New firm gestation: an empirical analysis of entrepreneurial behavior, start-up activities and the use of machine learning methods to predict the gestation outcome

D I S S E R T A T I O N
of the University of St. Gallen
School of Management,
Economics, Law, Social Sciences
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Management

submitted by

Paris Koumbarakis

from

Egg (Zürich)

Approved on the application of

Prof. Dr. Urs Fueglistaller

and

Prof. Dr. Thierry Volery

Dissertation no. 5056

Difo-Druck GmbH, Untersiemau 2021

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, October 23, 2020

The President:

Prof. Dr. Bernhard Ehrenzeller

Acknowledgment

This dissertation would not have been possible without the help of many wonderful people in my life who always supported me. I would like to mention some of them in particular.

First, I would like to thank Urs Fueglistaller and Thierry Volery for their support and providing the possibility of conducting the research for this dissertation under their guidance. Especially, I owe great thanks to Thierry Volery, whose supervision, expertise and persistent feedback were indispensable to the successful completion of my dissertation. Thank you for guiding and supporting me on this journey.

Furthermore, I'm deeply grateful to Heiko Bergmann whose insightful comments, recommendations, enthusiasm and generous support were of tremendous value to me. Thank you very much for the guidance, all the discussion sessions, the constructive feedback and the encouragement. Over and above, I want to thank the whole Swiss Research Institute of Small Business and Entrepreneurship for providing valuable feedback during the various workshops.

Finally, I owe my deepest gratitude to my family who continuously encouraged me over the course of my doctoral thesis. I am truly grateful for their support throughout my life. This dissertation is dedicated to my parents.

St.Gallen, June 2020 Paris Koumbarakis

Table of contents

ACKNOWLEDGMENT ..... **I**
TABLE OF CONTENTS ..... **II**
LIST OF FIGURES ..... **VI**
LIST OF TABLES ..... **VII**
LIST OF ABBREVIATIONS ..... **VIII**
EXECUTIVE SUMMARY ..... **X**
ZUSAMMENFASSUNG ..... **XI**

1. INTRODUCTION **1**
1.1 OVERARCHING TOPIC AND RESEARCH AIM 1
1.2 DEFINITION OF MAIN THEORETICAL CONCEPTS 2
1.2.1 THE ENTREPRENEURIAL PROCESS, NASCENT ENTREPRENEUR(SHIP) AND THE VENTURE OUTCOME 2
1.2.2 START-UP ACTIVITIES AND THE DISCOVERY–EXPLOITATION FRAMEWORK 4
1.2.3 ENTREPRENEURIAL BEHAVIOR AND COGNITION 6
1.2.3.1 SELF-REGULATION - REGULATORY FOCUS THEORY 6
1.2.3.2 SOCIAL IDENTITY THEORY 7
1.2.4 SUPERVISED LEARNING AND ARTIFICIAL NEURAL NETWORKS 7

2. OVERVIEW OF THE CUMULATIVE DISSERTATION **11**
2.1 KNOWLEDGE AND RESEARCH GAPS 11
2.2 PROCEDURE AND RESEARCH METHODOLOGY 13
2.3 STRUCTURE AND MAIN CHARACTERISTICS OF THE THREE PAPERS 14

3. THE IMPACT OF FOUNDER'S REGULATORY FOCUS ON FIRM BIRTH AND FIRM ABANDONMENT **17**
3.1 ABSTRACT 17
3.2 INTRODUCTION 17
3.3 THEORY AND HYPOTHESES DEVELOPMENT 19
3.3.1 DISCOVERY AND EXPLOITATION FRAMEWORK AND THE VENTURE OUTCOME 19

3.3.2 Regulatory focus, exploitation activities and its relation to firm birth 20
3.3.3 Regulatory focus, exploitation activities and its relation to firm abandonment 22
3.4 Research methodology 24
3.4.1 Data and sampling procedure 24
3.4.2 Common method bias 25
3.5 Measures 25
3.5.1 Dependent variables 25
3.5.2 Independent variable 26
3.5.3 Moderator variable 26
3.5.4 Control variables 27
3.6 Analysis and results 27
3.6.1 Descriptive statistics and correlations 27
3.6.2 Hypotheses testing results 27
3.7 Discussion and contribution 33
3.7.1 Practical implications 35
3.7.2 Limitations and future research 36
3.8 Conclusion 37

4. Imitate to emerge – who is not imitating? 38
4.1 Abstract 38
4.2 Introduction 38
4.3 Theoretical background and hypotheses development 40
4.3.1 Innovation and imitation in the firm creation process 40
4.3.2 Social identity theory and imitation / innovation 42
4.4 Methodology 45
4.4.1 Sample and data collection 45
4.4.2 Common method bias 46
4.5 Measures 46
4.5.1 Firm birth 46
4.5.2 Imitation / innovation 47
4.5.3 Social identity theory 48
4.5.4 Control variables 49

4.6 Analysis and results 50
4.6.1 Descriptive statistics 50
4.6.2 Hypothesis testing 50
4.7 Discussion and conclusion 54
4.7.1 Summary and discussion 54
4.7.2 Theoretical and practical implications 55
4.7.3 Limitations and future research 56

5. Predicting the venture gestation outcome with machine learning methods and artificial neural networks 57
5.1 Abstract 57
5.2 Introduction 57
5.3 Related work 59
5.3.1 Meta-analysis of the empirical literature on factors influencing the venture outcome during the gestation phase 59
5.3.2 Literature review of empirical entrepreneurial success and failure prediction models 62
5.4 Data and methodology 64
5.4.1 Dataset and variables 64
5.4.2 General analytical approach 68
5.5 Classification methods 72
5.5.1 K-nearest neighbors algorithm 72
5.5.2 Decision tree 73
5.5.3 Random forest 73
5.5.4 XGBoost 73
5.5.5 Support vector machine 74
5.5.6 Artificial neural networks 74
5.6 Results and discussion 75
5.7 Practical implications and limitations 88
5.8 Conclusions and future work 89
5.9 Appendix 91

