with Demand.

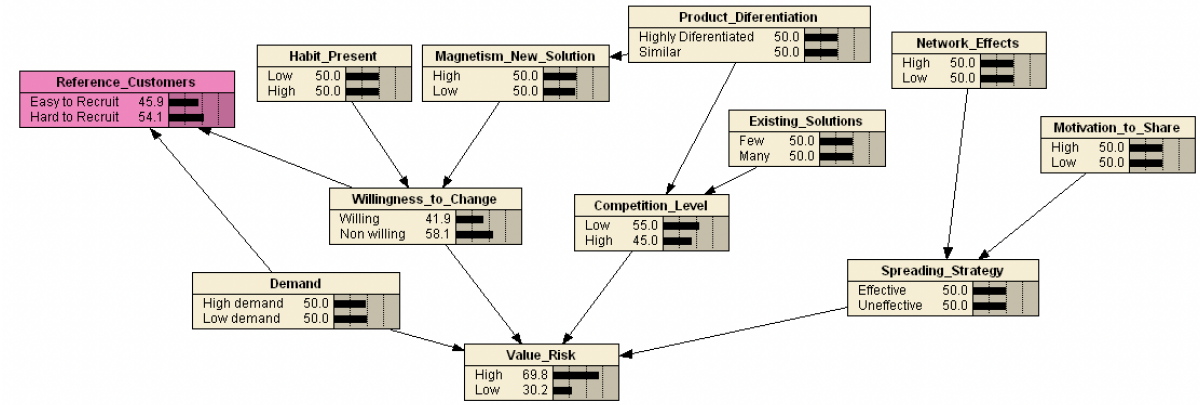

Figure 5.22: Value Network Version 2.1
Source: Second and Third Interview with the Expert - Screenshot from Netica

5.5 Viability Network

"Does this solution work for our business?" - is the the broad question Cagan [Cagan, 2008] asks when assessing the business viability of a product. While value 5.4 is centered around the user, viability means having a solution that works for the business, that both generates profits and complies with all the constraints of the organisation. A PM or PMM might ask the following questions to understand the business viability of a product:

- Are the costs of producing, marketing and selling the product significantly less than the revenue it generates?
- Does it comply with law of all the countries in which we want to launch?
- Are there security concerns that put the business at risk? (also evaluated within feasibility 5.2)
- Does this product fit within our current brand?
- Will it conflict or harmonize with our other offerings?

Business Viability correlates directly with the ability of a business to make profit (which is its ultimate goal) so, unlike usability or user value, it has always been a focal point for organizations building products. Where modern technology companies differ from more "traditional enterprises" is in the variety of business models that are possible for software products and in the focus on the cost of marketing and selling the product. Different business models allow for different ways to make a profit that are more aligned with the cost structure of highly technological products - usually characterized by high fixed costs and very low variable costs.

SaaS is a great example of new business models made possible through technology. A report from McKinsey [Dubey and Wagle, 2007] from 2007 introduces SaaS as *"A new delivery method is shaking*

the software industry's foundations. Traditional vendors should take heed.". Nowadays, less than 15 years later, this business model is so wide spread that the top 4 companies from the S&P500 all offer some sort of SaaS products. (Apple, Microsoft, Amazon, Alphabet)

It is also important to consider that the business viability is important in the context of the organisation and not just "within the product itself". It is not enough for the product to be profitable - it should make the whole business more profitable in a sustainable way - so it should not cannibalize other products, "dilute the brand" or confuse customers through an inconsistent offering.

There are components to business viability that are well suited for objective, quantitative study - such as the cost of manufacture and the retail price. The profit or break-even point for a given product can be objectively calculated using data analysis tools such as Microsoft Excel when given the price and cost functions. We will not focus on these components of business viability because a bayesian network approach is not particularly well suited for factors that are not uncertain. We will instead focus on the components that affect the business viability risk but that are harder to measure, evaluate and are inherently uncertain.

Looking at Unbabel - which delivers translations - the infrastructure (consider web servers, GPUs for Machine Learning purposes and similar costs) and translator costs are relatively easy to predict as they will be directly proportional to the number of translations delivered. Marketing and sales (customer acquisition) costs, on the other hand, are very hard to predict and will highly change depending on a number of product and market components. It is in the latter type of factors, which can account for a large percentage of the Cost of Good Sold, that a bayesian network approach is very suitable - given how uncertain they are before the launch of a new product.

Cagan [Cagan, 2008] describes the "modern" PMM as the representation of the market for the product team. The author considers the function to be a critical partner when building a viable product. At Unbabel the PMM partners up with the PM to define the positioning, messaging and go-to-market plan of the product. To build this network, a senior employee within the function was interviewed.

A total of 4 interviews of about 1 hour each were run. More time was needed than for the other networks particularly to get to the initial structure of the network - it was more challenging to identify the uncertain factors that are more interesting to be studied using this methodology.

Table 5.4 helps navigate the iterative process that was used to generate the Viability Network.

Table 5.4: Viability Risk Network Versions

Version	Expert Interview	Summary
1.0	First and Second	First network structure
2.0	Third and Fourth	Significant Changes to Network Structure
2.1	Third and Fourth	Addition of CPT

5.5.1 Version 1.0

Similarly to what was done for the other networks, the process started with a brainstorming exercise on all factors that affect the viability of a product. While the expert explained the different components the interviewer shared the screen and noted down all the ideas in the virtual white board.

The expert started by noting that price should be aligned with the business model, proportional to the persistent pain of the customer and competitive when compared to alternative solutions. He then moved to highlighting the positioning as a key factor in marketing a product while identifying a number of factors related to it - such as the messaging and the coherence with the product offering. The interviewee explained that a new product which is not consistent with the current offering would require a higher investment to market and sell. From this concept the cost of sales and marketing became relevant and the expert enumerated several factors that could impact the cost of acquiring the customer such as the differentiation of the product and the repeat-ability of the sales process. Behind the customer acquisition cost are not only the attributes of the product (such as differentiation) but also the effectiveness of the marketing organisation - the expert explained that a better target audience understanding and go-to-market preparation would help achieve a lower customer acquisition cost.

Through the explanation of the different components the interviewee mentioned the “four marketing P’s” which are: Place, Promotion, Price and Product.

The first attempt at building a network generated from the brainstorming generated a very complex structure so the process was continued in a second interview with the expert. This second interview was conducted after the other networks were generated so it was run in a more efficient way and the interviewer made sure to nudge the expert into providing all the definitions as the network was built. Figure 5.23 provides the resulting network, in green are some notes from the discussion.

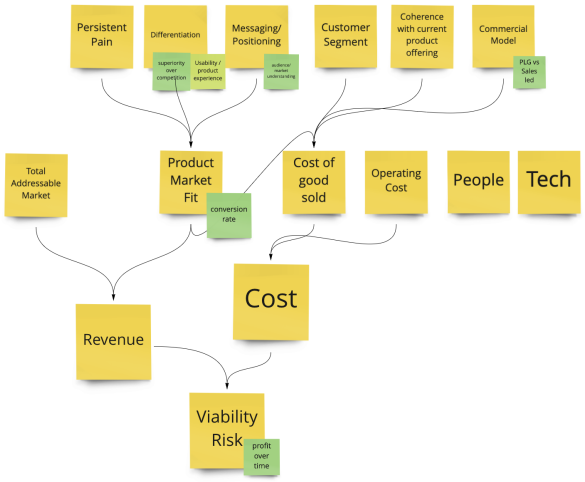

Figure 5.23: Viability Network Version 1.0
Source: Second Interview with the Expert - Screenshot from Miro

Business Viability implies making profit over time in a sustainable way. The key components of profit are revenue and cost - which immediately appear as the first level parent nodes. The revenue then ends up being proportional to the size of the market (Total Addressable Market) times the conversion rate (what % of the Total Addressable market is going to choose to buy this specific Product).