6. CONCLUDING CHAPTER 98
6.1 CONTRIBUTIONS 98
6.2 LIMITATIONS AND FUTURE RESEARCH 100
6.3 CONCLUSION 101

REFERENCES 102
CURRICULUM VITAE 127

List of figures

Figure 1. Graphical representation of the start-up process 3
Figure 2. Model of an artificial neuron 8
Figure 3. Graphical representation of a neural network using start-up activities 10
Figure 4. Structure of the dissertation 15
Figure 5. Hypothesized model 24
Figure 6. Interaction effect on firm birth (compared to ongoing gestation) 32
Figure 7. Interaction effect on firm abandonment (compared to ongoing gestation) 32
Figure 8. Outcome status by month 65
Figure 9. Schematic outline of the analytical approach 70
Figure 10. Feature importance firm birth 87
Figure 11. Feature importance firm abandonment 87

List of tables

Table 1. Discovery-exploitation framework 5
Table 2. Main characteristics of the three papers 16
Table 3. Descriptive statistics and correlations 29
Table 4. Results logistic regression 31
Table 5. Start-up status 47
Table 6. SwissPEB survey questions on innovation and imitation 48
Table 7. Founder identity scale 49
Table 8. Descriptive statistics and correlations 51
Table 9. Results binary logistic regression 53
Table 10. List of predictor variables 66
Table 11. Performance evaluation of classification models on test dataset 77
Table 12. Previous empirical research on venture outcome 91

List of abbreviations

ADASYN	adaptive synthetic sampling approach
ANN	artificial neural networks
β	correlation coefficient
CART	classification and regression trees
CAUSEE	Comprehensive Australian Study of Entrepreneurial Emergence
CEO	chief executive officer
CFA	confirmatory factor analysis
cf.	confer (compare)
CV	cross validation
df	degrees of freedom
DT	decision tree
EC	entrepreneurial cognition
EDN	edited nearest neighbors
e.g.	exempli gratia (for example)
et al.	et alii (and others)
etc.	et cetera (and so on)
FN	false negative
FP	false positive
i.a.	inter alia (among others)
i.c.	in casu (in this case)
i.e.	id est (that is)
KNN	k-nearest neighbors
p	probability (level of significance)
p.	page
N	number of observations

MLP	multilayer perceptron
NCR	neighborhood cleaning rule
n.s.	not significant
OR	odds ratio
PSED	Panel Study of Entrepreneurial Dynamics
RF	random forest
RFE	recursive feature elimination
RFT	regulatory focus theory
RMSEA	root mean square error of approximation
ROS	random over-sampling
RUS	random under-sampling
SB	Satorra-Bentler scaled chi-squared test
SD	standard deviation
SME	small and medium-sized enterprise
SMOTE	synthetic minority oversampling technique
SPSS	statistical package for the social sciences
SVM	support vector machine
STATA	software for statistics and data science
TN	true negative
TP	true positive
TLI	Tucker-Lewis index
U.S.	United States
VIF	variance inflation factor
vs	versus (against)
XGBoost	eXtreme gradiant boosting

Executive summary

Scholars agree on the central roles that start-up activities and cognition play with regard to the successful launch of a business. Despite their importance, key questions about how behavioral differences between nascent entrepreneurs impact the outcomes of their ventures remain unanswered. This dissertation attempts to answer some of these fundamental questions by providing empirical evidence of the role that cognition plays in the gestation process and by evaluating machine learning algorithms to predict a venture's outcome at an early stage.

While cognitive and motivational aspects driving start-up activities remain relatively underdeveloped in the current literature, the first paper addresses part of this shortcoming by showing how self-regulation influences the relationship between nascent entrepreneurial exploitation activities, firm birth and firm abandonment. This paper provides evidence of an increase in persistence as well as an increase in the likelihood of achieving firm birth due to a regulatory fit between exploitation activities and a promotion orientation.

Next to the conducted start-up activities, the impact of innovation and imitation on a firm's emergence and survival remains theoretically and empirically inconclusive. The second paper seeks to shed some light on the impact of the rather under-researched imitative business ideas on the likelihood of a firm's emergence. By drawing on social identity theory, this paper further delineates how founder identity types impact the likelihood of imitating and innovating. The investigations conducted confirm that imitation can be beneficial in an early stage and that certain identity types (i.e., Missionaries) are rather prone to innovate.

Finally, this research seeks to predict the likelihood of entrepreneurial success. While few attempts have been made to provide an early stage venture outcome prediction model, the third paper set forth to apply and evaluate a diverse set of machine learning algorithms as well as artificial neural networks to predict the likelihood of firm birth and firm abandonment. The analyses reveal that neural networks can provide a comparable performance in predicting the venture outcome, while especially the XGBoost algorithm provides good performance metrics in predicting the likelihood of a firms' birth.

In sum, the three papers offer valuable contributions to the extant entrepreneurship literature with respect to the impact of cognition on different venture outcomes. By using different machine learning techniques in predicting the venture outcome, this dissertation provides practitioners with a decision support system to optimize corresponding resource allocation.

Zusammenfassung

Obwohl in der Forschung Einigkeit über die Wesentlichkeit von Gründungsaktivitäten herrscht, bleiben zentrale Fragen bezüglich des Einflusses unterschiedlicher Verhaltenskognitionen und dem erfolgreichen Gründungsvorhaben ungelöst. Diese Dissertation untersucht den Einfluss einiger solcher Verhaltenskognitionen innerhalb des Gründungsprozesses und versucht mittels Methoden maschinellen Lernens den Ausgang von Gründungsvorhaben vorherzusagen.

In der aktuellen Literatur finden sich noch eine Vielzahl an Unzulänglichkeiten, was den Einfluss kognitiver sowie motivationaler Aspekte auf Gründeraktivitäten betrifft. Der erste Beitrag betrachtet, wie anhand der Regulationsfokustheorie die Beziehung zwischen Aktivitäten der Implementierungsphase und dem erfolgreichen sowie erfolgslosen Gründungsvorhaben beeinflusst wird. Der Beitrag legt nahe, dass der Promotionsfokus die Persistenz sowie den Prozess erfolgreicher Unternehmensgründungen positiv beeinflusst.

Der Einfluss von Innovation und Imitation auf die erfolgreiche Entstehung eines Unternehmens ist sowohl theoretisch als auch empirisch unschlüssig. Der zweite Beitrag beleuchtet die Auswirkungen der weniger erforschten imitativen Unternehmenstypen auf erfolgreiche Gründungsvorhaben. Anhand der Theorie der sozialen Identität wird ausserdem der Einfluss verschiedener Gründeridentitäten auf ein nachahmendes und innovatives Verhalten untersucht. Die Untersuchungen zeigen, dass Imitation in einem frühen Stadium von Vorteil sein kann und dass bestimmte Identitätstypen eher zu Innovationen als zu Imitation neigen.

Schließlich ist die Forschung daran interessiert, erfolgreiche Gründungsvorhaben vorherzusagen. Während nur wenige Versuche unternommen wurden, um in einem frühen Stadium der Unternehmungsgründung Vorhersagemodelle bereitzustellen, wurden im dritten Artikel unterschiedliche Methoden des maschinellen Lernens sowie künstliche neuronale Netze angewandt, um die Wahrscheinlichkeit erfolgreicher sowie erfolgloser Gründungsvorhaben vorherzusagen. Die Resultate zeigen, dass nebst künstlicher neuronaler Netze insbesondere der XGBoost Algorithmus zur Vorhersage erfolgreicher Gründungsvorhaben geeignet erscheint.

Insgesamt leistet diese Dissertation einen wesentlichen theoretischen Beitrag zum Verständnis des Einflusses kognitiver Verhaltensmuster innerhalb des Gründungsvorhabens. Durch die Verwendung verschiedener Techniken des maschinellen Lernens bei der Vorhersage des Gründungsvorhabens bietet diese Dissertation Praktikern ein Entscheidungshilfesystem zur Optimierung der entsprechenden Ressourcenallokation.

1. Introduction

1.1 Overarching topic and research aim

The emergence of a new organization forms the core of the entrepreneurial process. Extant literature suggests that a higher number of start-up activities lead to a higher likelihood of a firm to be born (Brush et al., 2008; N. M. Carter et al., 1996; Manolova et al., 2012; Reynolds, 2016). These actions are grounded in the individual entrepreneurial behavior (e.g., Fisher, 2012) and involve various mediators and moderators (e.g., Baron, 2012; Johnson et al., 2015). As such, cognition plays an essential roles in explaining entrepreneurial action and the successful creation of a firm (Brockner et al., 2004; Davidsson, 2016; Fauchart & Gruber, 2011; Gatewood et al., 1995; Pollack et al., 2015).

Despite the understanding of the importance of start-up activities and these behavioral constructs, only a handful of studies empirically link behavior and characteristics of the nascent entrepreneur with the design of the venture gestation process (e.g., Brockner et al., 2004) in order to explain the successful birth of a firm. If they did so, they are based either on (a) contextual factors (e.g., Giotopoulos et al., 2017; Steffens et al., 2012), (b) specific activities such as business planning (e.g., Brinckmann et al., 2010; Chwolka & Raith, 2012; Liao & Gartner, 2006); (c) a conceptual approach (e.g., Brockner et al., 2004); or (d) a specific part of the gestation process, for example, the opportunity recognition phase (e.g., Tumasjan & Braun, 2012). This is among others attributable to a lack of data availability providing information on simultaneous behavioral and cognitive constructs as well as on start-up activities and firm emergence.

Considering the importance of a multitude of cognitive and other psychological processes (Baron, 2007; Forbes, 1999; Grégoire et al., 2015) on the founding of a company, this cumulative dissertation seeks to contribute to a more comprehensive understanding of the nature and influence of behavioral differences within the gestation process. Each academic paper of the dissertation furthers our understanding of the successful firm emergence process and addresses specific unresolved research questions and knowledge gaps in this domain. The first paper attempts to demonstrate how self-regulation impacts the relationship between the number of exploitation activities and the outcome of the venture. Paper two analyzes how the type of the business idea - imitative or innovative - influences the likelihood of establishing a firm and further delineates, based on the social identity theory, who the imitators and innovators

are. Finally, paper three proposes a firm emergence prediction model using more advanced machine learning techniques such as artificial neural networks.

The analyses of the first two papers utilize data from the Swiss Panel of Emerging Businesses (SwissPEB), a longitudinal study of Swiss nascent entrepreneurs. This dataset consists of 298 Swiss nascent entrepreneurs. The analysis for the third paper draws on the dataset of five longitudinal studies that have been harmonized into one dataset, consisting of 3,537 nascent entrepreneurs from the United States, China, Sweden, and Australia. The contributions of this cumulative dissertation are manifold and relevant to the literature on start-up gestation activities, entrepreneurial cognition and practitioners in particular.

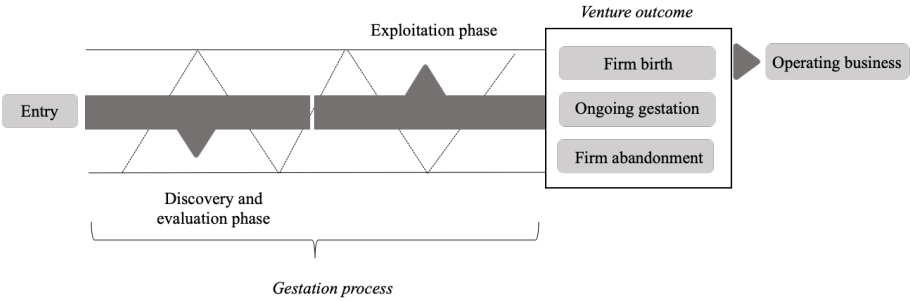
The organization and comprehensiveness of this dissertation are ensured by the use of clearly structured and coordinated chapters. The remainder of Chapter 1 provides further information on key theoretical constructs. Chapter 2 offers an overview of the three academic papers focusing on the addressed knowledge gaps and research methods. Chapters 3, 4 and 5 present the three academic papers. Finally, Chapter 6 summarizes the overall implications of the study's findings for theory and practice, outlines their key limitations, and offers recommendations for future research before finishing with some concluding remarks.