The expert then identified Product Market Fit as essential to any product, first at the same level as revenue and cost but after some discussion it was realized that it was actually a contributing factor to both revenue (as a good proxy for conversion ratio) and cost. The PMM then explained that the main contributing factors to achieving product market fit were:

- Addressing an identified persistent pain
- Solving the pain in a way that is novel or superior to competition - driving differentiation
- There is a well crafted message targeted at a well defined audience/persona - having strong messaging and positioning

A product that has achieved this state should be "flying of the shelves" because the right audience knows it exists and solves their problem. It is also a contributing factor to cost, particularly to sales and marketing. There is a higher efficiency when selling products that address the right pain, audience and use the correct messaging because the same profit can be achieved with lower costs. A smaller investment in sales employees or marketing campaign will achieve the same business generation.

The expert then explained that besides the Product the key components that impacted the viability of acquiring customers were the nature of the segment - exemplifying that selling to government requires very specific contacts and formal processes which can be more expensive - and the commercial model. During the chapter on value 5.4 the concept of PLG was explored - and it is a good example of a model that is best fit for scale and that requires less investment.

As a closing remark the interviewee noted that the team and the tools they use ("People and Tech") are important components to the viability - because very competent people with the right software could be more efficient in selling the same product, under the same commercial model to the same customers.

5.5.2 Version 2.0

While Version 1.0 of the network already provides a highly useful model to explain Product Viability, it presents a few challenges:

- Separating Cost and Revenue as different nodes might not be the best way to represent a Viability Network - all components that go into making a product viable will be a combination of a cost and revenue that they generate. As such, we can assume that for each factor we are already looking at the efficiency (profit) instead of duplicating all factors between the "revenue branch" and the "cost

branch" of the three. Rather than looking at the costs of acquiring the customers and then having to model the revenue that the customer acquisition generates - we can just represent the customer acquisition process and evaluate it based on the revenue, cost ratio. In Version 2.0, Revenue and Cost are implicit to all nodes.

- The process to generate CPT for the Cost of good sold node would not be efficient - because it has 4 parent nodes. As noted during the construction of the Value Network - building the CPT considering 16 states is highly challenging for the expert due to the large number of variations. This situation should be avoided every time possible.
- It is not fully clear what are the key cost components (Operation and Acquisition)

In order to address these challenges the researcher iterated a few different versions of the network which were then discussed with the expert. Figure 5.24 is the result of the collaborative effort and, while it keeps the key components from Version 1.0, presents improvements that make it easier to read, understand and use.

Some particularly relevant quotes from the expert were left in the image under the green and orange " virtual sticky notes".

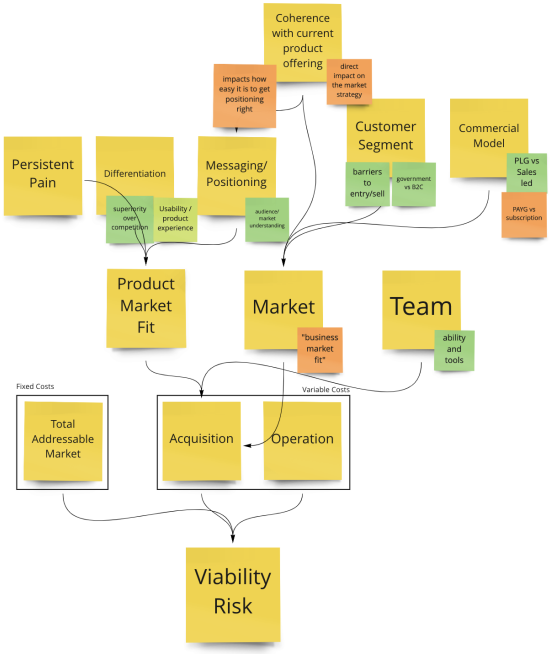

Figure 5.24: Viability Network Version 2.0
Source: Third Interview with the Expert - Screenshot from Miro

They key changes from Version 1.0 to Version 2.0 were:

- Viability was split into the two main processes that impact it - the Acquisition (how expensive it

is to acquire a customer when compared to the lifetime value) and the operation (what are the on-going costs to keeping the customer). It can be noted in the image that the Total Addressable Market is highly relevant to evaluating the Fixed Costs (is the market large enough to justify the initial investment?) while the acquisition and operation are mostly related to variable costs (such as marketing, sales, infrastructure and customer support).

- Revenue and Cost were abstracted from the model. They are now implicit in the Acquisition and Operation Nodes. A high margin Acquisition implies that the costs to generate it are significantly lower than the revenue it will generate during the lifetime of the customer. The same concept applies to Operation.
- The main focus was given to the acquisition process because of its uncertain nature. While the Operating margin can be calculated accurately it is particularly challenging to predict what will be the costs to acquire a customer. Moreover, for software companies it is particularly interesting to study acquisition because its costs represent a very significant portion of the total costs. Acquisition was then separated into three key components that affect its efficiency:
  - The Product - by achieving (or not) Product-Market-Fit
  - The Market - how easy it is to penetrate
  - The Team - which includes the people, their ability and the tools they use
- An unidentified relationship was the impact that coherence with current product offering has on messaging/positioning. It is much easier for a marketing team to message and position a product in an effective way if it is aligned with the current brand and offering from the company.

5.5.3 Version 2.1

Following the same process that was used for the previous networks the expert was asked to further define each node and state before providing probability values. Throughout the session the interviewer and interviewee tried to identify products that would fall under each scenario to help elicit the probabilities. Bellow are the definitions for all nodes:

- Product Market Fit - As becomes clear through the parent nodes, the expert defines that a Product has achieved Market Fit if it uniquely meets the needs of a specific target audience by solving a strong and persistent pain (which relates to Value), and is made available in a clear and accessible way (more unique to viability).
  - The expert used Miro `miro.com` as an example of a product that solves the strong and persistent pain that is collaborating online in an engaging way. (which was highly augmented by the pandemic)

  - Apple products apple.com are known for their ease of use and design when compared to competitors - which drives differentiation
  - Gong https://www.gong.io, a tool to analyse customers interactions and identify risks and opportunities, has a very clear messaging and positioning because they are highly focused on the right key audience (Sales Leadership and Management) and clearly communicate the product value to this audience.

- Coherence with Current Offering - A product is highly coherent with the current offering when it targets the same audience, within the same industry and brand. An example would be to have Unbabel release a new translation product within the Customer Support industry.

- Market - This node defines how challenging a given market is going to be for a product. Launches with "similar value" can have harder penetration into the market because of different characteristics. Being coherent with the current offering is going to help because the company already has the knowledge around how to reach and sell to the specific audience or industry. It should not be confused with the node "addressable market" which models only the size of the market (not its characteristics or how hard it is to penetrate).
  - Customer Segment - A specific segment is going to be harder to reach depending on the seniority and industry of the user and buyer "persona". As an example - it is much more challenging to reach a Chief Executive Officer (CEO), who has limited time, attention spawn and is usually "protected" by layers of managers and assistants than to reach a mid level contributor. Looking at industries, government, for example, requires connections with very specific entities and very formal and complex acquisition process.
  - Commercial Model - Some products such as Salesforce salesforce.com and Oracle oracle.com, due to their complexity, might require senior (and expensive) sales executives to be sold. This makes the user acquisition process much more expensive. Other products "target" more technical users (individual contributors) with free or low cost versions that require no approval, and then count on these users to "champion" the acquisition of the full premium product - which is a product-led approach. Miro miro.com is an example of a company following such an approach. This is the same concept that is explored in the Value Network 5.4 under the "spreading strategy". When there is a product-led commercial approach - a PLG strategy is being used.

- Team - A strong team is aligned within a specific goal and all resources and tools are used in the most efficient way possible. There is no redundancy and people are experienced in both their own domain and in the overall industry.

- Acquisition ROI - A product with a good ROI on acquisition will have a much **CLV!** (**CLV!**) (total worth to a business of a customer over the entire period of their relationship) than the total cost to acquire the customer (including marketing and sales). The positive extreme is a product that is so easy to sell that almost no hours need to be spent on having a Sales Team approaching prospects and it spreads organically (with minimal investment).
- Operation - Relates to the margin the business gets on operating the product. Using Unbabel as an example, it will include the cost to have humans translate the content and have a dedicated success and support team for the customer.
- Addressable Market - Refers to the total market demand for the product. It is the maximum amount of revenue the business can possibility generate by selling the product to a specific market.

The overlap between the definition of Product Market Fit and Product Value 5.4 is significant, particularly in the fact that both identify pain and differentiation as key causal factors. This makes complete sense because it is much easier to drive a customer to acquire a product that clearly provides value than otherwise. From the insights coming from the experts it can be stated that a Product has achieved Market Fit if it provides Value (by solving a strong and persistent pain in a differentiated way) and has a clear messaging and positioning ("is made available in a clear and accessible way").

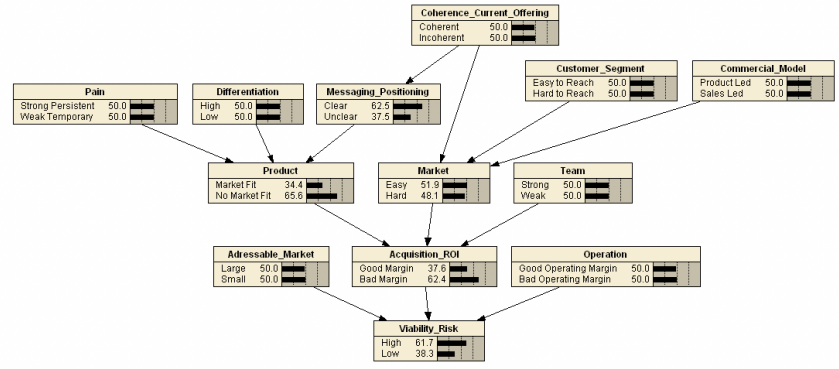

Figure 5.25: Viability Network Version 2.1
Source: Fourth Interview with the Expert - Screenshot from Netica

It should be noted that the optimism of the expert is within the 60/70% range, similarly to what was seen for all networks asides from Feasibility. Value seems to be a highly important causal factor of Viability (which is surfaced by the Product Market Fit node), so it is consistent to have optimism within the same company be within the same order of magnitude.

Starting the sensitivity analysis with the Product node we realize that addressing a strong and persistent Pain is an absolute requirement to achieve Product Market Fit. As can be seen in figure 5.26, any scenario where the Pain is evaluated as weak and temporary the confidence that there is market fit

is always equal or lower to 10%. There is no amount of differentiation or messaging/positioning that can compensate for a product not targeting a large enough pain.

Messaging_Positioning	Differentiation	Pain	Market Fit	No Market Fit
Clear	High	Strong Persistent	85	15
Clear	High	Weak Temporary	10	90
Clear	Low	Strong Persistent	60	40
Clear	Low	Weak Temporary	5	95
Unclear	High	Strong Persistent	65	35
Unclear	High	Weak Temporary	5	95
Unclear	Low	Strong Persistent	30	70
Unclear	Low	Weak Temporary	0	100

Figure 5.26: CPT for the Product Node in the Viability Network
Source: Fourth Interview with the Expert - Screenshot from Netica

This observation is confirmed with the Mutual Info scores for the parent nodes in which Pain is an order of magnitude higher: 0.297 for Pain; 0.022 for Differentiation; 0.017 for Messaging/Positioning.

The parent nodes for Market, on another hand, are quite balanced in level of importance. This is consistent with the evaluations of the expert under which there is no node that is preventive of at least 50% confidence of having an Easy Penetration into the Market. This is, even if Customer Segment is Hard to Reach, if the two other parent nodes are in the positive extreme, it is more likely than not that the Market is easy to penetrate. (the same applies to Coherence with Current Offering and Commercial Model).

```
Sensitivity of 'Market' to a finding at another node:

Node                        Mutual      Percent   Variance of
----                        Info                  Beliefs
Customer_Segment            0.06057     6.06      0.0206641
Coherence_Current_Offeri    0.05037     5.04      0.0172266
Commercial_Model            0.02555     2.56      0.0087891
```

Figure 5.27: Sensitivity to Findings for Market Node
Source: Fourth Interview with the Expert - Screenshot from Netica

Looking at figure 5.28 Coherence with Current Offering is a really good predictor of clear Messaging/Positioning. If there is coherence, there is almost absolute confidence that messaging/positioning will be clear (95%), which is consistent with intuition given the marketing team is familiar with the context. But being incoherent with the current offering does not necessarily mean that the messaging/positioning will be unclear - which is also logical - because the marketing team can adapt to a new situation and context although it is much more challenging.

Evaluating the causal factors for the Acquisition ROI Product stands out as the most important by a large margin, as can be seen in figure 5.29. This is an interesting insight from the PMM that shows

Coherence_Current_Offering	Clear	Unclear
Coherent	95	5
Incoherent	30	70

Figure 5.28: CPT for Messaging/Positioning Node and Coherence with Current Offering
Source: Fourth Interview with the Expert - Screenshot from Netica

how even under a really good team within an “easy market” it is going to be very difficult to sell a weak product efficiently. This evaluation is consistent with intuition, the better a product is at solving an important problem in a unique way, the more efficient both the sales and marketing process will be. The same account executive will enjoy a much higher conversion ratio under the same number of prospects if selling a product that has achieved market fit, so the ROI will be higher.

```
Sensitivity of 'Acquisition_ROI' to a finding at another node:

Node                    Mutual      Percent   Variance of
----                    Info                  Beliefs
Product                 0.17702     18.5      0.0570536
Market                  0.05349     5.6       0.0170486
Team                    0.03182     3.33      0.0102481
```

Figure 5.29: Sensitivity to Findings on the Acquisition ROI Node
Source: Fourth Interview with the Expert - Screenshot from Netica

When looking at the Viability Risk it becomes evident that a small addressable market or weak acquisition ROI are very important risk factors. This is consistent with the scenario that a technology scale-up faces - software companies usually have very low variable costs because there are no costs to replicate the product (the opposite of what happens with physical products in traditional industries). Even when there are costs to scale the number of products - Unbabel is a good example because delivering more translations requires hiring more translators - this is a cost that is much less uncertain because it is easier to measure and predict. Unbabel can calculate how much it spends on translators, serving twice as many customers will mean spending roughly twice as much on translators. But when it comes to the costs to acquire new customers it is much harder to predict if $10 or $1.000 will be needed to lock a sale, so the viability of the process is much more uncertain.

The size of the addressable market has the highest Mutual Info score - technology scale-ups chase exponential growth so a large return on investment is expected on all product decisions - investing in a small addressable market will rarely aligned with the expected strategy even if there is a high profit margin on each sale.

Sensitivity of 'Viability_Risk' to a finding at another node:

Node	Mutual Info	Percent	Variance of Beliefs
Adressable_Market	0.15536	16.2	0.0485597
Acquisition_ROI	0.10786	11.2	0.0352144
Operation	0.05729	5.97	0.0184576

Figure 5.30: Sensitivity to Findings on the Viability Risk Node
Source: Fourth Interview with the Expert - Screenshot from Netica

6

Discussion and Conclusion

6.1 Scenario Analysis and Stress Testing

Scenario analysis is a process through which the outcomes and consequences of possible future conditions are studied. Through it we can evaluate the model predictions under different probable scenarios and compare them against intuition and logic. It is a useful tool to evaluate the resilience and consistency of the model. When doing scenario analysis - two or more input variables will be changed to model

a likely scenario. This is because under “real-world” business situations, the decision maker will usually have knowledge over several factors (although not all) - he might know the typical performance of the team and the complexity of the technology stack that will be used - but not if there are any available libraries with code that can be reused.

Stress testing is a related but different technique. It focuses on assessing the consequences of extreme impact (and low likelihood) events on the model. When performing stress tests - only one input variable will be changed to an “extreme” state. This will allow us to understand the impact of an isolated risk factor on the behaviour of the entire network.