1.2 Definition of main theoretical concepts

1.2.1 The entrepreneurial process, nascent entrepreneur(ship) and the venture outcome

The creation of a new firm is a process rather than an instant event (Davidsson, 2016; Reynolds & White, 1997). This process begins when an individual or a team decides to become entrepreneurially active; it can be divided into discovery, evaluation, and exploitation phases (Shane & Venkataraman, 2000). While being interrelated and dynamic, discovery and evaluation refer to the conceptual side of the venture development, and exploitation deals with the realization of the business idea (Davidsson, 2016). Figure 1 provides an adapted graphical summary of this overlapping process based on Davidsson (2016, p. 29) and Reynolds (2017, p. 42).

Figure 1. Graphical representation of the start-up process



Note. This figure represents the firm creation process and outlines the different stages with the corresponding outcome. Adapted from "When is a Firm Born? Alternative Criteria and Consequences", by P. D. Reynolds, 2017, *Business Economics, 52*(1), p. 42. (https://doi.org/10.1057/s11369-017-0022-8). Copyright 2017 by Springer Nature. Adapted with permission.

The lines in Figure 1 that go back and forth from the discovery and evaluation phase to the exploitation phase reflect part of this dynamism. For example, the process possibly starts with a business idea, which is evaluated via a diverse set of activities. After the evaluation, the entrepreneur attempts to exploit the idea by researching the market and accumulating resources from investors, thus transforming the business idea into a reality. The feedback from the market can provide the entrepreneur with new information, which may lead to a change in business concept, thereby possibly placing him or her back to the discovery and evaluation phase. While the process can also evolve differently, it ends in a first step with the nascent entrepreneur either having successfully created a new firm, still persisting in the process (defined as "ongoing gestation" in this cumulative dissertation), or abandoning the firm. "Firm birth" is defined as the presence of monthly profits that cover expenses and owner salaries, "firm abandonment" is defined as having quit working on the start-up venture, and "ongoing gestation" relates to entrepreneurs who are still in the process, neither having become a profitable new business or having abandoned the venture (Reynolds et al., 2016a). In the literature the terms firm birth, firm creation, firm emergence, venture emergence, organizational emergence, and emergence success are often used interchangeably to describe the process of a business come into being (Brush et al., 2008; Davidsson, 2016; Gordon, 2012; Katz & Garnter, 1988; Manolova et al., 2012; Reynolds, 2017). While "nascent entrepreneurship" relates to the first intent to pursue an opportunity and the ongoing effort to create the firm (graphically represented in Figure 1 as the

gestation process), “nascent entrepreneurs” are defined as individuals or teams who initiate activities and are in the process of establishing a business venture (Dimov, 2010; Reynolds & White, 1997). In this dissertation, the entry point - and, thus, the definition of being considered a nascent entrepreneur - builds upon the Panel Study of Entrepreneurial Dynamics (PSED) and implies that: “i) they consider themselves as involved in the firm creation process, ii) they have engaged in some start-up activity in the past 12 months, iii) they expect to own all or part of the new firm, and iv) the initiative has not progressed to the point it may be considered an operating business” (Reynolds & Curtin, 2007, p. 9).

1.2.2 Start-up activities and the discovery–exploitation framework

The definition of the single start-up activities that nascent entrepreneurs conduct during the gestation process was in some way or another inspired by the framework of Katz & Gartner (1988) and over the past decade has been itemized in a diverse set of start-up activities. Since the inclusion of these start-up activities into various datasets - such as CAUSEE, PSED I and II, SE-PSED, CH-PSED and SwissPEB[1] - researchers have attempted to analyze the relationships between single and grouped activities on firm emergence to determine how a successful start-up process evolves (Brush et al., 2008; N. M. Carter et al., 1996; Gordon, 2012; Manolova et al., 2012). While results regarding the impact and necessity of specific activities (e.g., business planning) are inconsistent (Chwolka & Raith, 2012; Daft & Weick, 1984; Delmar & Shane, 2003; Honig & Karlsson, 2004; Honig & Samuelsson, 2012; Liao & Gartner, 2007) some level of activity is needed. Moving beyond single activities, scholars further began to research how dynamics as well as sequences of activities influence the successful firm creation process (Gordon, 2012; Lichtenstein et al., 2007). On this note, Gordon (2012) conceptualized the single activities into the discovery, evaluation and exploitation framework (Shane & Venkataraman, 2000) to further nurture the understanding of the evolvement process. Table 1 summarizes the different activity types into the discovery and exploitation framework (following Gordon, 2012, p. 238). Note that Gordon (2012) combines the discovery and evaluation process into one phase.

[1] CAUSEE stands for Comprehensive Australian Study of Entrepreneurial Emergence, SE-PSED for Swedish Panel Study of Entrepreneurial Dynamics, CH-PSED for Chinese Panel Study of Entrepreneurial Dynamics and SwissPEB for the Swiss Panel of Emerging Businesses.

Table 1. Discovery-exploitation framework

DISCOVERY ACTIVITIES	EXPLOITATION ACTIVITIES
Started thinking about business	Registered business name
Began product development	Decided location for business
Developed proprietary technology	Established legal form
Commenced customer discussions	Signed ownership agreement
Collected competitor information	Began marketing
Defined market opportunities	Leased major equipment / facilities
Produced financial projections	Purchased liability insurance
Determined regulatory requirements	Registered business number (ABN)
Began developing business plan	Registered for goods and service tax (GST)
	Registered for payroll tax (PAYG)
	Sought outside funding
	Established supplier credit
	Hired employee
	Opened business bank account
	Invested own money
	Retained an accountant
	Retained a lawyer
	Made business contactable
	Created business website

Note. This table allocates the different start-up activities into the discovery-exploitation framework. Adapted from *Dimensions of the venture creation process: Amount, dynamics, and sequences of action in nascent entrepreneurship* (p. 238), by S. Gordon, 2012, Queensland University of Technology. Copyright 2012 by S. Gordon. Adapted with permission.