For both methods, the target network nodes will either be changed to one or another state with 100% confidence. **In summary: under scenario analysis, two or more nodes will be changed to model a likely scenario, under stress testing a single node will be changed at a time.**

Figure 6.1 illustrates the stress testing process in Netica for the Feasibility Network under the Weak Team Scenario:

1. Under middle-term uncertainty [1] the probability of the project not being feasible is 39.3%
2. Team Node is changed to 100% probability of being Weak
3. As a consequence, the probability of the project not being feasible increases to 56.7%

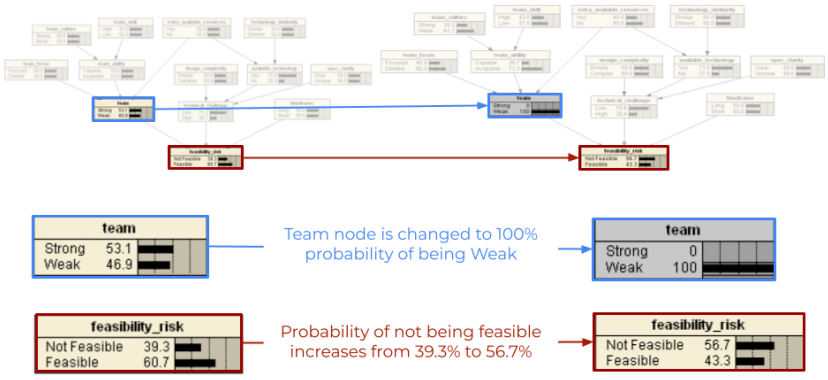

Figure 6.1: Visual Example of stress testing the Feasibility Network by changing the Team Node

It is also useful to introduce a concept which we propose calling *“middle-term uncertainty”*. Middle-term uncertainty happens when all the “parentless nodes” of the network have 50% chance of being in either one state or another. A node has no parents when there are no other nodes “pointing to it” and it means that it is conditionally independent from all other “parentless nodes”. In figure 6.2 we can see that the “parentless nodes” Team Focus and Technology Similarity are conditionally independent - knowing

[1] Middle-Term uncertainty is defined in 6.2

that a team is focused affects in no way the probability of technology being available. Middle-term uncertainty is the "default state" of the network if the decision maker has absolutely no knowledge over the state of any factor. It is useful because it can be used for comparison purposes. When looking at the effect of changing the state of one or more nodes, we can compare how the risk level changes versus the middle-term uncertainty scenario. All networks in chapter 5 were presented under the middle-term uncertainty scenario, examples are figures 5.3; 5.9; 5.22; 5.25.

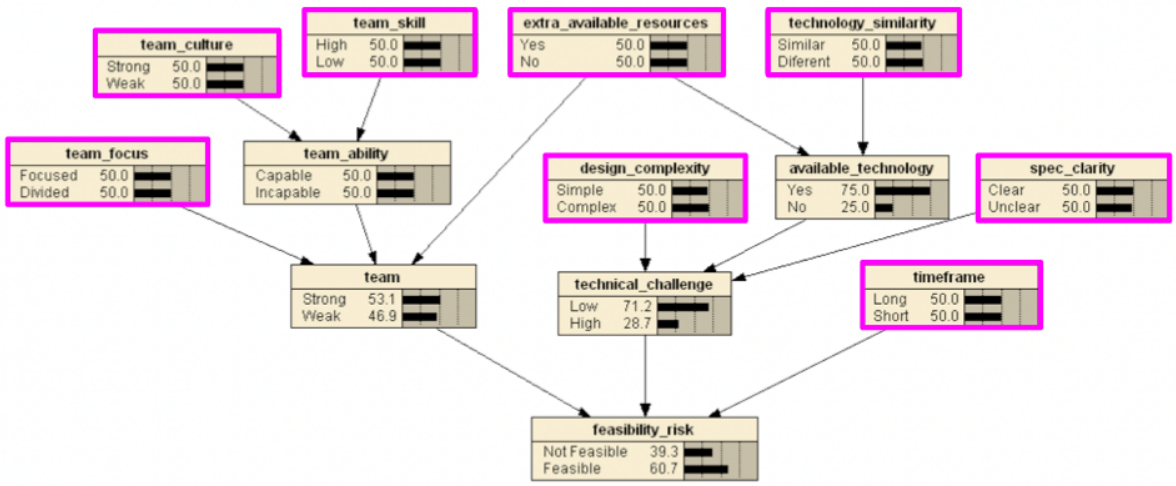

Figure 6.2: Middle-term uncertainty happens when all parentless nodes have 50% probability of being in either state. Notice how this applies to Team Focus, Team Culture, Team Skill, Extra Available Resources, Technology Similarity, Spec Clarity, Design Complexity and Time Frame.

In order to assess the robustness of the models a set of scenario analysis and stress testing will be used for each key risk (Feasibility, Usability, Value, and Viability). For each network a combination of likely and extreme scenarios will be input into Netica and the results will be compared against logical reasoning.

On each iteration of the scenarios different nodes will be changed to evaluate the behaviour of the network. To make sure that results are as meaningful as possible for a limited number of tests, we will prioritise changing the nodes that have the highest impact on the network and/or that are the least found in the literature as measured by their estimated variability [2], sensitivity to findings [3] and "uniqueness" [4].

For each network a summary table will be presented with the scenario description, scenario risk level and risk delta:

- Scenario description contains the indication of which nodes will be changed as input.

[2] Variability refers to how spread out a set of data is. The higher the variability the more important it is to study extreme events as they are more likely to happen

[3] Sensitivity to Findings studies how the uncertainty in the output of a model can be attributed to different inputs of a model. In this project it is measured through the Mutual Info score, defined in chapter 5

[4] "Uniqueness" will be informally used by the author to identify causal factors that are less frequently found in the literature. Discovery Risk, for example, was often mentioned in the interviews but is rarely found in publications

- Scenario risk level is defined as the probability of the Feasibility (Usability/Value/Viability) risk being high under the scenario. It can be thought of as a measure of how likely the product or feature is to fail under that specific area. If a scenario risk level is higher then it means there is a higher risk of it not being Feasible (Usable/Valuable/Viable).
- Risk delta is defined as the difference between the scenario risk level and middle-term uncertainty risk level. This is a measure of how much the inputs will affect the risk of the scenario (compared to the default state of middle-term uncertainty for all parentless nodes). A high risk delta means that the node(s) that were changed have a very large impact on the network. For example the risk delta under table 6.2 for High Discovery Risk is 15.2% while for High User Risk it is only 5.7% - it means that both scenarios have a higher risk level than a fully middle-uncertain scenario but when Discovery risk is high it is even more likely that the user will not understand how to use the product (usability risk).

In the example from figure 6.3 we can see that the middle-term uncertainty risk level is 39.3%, the scenario risk level is 56.7%, and the risk delta is 17.4%.

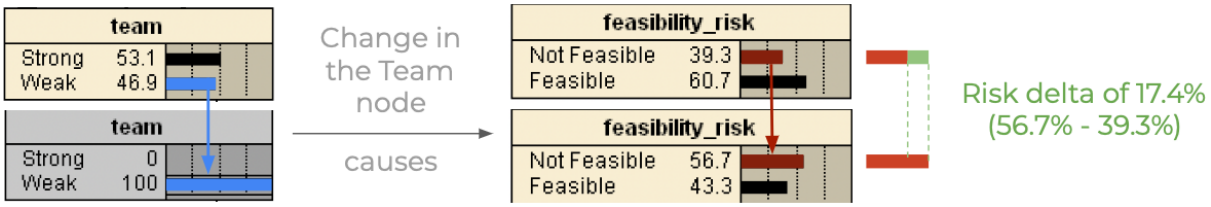

Figure 6.3: Illustration of the scenario risk level (56.7%), middle-term uncertainty risk level (39.3%), and risk delta (17.4%) when Team node was changed to “Weak” state in the feasibility network

6.1.1 Feasibility Network

Under a middle-term uncertainty scenario the feasibility risk level is 32.1%

Table 6.1 provides a summary of the different scenarios, the risk level under them and the risk delta against the middle-term uncertainty risk level (in probability). For each scenario only the nodes that are identified are changed, all other parentless nodes are kept under middle-term uncertainty. We can see in the figure 6.1 example that the Team node is changed to 100% probability of being Weak, but all other parentless nodes are kept in middle-term uncertainty - the “output node” (feasibility risk) changes because it updates the probabilities with the new knowledge.

We can remember from chapter 5.2 that feasibility risk is the risk of the team not being able to build the product/feature.

Table 6.1: Feasibility Risk Scenarios

Scenario (input)		*Feasibility Risk Level*	*Risk Delta*
Node(s)	*State*		
Middle-Term Uncertainty		32.1%	/
Team	100% Weak	48.8%	+16.7%
Technical Challenge	100% High	49.1%	+17%
Time Frame	100% Long	20.2%	-11.9%
Technical Challenge and Team	100% High and 100% Strong	30%	-2.1%
Time Frame and Technology	100% Short and 100% Available	41%	+8.9%

6.1.1.A Stress Test - Weak Team affecting Feasibility

Under the sensitivity to findings analysis the Team node is, by a large margin, the most important causal factor of feasibility risk. As such, we expect it to have a huge impact on the outcome of the project.

Looking at this scenario, the probability of the project not being feasible moves almost to 50% - which means that we expect a project to fail half of the times if the team is weak.

6.1.1.B Stress Test - High Technical Challenge affecting Feasibility

The network behaves in a very similar way between having a Weak Team or a High Technical challenge - the risk delta is of 17% and feasibility risk approaches 50%. We can conclude that for the expert (and the model) having a project with a very high technical challenge and an average team represents around the same risk as having a weak team facing an average challenge.

6.1.1.C Stress Test - Long Time frame affecting Feasibility

During interviews the expert stated that most feasibility challenges can be overcome if there is enough time available to attempt different approaches and research the subject. The model behaves as expected because under a long time frame we expect an “average” product feature to be feasible 4 out of 5 times.

It is consistent with logical reasoning that some product features (1/5) might still not feasible - there will be scenarios where the technology simply does not exist

6.1.1.D Scenario Analysis - High Technical Challenge and Strong Team affecting Feasibility

A strong team presented with a high technical challenge will result in 70% probability of being feasible. This means that a strong team can be trusted to succeed more than half of the times even when presented with the most challenging projects. It is consistent with intuition that senior and focused engineers are able to solve almost any problem they are given.

6.1.1.E Scenario Analysis - Short Time Frame and Available Technology affecting Feasibility

Under a scenario where we are confident that some of the available technology can be used for a project but the time frame for development is short - the feasibility risk is 41%. This is still almost a 10% risk delta when compared to an uncertain scenario. This means that a decision maker cannot drastically shorten the delivery date just because components can be reused from other projects. It makes sense that having time to properly develop for a project is more important than having a library of previous available technology.

6.1.2 Usability Network

Under middle-term uncertainty there is a 71.8% likelihood of the Usability Risk being high.

Table 6.2 provides a summary of the different scenarios, the risk level under them and the risk delta against the middle-term uncertainty risk level.

Table 6.2: Usability Risk Scenarios

Scenario (input)		*Usability Risk Level*	*Risk Delta*
Node(s)	*State*		
Middle-Term Uncertainty		71.8%	/
Discovery Risk	100% High	87.9%	+15.2%
User Risk	100% High	78.4%	+5.7%
Team	100% Incapable	84.4%	+11.7%
Domain and User Risk	100% Complex and 100% High	73.4%	+0.7%
Discovery and User Risk and Team	100% Low and 100% Low and 100% Capable	17.1%	-54.7%

From chapter 5.3 we have that usability risk is the risk of the user not understanding how to use the product.

6.1.2.A Stress Test - High Discovery Risk affecting Usability

When Discovery Risk is high the Usability Risk moves to almost 90% probability of being high. This means that under almost no circumstance should a decision maker invest in a product feature before running comprehensive discovery work. We can then conclude that a very effective way to decrease the risk of a product is to research the user and problem before committing to delivering.

6.1.2.B Stress Test - High User Risk affecting Usability

Under a scenario where User Risk is high the risk delta is only of 5.7%. From this value we can understand that the expert believes the type of user does not have a very significant impact on the risk. This could potentially be explained by the fact that there are now a number of tools that help

designers make products more accessible. (A good example used by Unbabel is EqualWeb https://www.equalweb.com)

6.1.2.C Stress Test - Incapable Team affecting Usability

Under an incapable design team the probability of the customers not knowing how to use the product is 84.4%. This means that we can be almost certain that an inexperienced designer with little domain expertise will not be able to create interfaces that can be easily used and leveraged by the users - and the product experience will be significantly hindered.

6.1.2.D Scenario Analysis - Complex Domain and a Low User Risk affecting Usability

A highly complex domain moves the risk to almost 80%, but having a low user risk shifts the risk back to almost the initial value.

We can then conclude that designing for a highly digital, capable and expert user in a very complex domain represents approximately the same level of challenge as designing under a fully uncertain scenario.

6.1.2.E Scenario Analysis - Low Discovery and User Risk with a Capable Team affecting Usability

This scenario is meant to model a situation where a capable team has invested into doing discovery. More than 80% of the time the customers will not have any problem using a product with average domain complexity. We can then conclude that we can eliminate almost all usability risk if we invest in a good team and if they conduct good discovery work.

6.1.3 Value Network

Under a middle-term uncertainty scenario there is a 69.8% Value Risk level.

Table 6.3 provides a summary of the different scenarios, the probability of the user not choosing to buy/use the product and the risk delta in probability it represents against the middle-term uncertainty scenario.

From chapter 5.4 we have that value risk is the risk of the user not choosing to use/buy the product.

Table 6.3: Value Risk Scenarios

Scenario (input) *Node(s)*	*State*	*Value Risk Level*	*Risk Delta*
Middle-Term Uncertainty		69.8%	/
Demand	100% Low	91%	+21.2%
Competition Level	100% High	74.8%	+5%
Willingness to Change	100% Willing	54.8%	-15%
Reference Customers	100% Easy to Recruit	56.1%	-13.7%
Habit of the Present and Magnetism of the new solution	100% Strong and 100% High	71.8%	+2%

6.1.3.A Stress Test - No Demand affecting Value

Demand is the "most important factor" if measured by the Mutual Info Score [5]. As such, and under any logical reasoning we expect a product to fail if there is no demand in the market - this is confirmed because under this scenario Value Risk is higher than 90%.

6.1.3.B Stress Test - Highly Competitive Market affecting Value

A very common scenario in technology is to have solutions trying to disrupt crowded markets. Uber `uber.com`, Lyft `lyft.com`, Bolt `bolt.eu` are all competing in the "ride hailing" space - demand which is validated by the number of people with a need for fast and convenient transportation.

While the author expected Competition Level to have a very significant impact - there is only a risk delta of 5% under this scenario. While this is not fully consistent with intuition - it might be explained by the expert's optimism and availability bias - his previous experience is in the innovation space where product differentiation is typically high.

6.1.3.C Stress Test - Willingness to Change affecting Value

Conducting Discovery work allows us to assess if there is Willingness to Change from the users. If we are confident there is then value risk moves to almost 50%. This means that just by being confident there is Willingness to Change - we already expect the product to succeed almost half of the time.

6.1.3.D Stress Test - Easy to Recruit Reference Customers affecting Value

One of the least "expensive" data points that can be acquired to scope the value of a product is seeing how easy it is to recruit reference customers[6]. Under the scenario where it was easy to recruit them Value Risk moves to 56.1%, almost the same as if we have confirmed there is willingness to change

[5] Mutual Info (I) - defined as the expected reduction in entropy of node Q due to findings at node F. Consult 5 for a more comprehensive definition

[6] A reference customer is an early "real customer" (with no hidden motivations) that is willing to run a product in production and both make some sort of investment (monetary or in time) and willing to tell others how much they love the product

and significantly better than under a fully uncertain scenario. The conclusion is that investing in finding out the willingness to change of the customers or investing in recruiting reference customers provides almost the same information on the uncertainty - but reference customers are probably a much more efficient (cheaper) approach.

6.1.3.E Scenario Analysis - Strong Habit of the Present and High Magnetism of the New Solution affecting Value

Some market opportunities rely on disrupting highly established products. Slack https://slack.com is an example of a product that solved a large problem (communicating with colleagues) in a different way (flexible chat channels) when most companies had been using email for formal communication for many years.