There are a few benefits in considering the gestation process in relation to start-up activities as such dual process of discovery and exploitation. A dichotomous conceptualization is the least intricate but still allows for an in-depth analysis of the internal temporal order of the venture creation process. This temporal order relates to the different phases (discovery and exploitation) of the entrepreneurial process (Shane and Venkataraman, 2000) and delineates a possible sequence in which the process unfolds - that is, the early stage is dominated by discovery activities and the later stage is dominated by exploitation activities (Gordon, 2012). Such

contemplation allows the study of start-up activities and the venture outcome from a broader perspective and permits the incorporation of how cognition impacts the process itself.

1.2.3 Entrepreneurial behavior and cognition

"Behind entrepreneurial action are entrepreneurial intentions, behind entrepreneurial intentions are known entrepreneurial attitudes; Behind entrepreneurial attitudes are deep cognitive structures; Behind deep cognitive structures are deep beliefs" (Krueger, 2007, p. 124). While entrepreneurial behavior is a complex construct with different facets, this dissertation draws on the ideas of Krueger (2007) who outlines the importance of integrating cognitive development psychology in order to better understand the background of certain entrepreneurial action. Against this background, entrepreneurial cognition (EC) research provides answers to questions such as, how entrepreneurs think (Mitchell et al., 2007) and why differences in action and duration among entrepreneurs exist (Baron, 2004). Additionally, EC research helps not only to explain the effects of cognitive variables for entrepreneurial action, "but also to articulate the relationships between mind, environment, and entrepreneurial action that influence the development, use, and transformation of these variables" (Grégoire et al., 2011, p. 1448). With the renewed interest on the individual entrepreneur (Baron, 2004), cognitive theories have become a significant way to study the entrepreneurial process (Grégoire et al., 2011; Randolph-Seng et al., 2015; Sassetti et al., 2018). Within this cognitive and behavioral realm, this dissertation focuses on two influential and specific theories: (1) regulatory focus theory and (2) social identity theory.

1.2.3.1 Self-regulation - regulatory focus theory

Self-regulation is widely seen as a process that enables people to guide their goal-related activities over time and across different contexts. An effective self-regulatory process helps entrepreneurs select appropriate goals, maintain consistent focus on them, and adjust their actions during the start-up process (Karoly, 1993). One of the most influential self-regulation theories is the regulatory focus theory (Higgins et al., 2001).

Regulatory focus theory (RFT) provides a framework toward a fuller explanation of how the entrepreneurial mindset is constructed, and how entrepreneurs go about accomplishing their tasks and objectives (Brockner et al., 2004). The theory describes two different orientations - a promotion focus and a prevention focus. A promotion focus describes preferences towards growth, development and advancement needs. A prevention focus refers to preferences towards security and safety, especially with regards to tasks such as the avoidance of risks (Higgins et

al., 2001). Specifically, RFT draws attention to the motivational and strategic inclinations that people employ during pursuit of their goals (Brockner et al., 2004).

1.2.3.2 Social identity theory

Identity theory attempts to explain the human "self" and behavior in terms of people's identities (Laird, 2012). In the social identity theory, the definition of one's self is either related to being a member of an in-group that has significantly different attributes from an out-group (Tajfel & Turner, 1979). The own comparison of the positive in-group attributes with the perceived negative aspects of the out-group can enhance self-efficacy and increase the self-esteem (Abrams & Hogg, 1988; Stryker & Burke, 2000). Identity development begins from early childhood, is constantly questioned, and grows over the course of a person's life (Fauchart & Gruber, 2011). In an entrepreneurial context, these social identities can shape the goals and strategies around which the ventures were initially founded (Laird, 2012) and play a central aspect in the behavioral process that drives venture creation (Cardon et al., 2009). Fauchart and Gruber (2011) proposed a categorization into three social identity types: *Darwinian*, *Communitarian* and *Missionary*. Darwinians focus on the financial success, are driven by economic self-interest, see their frame of reference in competing firms, and follow an attitude of being successful via a "business school" approach (Fauchart & Gruber, 2011). The central aspect for Communitarians is the contribution towards a community or group with whom they strongly identify. The benefit for and the feedback from the group are essential. Missionaries seek to improve the society as a whole and see success as making the world a better place. As such they focus on understanding the firm as an agent of change (Fauchart & Gruber, 2011; Sieger et al., 2016). These identities can influence entrepreneurial decisions differently (Shepherd et al., 2015), thus possibly impacting the successful birth of a firm.

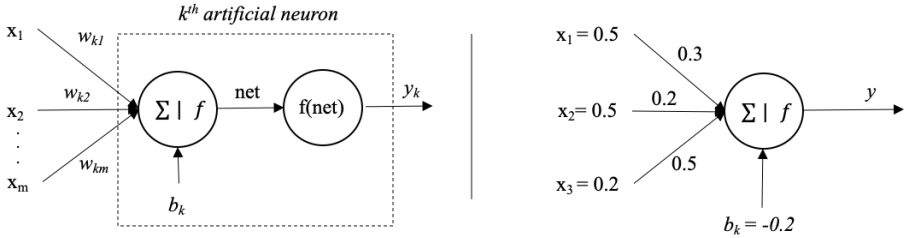
1.2.4 Supervised learning and artificial neural networks

Based on the type of learning process, data-mining techniques can be classified either as (semi-)supervised or unsupervised learning methods. Simplified, supervised learning algorithms learn to estimate an unknown dependency by being trained on input and output data (Mehmed, 2011). Similar to being supervised by a teacher who knows the output values, the system first learns this knowledge and then applies it to a new, similar problem with the corresponding input data. This learning process unfolds via the training dataset and the error signals that relate to learning based on differences between the desired response and the actual response (Mehmed, 2011). In reducing this error, the system learns to better map the input data

to the output data. On the other hand, unsupervised learning models do not build on this teaching process via the training data. They are used to identify unknown patterns in the data, with clustering techniques often being referred to as typical unsupervised learning (Aggarwal, 2015). While different techniques exist for either supervised or unsupervised learning, artificial neural networks (ANN) are currently popular and are a common algorithm in supervised learning (e.g., Heydari et al., 2012; Sevim et al., 2014; Topuz et al., 2018). Given the research goal (see Chapter 1.1), this dissertation focuses on supervised learning methods.