Under the scenario where there is already a strong habit of the present but the new solution is highly magnetic we see a risk delta of 2% - almost 0. We can then conclude that a deeply rooted product habits can be countered by releasing a strong new product that is highly appealing to the users.

6.1.4 Viability Network

Under middle-term uncertainty there is a 61.7% likelihood of the Viability Risk being high.

Table 6.4 provides a summary of the different scenarios, the viability risk level and the risk delta it represents against middle-term uncertainty.

Table 6.4: Viability Risk Scenarios

Scenario (input) *Node(s)*	*State*	*Value Risk Level*	*Risk Delta*
Middle-Term Uncertainty		61.7%	/
Product	100% Market-Fit	68.4%	+6.7%
Addressable Market	100% Small	83.7%	+22%
Acquisition ROI	100% Good	37.5%	-24.2%
Addressable Market and Operating Margin	100% Small and 100% Good	78.7%	+17%
Addressable Market (size) and Market Characteristics	100% Large and 100% Hard	46.1%	-15.6%

From chapter 5.5 we have that viability risk is the risk of the product “not working for the business”.

6.1.4.A Stress Test - No Product Market Fit affecting Viability

Having no Product Market Fit increases the Viability Risk to almost 70% - this means that the product will not work for the business most of the time and a decision maker should avoid investing in it. While

the risk could be expected to be higher (if the product does not have market fit it will rarely have great adoption)

6.1.4.B Stress Test - Small Addressable Market affecting Viability

When the market is small the product will fail to become viable more than 4 out of 5 times. This judgment is consistent with logical reasoning because software development is typically tied to large fixed costs and small variable costs. This "cost profile" greatly benefit from scale, which is not possible if the addressable market is small.

6.1.4.C Stress Test - Good Acquisition Return on Investment affecting Viability

Under the scenario where acquisition has a great ROI the risk of the product not working for the business goes down very significantly - almost 25%. When this happens it means that the costs of acquisition are significantly smaller than the Lifetime Value of the customer.

While the risk delta might appear higher than intuition (it leads to questioning the impact of the other type of cost - service/operation) the expert explained that acquisition is typically harder to solve (and riskier) because it depends on many uncertain external factors. The operating margin, on the other hand, will usually improve with growth because software costs do not scale proportionally with revenue.

6.1.4.D Scenario Analysis - Small Addressable Market and Good Operating Margin affecting Viability

Having a good operating margin only slightly decreases the risk of targeting a small addressable market. We can then conclude that investing in an average size Addressable Market under an uncertain operating margin is a better investment than having a good operating margin within a known small addressable market.

6.1.4.E Scenario Analysis - Large and Hard to Penetrate Market affecting Viability

A product that addressees a large, even if hard to penetrate, market will "work for the business" more than half of the times. The conclusion is that the size of the market is more important to the Viability Risk than the ease of entry.

6.2 Applications

As stated in the literature review 3 this project addressees a largely unstudied area (modern Product Management) using a novel approach (bayesian networks.) Due to its "early nature" there are still

many improvements that can be done to get further and more interesting results - some suggestions will be explored in the next section of the current chapter. But there are already a number of possible applications for this approach with both academic and business value:

- **Using it to explain how product professionals make decisions**

  The research provides value by shedding a light on how professionals make product related decisions on their day-to-day and how their reasoning relates to modern and widely accepted product development frameworks [Cagan, 2008].

  Understanding how professionals actually make decisions in specific contexts is important because it provides for an opportunity to diagnose and mitigate biases and to build tools that have wider adoption. As explored in the Literature Review chapter 3 startups and scale-ups rarely adopt "formal" and "sophisticated" risk analysis tools - but by definition they are faced with very uncertain scenarios. A better understanding of "product professionals" could be the basis of new decision making software - not just used by business analysts but also by engineers, designers and managers on their day to day.

- **Using it to build customized decision making tools that evolve over time**

  Bayesian Networks have the benefit of being easy to customize and update over time. While this thesis provides networks that model Feasibility, Usability, Value and Viability Risk - these can very easily be provided as templates only that each organization and professional can adapt to their own "model of the world". Another startup can pickup the examples and change some of the nodes because they have a different perception on how to model Value or because their market has specific characteristics that should be surfaced in the model.

  While the Conditional Probabilities were all collected through direct elicitation - which is the fastest way to build a "functioning network" - a company that adopts this method as a tool can start collecting data on each product feature and use it to update the CPT over time. Tools such as Netica can build the probabilities from cases - so if a company makes the investment to rate each product feature on each parent node - and later registers what was the outcome under each area (Did the engineering delivery go smoothly? Did users find it easy to use? Was it widely adopted by the market? Did it generate a sustainable profit?) it can update the logic behind the network.

- **Using it to predict the risk in each area of developing a product (feature) under different teams**

  The output of each network provides a clear numeric prediction on the probability of the product succeeding or failing under each key area. These values can be used to decide on whether to invest or not on a specific product under the risk profile of the decision makers. It also has the potential to aid on making complex decisions such as studying how risk changes depending on the

allocation of each product to each team. A decision maker choosing to allocate product features A, B and C to teams 1, 2 and 3 can simulate the 6 possible combinations (A-1, B-2, C-3 / A-2, B-3, C-1/ ...) using the networks and decide based on the optimal risk level. Different teams will have different risk results for the same features because they have different engineers, designers, marketeers and domain knowledge.

As a note, it should be acknowledged that the predictive value of the bayesian networks generated throughout this project has not been proved. This validation would require following the decision, development and launch several products at Unbabel to assess success under each area. Which is one of important follow-up opportunities for this project.

- **Using it to assess which investments to make in order to reduce uncertainty**
  The Mutual Info score[7] is a great indicator of which knowledge has the most impact on the certainty around risk.

  This score can be used to assess whether it is worth investing or not into acquiring knowledge. If we take the example of the Viability Network - a market research project to understand the size of the market might cost $10.000, as a decision maker I can decide if it is worth the investment based on how much I value knowing the state of the "Addressable Market" node.

- **Using it compare how different product organizations work**
  After conducting the exercise of generating the networks for different organisations we are presented with a number of different variations on how to think about risk and make decisions. Each acylic graphic is a visual, and easily shared, representation of how the organisation is thinking about its product. *Do companies organized in product teams make decisions differently from functional divided companies?*; *Do startups think about risk in a different way from enterprises?* and is that why they are thought to have a higher innovative potential?

  There are a number of questions that can be raised and answered by conducting this exercise across multiple companies and comparing the results.

- **Using it to bridge the gap between academia and business practices**
  As stated in the literature [Pukala et al., 2018] [Crovini et al., 2020], startups and SMEs rarely adopt formal risk management practices that are proposed by the academia. On the other hand, product management publications from popular authors such as Cagan [Cagan, 2008] and Torres [Torres, 2021] have seen wide adoption from the ecosystems (Unbabel is a good example). Having methods that can formalize and crystallize the concepts already used by modern companies into structured frameworks provides for an opportunity to bring both communities (academia and startups) closer.

[7] Mutual Info (I) - defined as the expected reduction in entropy of node Q due to findings at node F. Consult 5 for more

6.3 Conclusion

Expert interview is a viable method to generate bayesian networks and create predictive models for complex and unstructured scenarios. With an investment of a few hours it is possible for a prepared interviewer to guide an expert into creating a visual representation of all the factors that go into evaluating product risk. Through direct elicitation the researcher can then complete the Conditional Probability Tables to generate a model capable of making predictions consistent with the expert opinion.

Both under qualitative and quantitative evaluation methods the networks behaved consistently with logical reasoning. The causal relationships described in the acylic graphics were reviewed in interviews with experts and the behaviour of the networks was submitted to sensitivity analysis, stress testing, and scenario analysis.

While there are risk factors that were expected to have an impact on product development, such as team and competition level, the experts also identified new important concepts that would benefit from further study like “Discovery” and “User Risk”.