Artificial neural networks originate from attempts to imitate the information processing system of the human brain (Fertsch, 2020) and have been studied for nearly more than five decades (Mehmed, 2011). Similar to the neural brain connections, a neural network structure consists of a number of nodes connected through directional links, which specify the causal relationship between connected nodes. Figure 2 and the explanations in the next paragraph built upon Mehmed (2011, p. 201-203) and are used for illustrative reasons to provide a basic understanding of the logic behind neural networks.

Figure 2. Model of an artificial neuron



Note. This figure illustrates the basic logic of an artificial neural network with an example. Reprinted from *Data Mining Concepts, Models, Methods, and Algorithms* (p. 202), by K. Mehmed, 2011, Hoboken, New Jersey: John Wiley & Sons Inc. Copyright 2011 by Institute of Electrical and Electronics Engineers. Reprinted with permission.

The process begins with the input variables (x_1-x_m). These variables are multiplied by a corresponding weight (w_k), with k reflecting the index of a given neuron in an ANN. When the network begins to "fire", the input variables are multiplied and summed, with *net* being denoted as the weighted sum of products x_i and w_{ki} for i = 1, ..., m. We further add a bias (b_k) which, to a certain extent, can be compared to the intercept of a normal linear equation and "has the effect of increasing or lowering the net input of the activation function, depending on whether it is

positive or negative" (Mehmed, 2011, p. 201). Mathematically this process can be described the following:

$$net_k = x_1 w_{k1} + x_2 w_{k2} + x_m w_{km} + b_k$$

Based on this calculation, the next step involves the artificial neurons to compute the output y_k as a certain function of net_k value.

$$y_k = f(net_k)$$

This function is called activation function and defines the output of that node given the set of inputs. Different activation functions exist. For example, for a binary classification problem sigmoid functions have often been applied. However, depending on the problem, newer functions are developed to increase the model performance (e.g., Ramachandran et al., 2017). Translating these elements into a practical example with a sigmoid activation function (Figure 2), the calculation would be the following:

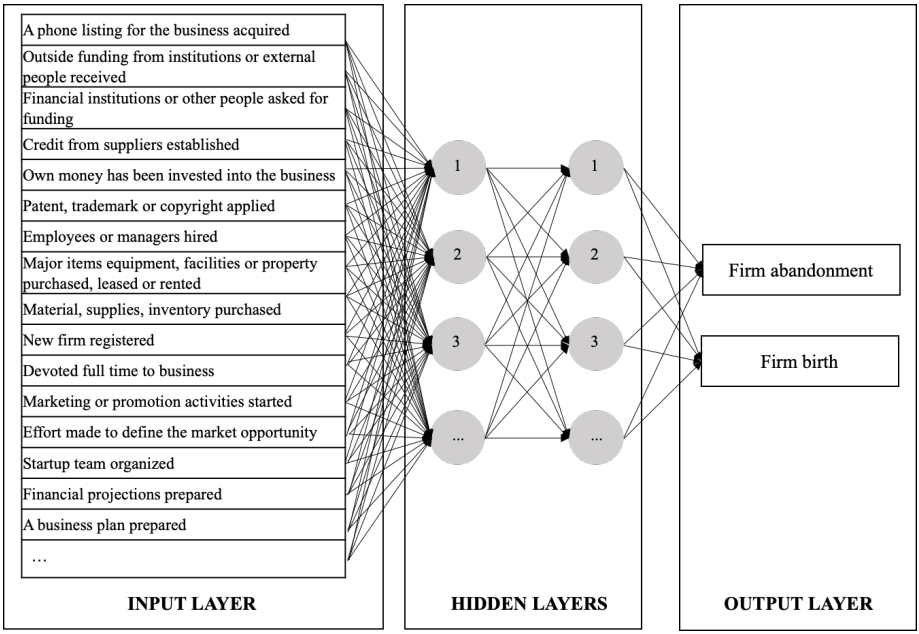
$$net = 0.5 * 0.3 + 0.5 * 0.2 + 0.2 * 0.5 + (-0.2) = 0.15$$

$$y_k = \frac{1}{(1+e^{-0.15})} = 0.54$$

This value would equal the first prediction. With this, the learning process begins with the system comparing this estimated output with the actual results, produces an error signal (the difference between the predicted value and the actual value), and starts to adjust the input weights of a neuron to make the output signal y_k come closer to the desired response in an iterative way. This is done by minimizing the cost function, which is the instantaneous value of the error energy. This iterative weight adaption is done using an algorithm called gradient decent, while estimating the influence of the individual weight on the error function is determined via an error backpropagation algorithm (propagate back from the last layer of the network). This process allows neural networks to learn to accurately predict the desired output.

The outlined highly simplified example illustrated a processing step with three variables and a direct link to the output variable. In practice, these networks and the corresponding architecture are more complex. As such, a typical neural network consists of an input layer, one or more hidden layers and an output layer. The hidden layer identifies features in the patterns presented at the input layer. Neural networks with more than one layer are considered to be deep neural networks. Figure 3 illustrates an excerpt of a neural network consisting of start-up activities with two hidden layers as used within this dissertation.

Figure 3. Graphical representation of a neural network using start-up activities



2. Overview of the cumulative dissertation

2.1 Knowledge and research gaps

In each of the academic papers, this cumulative dissertation addresses specific research gaps identified in the entrepreneurial literature. Scholars have confirmed the importance of entrepreneurial cognition and its impact on entrepreneurial action on firm emergence (e.g., Mitchell et al., 2007; Wood et al., 2012). Such insights are particularly relevant to entrepreneurship research, as they can contribute to a better understanding of the success or failure of an entrepreneur (cf. Hmielseki & Baron, 2008; Malmström et al., 2015), thus further nurturing from a theoretical and practical perspective our understanding of successfully creating a new firm. While important, scholarly understanding of differences between entrepreneurial actions - and how, when, and why relevant cognitive variables exert the influence they do on entrepreneurship - remains severely limited to date (cf. Davidsson, 2016; Grégoire et al., 2011).

For example, extant literature suggests that entrepreneurs who establish a higher number of start-up activities are more likely to achieve firm birth (Brush et al., 2008; Carter et al., 1996; Reynolds, 2016), with discovery activities typically conducted at an early stage and exploitation activities at a later stage (Gordon, 2012). However, recent research by Arenius et al. (2017) and Shim and Davidsson (2018) indicate that the gestation duration might be shorter than generally expected and that completing a small number of start-up activities may suffice for venture emergence. While only a handful of studies have sought to explain such action and performance difference based on cognition, for example self-regulation (Brockner et al., 2004), few empirical studies exists that relate differences in regulatory focus to the entrepreneurial process with the discovery and exploitation phase. Among others, this limited understanding can be attributed to a lack of longitudinal data that combines activity as well as cognitive elements. In the *first academic paper*, this research attempts to fill this gap by answering the following concomitant research question: How does self-regulation impact the relationship between the number of exploitation activities and the venture outcome?

The *second academic paper* examines how the type of the business idea, innovative or imitative, influences the likelihood of achieving firm birth and further attempts, based on the social identity theory, to narrow who the imitators are and who they are not. While various studies delineate a positive effect of innovativeness on the firms and their survival rates (e.g., Cefis & Marsili, 2005, 2006, 2012; Colombelli et al., 2013; Helmers & Rogers, 2010; Wagner & Cockburn, 2010), an emerging empirical strand within the entrepreneurial literature

highlights a context dependency with respect to innovativeness and firm survival (Boyer & Blazy, 2013; Buddelmeyer et al., 2009; Reid & Smith, 2000). While reasons for these different results are manifold, especially selection and survival bias are considered a major issue in related past studies (cf. Hyytinen et al., 2015). Moreover, while there is an abundant number of studies related to innovation, those that focus on the not less important imitative business types are rather scarce (Casadesus-Masanell & Zhu, 2013; Ethiraj et al., 2008; Fu & Tiez, 2015), and the "knowledge about what exactly leads to innovative rather than purely imitative business ideas is still quite limited" (Koellinger, 2008, p. 21). While some studies provide evidence of the influence of cognitive factors such as self-regulation (Amir, 2015; Cowden & Bendickson, 2018) or effectuation (Guo, 2019) on the degree of innovativeness, a vast number of unknown factors - such as founder identity types - exist, that can influence the innovative propensity of nascent entrepreneurs (Koellinger, 2008). Consequently, the second academic paper attempts to address some of these shortcomings by answering the following research questions: i) How does an imitative business idea impact the likelihood of achieving firm birth and ii) how do different founder identity types influence the likelihood of starting an innovative or imitative business?

The *third academic paper* provides an overview of key success and failure factors within the gestation process based on recent empirical results. Additionally, it investigates the applicability of a firm emergence prediction model by applying a diverse set of machine learning techniques and evaluating the applicability of artificial neural networks. Success or failure prediction models have previously been applied in the context of emerged start-ups as well as small and medium sized companies (Andrikopoulos & Khorasgani, 2018; R. Carter & Auken, 2006; Gupta et al., 2018; Lussier, 1995; Pompe & Bilderbeek, 2005; Y. Wu et al., 2010). However, these models demonstrate a variety of limitations. First, they relate to the later stage of the start-up process, that is the phase from firm birth to becoming an operating business (Reynolds, 2017), and not to the gestation phase. Moreover, most of these studies use data collected at a single point in time with no information included about the time, duration and event development since the start-up became operational. Since the venture process is considered to be a dynamic process (Packard et al., 2017), the usage of a longitudinal dataset, as employed in this study, allows a better grasp of impact factors that affect the likelihood of firm birth and firm abandonment over time. Finally, logistic regression models are currently the most common method used in the entrepreneurial literature to predict respective firm outcome in the context of an operating business (e.g., Baidoun et al., 2018; Lussier, 1995) as well as in the nascent entrepreneurial gestation literature. Despite the popularity of such classical

statistical methods, with recent developments in technology, newer models such as artificial neural networks appear to outperform classical models in different ways. Various scholars, for example, provide evidence of a higher prediction accuracy (Alharthi, 2019; Loureiro et al., 2018; Moro et al., 2014; Ravisankar et al., 2011; Sadat-Hashemi et al., 2005; Williams, 2017) of such models as well as outline the benefit of recognizing the appearance of non-linearity and interactions in input data (Gerasimovic & Bugaric, 2018). On this note, the third academic paper attempts to address the following research question: How do artificial neural networks, compared to other machine learning techniques, perform in predicting the likelihood of firm birth and firm abandonment?

2.2 Procedure and research methodology

The empirical analyses of the first two papers are built on the same random sample of n = 298 nascent entrepreneurs from Switzerland, drawn in 2015 through a professional business-to-business market research panel (LINK research institute) and interviewed over a period of three years. After data cleansing, the first paper consists of n = 181 cases, while for the second paper the sample consists of n = 180 cases. The third paper uses the publicly available harmonized PSED dataset (cf. Reynolds et al., 2016a, 2016b) and consists of n = 1,452 samples after data cleansing.