These models are useful for research and business alike because of their multiple applications. They provide for a visual representation of a mental model around product development risk but the underlying mathematical model built on top of the CPT can be used for more advanced and complex applications. By leveraging the features of software such as Netica it is possible to input different scenarios into the model to see how they affect the risk of the product failing.

An Engineering Manager can adjust the conditional probability tables to fit his experience and then use the resulting model to test how Feasibility Risk would change depending on which team he assigned a feature that needs to be developed; A Director of Design could quantitatively justify how targeting a different user could impact the risk of the product being “hard to use”; A Vice President of Innovation could choose to develop one product instead of another because there is a higher confidence that the first would provide more user value - these are all examples of business applications for the models.

Finally, the method was successful in formalizing the “modern concept” of the 4 main product risks: Feasibility, Usability, Value and, Viability, found in Inspired [Cagan, 2008] and used in the industry. This is a first step in bridging the gap between business and academia when it comes to technology product development.

6.4 Future Work and Research Opportunities

This section will address some of the challenges and limitations of the project - which was conducted under specific time and resource constraints - mostly as opportunities for follow-up work. Through the project many interesting ideas surfaced but not all of them were explored - so there are multiple paths for anyone interested in the area to continue the research.

- **Building a "universal" product risk bayesian network**

  Product risk was divided into 4 key areas for a few different reasons. First of all, there are restrictions on the Netica version that was used which restrict the network to a limited number of nodes. Secondly, each network was elicited to a different expert who is specialized in the area. Merging all of the networks would require a significant time investment from the experts to find consensus on the network and on the overlapping concepts. It would also be challenging to get to accurate conditional probabilities to combine Feasibility, Usability, Value and Viability Risk. Each expert is usually biased on the importance of their own area [8].

  A project that overcomes these limitations can present a "unified network" - a single visual model that contains all product risk factors and how they relate to each other. It would be possible to assess how impactful each type of risk is and for each professional to update knowledge with case findings to provide a single "risk score" for the product.

- **Experimenting with different elicitation methods**

  Direct expert elicitation was used because it is the fastest method to generate a CPT. But there are many disadvantages to this method - direct estimation is a very hard and error prone exercise for experts. Biases such as anchoring bias [9] can very easily impact the results, and the exercise "explodes" in complexity as the number of parent nodes increases. This limitation was very visible when the interviewer was eliciting probabilities for nodes with 4 parent nodes (there are 16 possible scenarios) - the experts struggled to provide assessments. The number of scenarios is not only proportional to the number of parent nodes but also to the number of possible states for each node - all had to be simplified to two extreme states (A team is either weak or strong with a given probability).

  There are a number of different elicitation methods that can be explored and that will, potentially, generate much more accurate and consistent results. Advanced methods are able to find inconsistent judgements which can be reviewed with the experts Monti for example, modified the original elicitation method used in the AHP process [Saaty, 1987] so that it could be used in the context of probabilities [Monti and Carenini, 2000]. A great example is the MACBETH software from Bana e Costa. et al [e Costa et al., 2012] - in an integrated way it helps structure the problem hierarchical; elicit judgments in an intuitive way (through pairwise comparisons); rank each factor; detect inconsistent judgments and choose between different options. While the bayesian networks do not necessarily follow the hierarchical structure that is needed for this method - the trade-of between

[8] Motivational Bias - Happens when there is a (usually hidden) incentive for the expert to provide a given assessment. It happens when the individual doing the estimation believes that the expressed judgements may affect them personally

[9] Anchoring Bias - The estimation of a probability is highly impacted (anchored) on the first assessment that is made. Providing a random number as an example during the interview can influence the interviewee to provide judgments within the same order of magnitude - even when there is no logical reason to do so. This is a very significant bias particularly related to how the researcher conducts the interview.

the flexibility of networks and end-to-end ease of use could be studied.

- **Further exploring "symptom" nodes**

  When using bayesian networks under a medical (diagnostic) context [Yet et al., 2014] symptoms are very frequently modeled as nodes in the network. Their existence (or not) is a great predictor of an underlying disease so the state of the nodes updates the knowledge of the network.

  The same concept can be applied to some areas of product development - in this thesis we explored how the existence of Reference Customers [10] can be considered a "symptom" of the existence of Demand and Willingness to Change by users. Only one of these nodes was identified between the 4 networks - but that is mostly because the interview format and questions did not prone the experts to focus on this type of nodes.

  Further research can focus on finding more "symptom" nodes which can be used to update the findings of the network. It is usually much more efficient to find information on the state of these type of nodes than in the risk factors. If we take reference customers and demand - it is hard and expensive to define the existence of demand directly - but it is cheap and fast to try to recruit reference customers from a mailing list. We suggest that it is valuable for businesses to find these "symptom nodes" which can update our beliefs on the risk of the product. So it is worth it conducting further research in the area.

- **Conditional Probability Generation through historical data**

  The Netica software has the capability to learn from case data `https://www.norsys.com/WebHelp/NETICA/X_Learning_from_Cases.htm`. When given a data set of enough historical findings with states for each node the software is capable of inferring a direction of causation (structure learning) and the conditional probability tables (parameter learning).

  An organization that has this data in the intended format can then automatically build the network (assuming the nodes are already identified in the data structure) and only fine tune-it with help from experts. The company could then find new causality relationships through the data. This process is also efficient to, over time, continuously update the Conditional Probability Tables (CPTs) and possibly increase the predictive accuracy of the network.

- **Building the acyclic graphs through group brainstorming exercises**

  Each network was generated through a series of interviews with a single subject matter expert. This means that the results are exposed to all the inconsistencies and biases of those specific individuals. The same exercise can be done with a group of subject matter experts.

[10] A reference customer is an early "real customer" (with no hidden motivations) that is willing to run a product in production and both make some sort of investment (monetary or in time) and willing to tell others how much they love the product. Note that within all the networks this is the only "symptom" node

References

[Akhavan et al., 2021] Akhavan, M., Sebt, M. V., and Ameli, M. (2021). Risk assessment modeling for knowledge based and startup projects based on feasibility studies: A bayesian network approach. *Knowledge-Based Systems*, 222:106992.

[Autor, 2022] Autor, U. P. (2022). Value proposition.

[Bevan et al., 2016] Bevan, A., Lewis, J., Bevans, T., and Smes, M. (2016). Why British businesses don ' t scale up.

[BGI and EIT Digital, 2019] BGI and EIT Digital (2019). Scaleup Portugal. (June).

[Bielen and Demoulin, 2007] Bielen, F. and Demoulin, N. (2007). Waiting time influence on the satisfaction-loyalty relationship in services. *Managing Service Quality*, 17(2):174–193.

[Bromiley et al., 2015] Bromiley, P., McShane, M., Nair, A., and Rustambekov, E. (2015). Enterprise Risk Management: Review, Critique, and Research Directions. *Long Range Planning*, 48(4):265–276.

[Brown and Mawson, 2013] Brown, R. and Mawson, S. (2013). Trigger points and high-growth firms: A conceptualisation and review of public policy implications. *Journal of Small Business and Enterprise Development*, 20(2):279–295.

[BSI Standards Publication, 2018] BSI Standards Publication (2018). BS ISO 31000 : 2018 BSI Standards Publication Risk management — Guidelines. *BSI Standards Publication*, ISO31000:26.

[Cagan, 2008] Cagan, M. (2008). *Inspired: How to Create Tech Products Customers Love*. John Wiley Sons, 2 edition.

[Cagan, 2017] Cagan, M. (2017). The four big risks.

[Cantamessa et al., 2018] Cantamessa, M., Gatteschi, V., Perboli, G., and Rosano, M. (2018). Startups' roads to failure. *Sustainability (Switzerland)*, 10(7):1–19.

[Chin et al., 2009] Chin, K. S., Tang, D. W., Yang, J. B., Wong, S. Y., and Wang, H. (2009). Assessing new product development project risk by bayesian network with a systematic probability generation methodology. *Expert Systems with Applications*, 36:9879–9890.

[Christensen et al., 2016] Christensen, C. M., Dillon, K., Hall, T., and Duncan, D. S. (2016). *Competing Against Luck: The Story of Innovation and Customer Choice*. Harper Business.

[Collective, 2022] Collective, P. G. (2022). What is product-led growth?

[Cooper, 2000] Cooper, L. G. (2000). Strategic marketing planning for radically new products. *Journal of Marketing*, 64:1–16.

[Creswell and Creswell,] Creswell, J. W. and Creswell, J. D. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*.

[Crovini et al., 2020] Crovini, C., Ossola, G., and Britzelmaier, B. (2020). How to reconsider risk management in SMEs? An Advanced, Reasoned and Organised Literature Review. *European Management Journal*.

[Davis, 2002] Davis, C. R. (2002). Calculated Risk: A Framework for Evaluating Product Development.

[de Souza Savian et al., 2018] de Souza Savian, F., Bisognin Garlet, T., and Mairesse Siluk, J. C. (2018). GESTÃO DE RISCOS ORGANIZACIONAIS EM EMPRESAS DE BASE TECNOLÓGICA: UMA ANÁLISE BIBLIOMÉTRICA. *Iberoamerican Journal of Industrial Engineering*, 10(19):144–156.

[Deloitte, 2013] Deloitte (2013). The Risk Intelligent technology company Managing risk to capture value. *Risk Intelligence Series*.

[Dimensional Research, 2013] Dimensional Research (2013). Customer Service and Business Results. *Dimensional Research*, (April):1–11.

[Diogo Ferreira Nunes, 2020] Diogo Ferreira Nunes (2020). Unbabel despede quase 100 pessoas em reestruturação interna.

[Dubey and Wagle, 2007] Dubey, A. and Wagle, D. (2007). Delivering software as a service. *TheMcKinseyQuarterly*.

[e Costa et al., 2012] e Costa, C. A. B., Corte, J.-M. D., and Vansnick, J.-C. (2012). Macbeth.

[e Costa et al., 2021] e Costa, C. B., Oliveira, M., and Mascarenhas, E. (2021). Why do we need decision analysis? (exploring heuristics and biases).

[Eisenhardt, 1989] Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review*, 14:532–550.

[Eisenmann, 2021] Eisenmann, T. (2021). Why start-ups fail. *Harvard Business Review*.

[Employee, 2021] Employee, U. A. (2021). What is product value?

[Ferreira de Araújo Lima et al., 2020] Ferreira de Araújo Lima, P., Crema, M., and Verbano, C. (2020). Risk management in SMEs: A systematic literature review and future directions. *European Management Journal*, 38(1):78–94.

[Gjerdum, 2015] Gjerdum, D. (2015). Risk Management ' s Standard of Practice –. (June).

[Grace et al., 2015] Grace, M. F., Leverty, J. T., Phillips, R. D., and Shimpi, P. (2015). The Value of Investing in Enterprise Risk Management. *Journal of Risk & Insurance*, 82(2):289–316.

[Hölzl, 2009] Hölzl, W. (2009). Is the R&D behaviour of fast-growing SMEs different? Evidence from CIS III data for 16 countries. *Small Business Economics*, 33(1):59–75.

[Institute of Risk Management, 2010] Institute of Risk Management (2010). A structured approach to Enterprise Risk Management (ERM) and the requirements of ISO 31000 Contents. *Risk Management*, 7(1):20.

[Kim and Vonortas, 2014] Kim, Y. and Vonortas, N. S. (2014). Managing risk in the formative years: Evidence from young enterprises in Europe. *Technovation*, 34(8):454–465.

[Lavie, 2020] Lavie, A. (2020). Why Facebook and Google's New Multilingual Models are Great Steps for AI.

[Miller and Friesen, 1984] Miller, D. and Friesen, P. H. (1984). A LONGITUDINAL STUDY OF THE CORPORATE LIFE CYCLE. *Management Science*, 30(10):1161–1183.

[Monteiro, 2019] Monteiro, G. F. A. (2019). High-growth firms and scale-ups: a review and research agenda. *RAUSP Management Journal*, 54(1):96–111.

[Monti and Carenini, 2000] Monti, S. and Carenini, G. (2000). Dealing with the expert inconsistency in probability elicitation. *IEEE Transactions on Knowledge and Data Engineering*, 12:499–508.

[Mosier et al., 2000] Mosier, S. P., Guenterberg, S. A., and Raphael, R. R. (2000). The relationship of technology change management to risk management.

[Myers, 2009] Myers, D. M. (2009). *Qualitative research in business management*. Sage, Los Angeles, 1. publ. edition.

[Myers, 1997] Myers, M. (1997). Qualitative Research in Information Systems. *MIS Quarterly*, 21.

[Nadali et al., 2018] Nadali, A., Grilo, A., and Zutshi, A. (2018). A conceptual framework of risk identification for scale up companies in transition period. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 2018-March:2346–2357.

[Nadkarni and Shenoy, 2001] Nadkarni, S. and Shenoy, P. P. (2001). A bayesian network approach to making inferences in causal maps. *Eur. J. Oper. Res.*, 128:479–498.

[Neil et al., 2009] Neil, M., Häger, D., and Andersen, L. (2009). Modeling operational risk in financial institutions using hybrid dynamic bayesian networks. *The Journal of Operational Risk*, 4:3–33.

[Palan, 2021] Palan, H. (2021). Do you really understand the concept of product value?

[Pedro Marins Freire Teberga, 2018] Pedro Marins Freire Teberga, F. L. O. (2018). Identification, Analysis and Treatment of Risks in the Introduction of New Technologies by Start-ups. *Emerald Publishing Limited*.

[Phillips, 2007] Phillips, S. (2007). A brief history of facebook.

[Picken, 2017] Picken, J. C. (2017). From startup to scalable enterprise: Laying the foundation. *BUSINESS HORIZONS*, 60(5):587–595.

[Pukala et al., 2018] Pukala, R., Sira, E., and Vavrek, R. (2018). Risk management and financing among Start-ups. *Marketing and Management of Innovations*, (3):153–161.

[PwC, 2017] PwC (2017). Enterprise Risk Management. Integrating with strategy and performance. *The Committee of Sponsoring Organizations of the Treadway Commission*, (June):16.

[Ryu et al., 2016] Ryu, C. H., Lim, S. Y., and Suh, M. (2016). PROJECT RISK MANAGEMENT IN R&D ORGANIZATIONS: A SURVEY ON RISK REGISTER FROM KOREAN COMPANIES. *Journal of Modern Project Management*, 4(2):10–23.

[Saaty, 1987] Saaty, R. W. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical Modelling*, 9:161–176.

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423.

[Spiek, 2012] Spiek, C. (2012). Unpacking the progress making forces diagram.

[Torres, 2021] Torres, T. (2021). *Continuous Discovery Habits: Discover Products that Create Customer Value and Business Value*. Product Talk LLC, 1 edition.

[Unbabel,] Unbabel. https://unbabel.com.

[van Dijk et al., 2015] van Dijk, M., Kruit, J. D., Mogendorf, G., and Scheper, W. (2015). Scale-Up: The Experience Game. *Startup Scaling Research Project* -, page 11.

[Verbano and Venturini, 2013] Verbano, C. and Venturini, K. (2013). Managing Risks in SMEs: A Literature Review and Research Agenda. *Journal of Technology Management & Innovation*, 8(3):186–197.

[Villa Todeschini et al., 2017] Villa Todeschini, B., Sosa Boelter, A., Siqueira, J., Souza, D. E., and Cortimiglia, M. N. (2017). Risk Management from the Perspective of Startups system. *European Journal of Applied Business Management*, 3(3):40–54.

[Yet et al., 2014] Yet, B., Perkins, Z. B., Rasmussen, T. E., Tai, N. R., and Marsh, D. W. R. (2014). Combining data and meta-analysis to build bayesian networks for clinical decision support. *Journal of Biomedical Informatics*, 52:373–385.

[Yin, 1981] Yin, R. K. (1981). The Case Study as a Serious Research Strategy. *Science Communication*, 3(1):97–114.

[Yin and SAGE., 2003] Yin, R. K. and SAGE. (2003). *Case Study Research: Design and Methods*. SAGE Publications.