The *first paper* of this cumulative dissertation examines how the regulatory focus moderates the relationship between the effect of exploitation activities on firm birth and firm abandonment. The scale used to capture the regulatory focus in an entrepreneurial setting was developed by von Grone et al. (2016). This scale is adapted from Higgins's et al. (2001) original measurement. The scale consists of five items for measuring regulatory focus (four prevention and five promotion items). Maximum likelihood estimation procedure was used to examine the applicability of all scales in a confirmatory factor analysis. Moderated binary logistic regression is employed with regulatory focus as an interaction term to test the hypotheses. The authors controlled for common method bias, and statistical tests reveal that the data did not suffer from multicollinearity (Hair et al., 2010). Furthermore, control variables on the individual level, venture level, and market level are used to further scrutinize the data.

In the *second paper*, the authors analyze differences in the likelihood of successful firm birth based on the perceived innovativeness and imitativeness of the business ideas. The degree of innovativeness and imitativeness is measured via five items that have also been used in past studies and within the Global Entrepreneurship Monitor (Koellinger, 2008; Reynolds et al.,

2005). Furthermore, the paper attempts to identify, based on the social identity theory, entrepreneurs who consider imitative rather than innovative business ideas. The scale to measure the identity type is based on the theory of Fauchart and Gruber (2011) as well as an early adaption of the developed items based on the paper of Sieger et al. (2016) and consists of nine items (three per identity type) with acceptable model fit statistics using confirmatory factor analysis (CFA). Binary logistic regression is employed to test the hypotheses. The authors controlled for common method bias, and statistical tests reveal that multicollinearity is not an issue in the data (cf. Hair et al., 2010). Furthermore, control variables on the individual level, venture level, and market level are used to further scrutinize the data.

The *third paper* attempts to predict the venture outcome (i.e., firm birth and firm abandonment) by applying different machine learning techniques and testing the use of artificial neural networks. The methodological process included mainly three steps: first, a data preprocessing; second, optimization; and third, an evaluation step. After the data preprocessing, the initial dataset was split into a training and validation set, which included 75% of the observation (n = 1,089), while 25% of the dataset was used as a test and hold-out set (n = 363) to validate the performance of the models. Seven different models (logistic regression, k-nearest neighbors, decision tree, random forest, XGBoost, support vector machine, and artificial neural networks) were employed and evaluated using standard performance metrics (Topuz et al., 2018; Veganzones & Séverin, 2018). Common techniques (Mehmed, 2011) within the machine learning literature (e.g., cross-validation, feature selection, over- and under-sampling, and hyperparameter tuning via RandomSearchCV and GridSearchCV) were applied to test and derive the optimized prediction models.

2.3 Structure and main characteristics of the three papers

The structure of this cumulative dissertation is illustrated in Figure 4. It outlines how various cognitive theories impact different elements (i.e., exploitation activities, type of business idea) of the gestation process. Table 2 compares and summarizes the three papers according to common criteria such as applied methods, main contributions and the publication status.

Figure 4. Structure of the dissertation

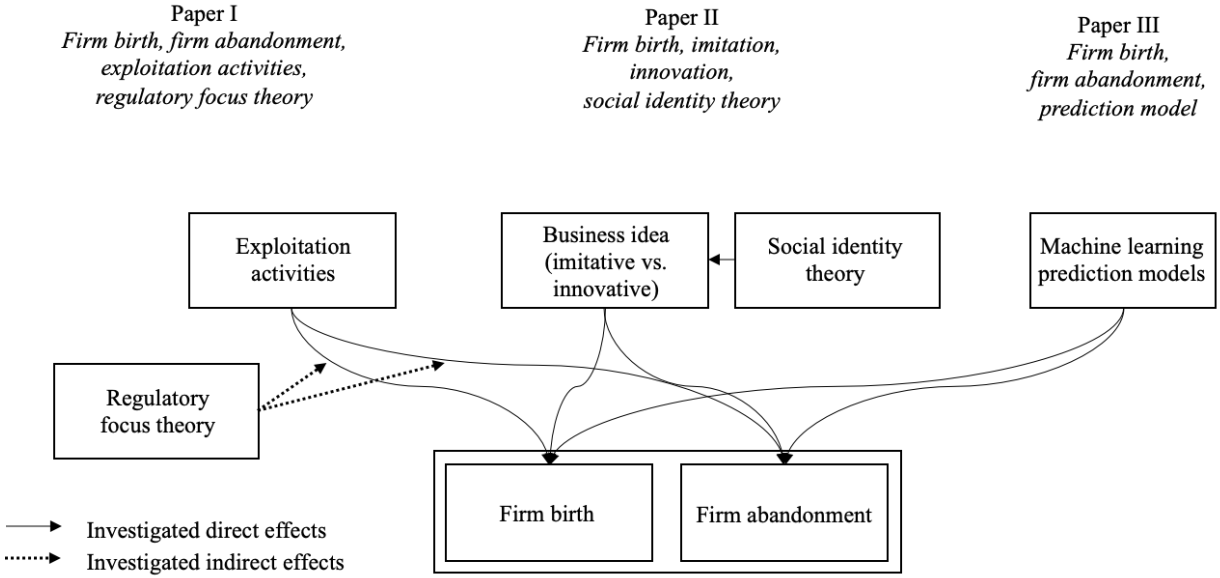


Table 2. Main characteristics of the three papers

Authorship	*Research gap*	*Main method(s) and sample*	*Main (theoretical/practical) contribution*	*Publication status / external recognition*
Paper 1: The impact of founder's regulatory focus on the process of firm birth and firm abandonment				
Paris Koumbarakis, Heiko Bergmann & Thierry Volery	Lack of understanding of the impact of founder's regulatory focus on the exploitation process and the likelihood of firm birth and firm abandonment	Moderated binary logistic regression, confirmatory factor analysis, 181 nascent entrepreneurs	Emergence success, regulatory focus theory, start-up activity literature; how differences in the entrepreneurial regulatory foci influence the relationship between the number of conducted exploitation activities and the venture outcome	Presented at Babson College Entrepreneurship Research Conference 2019 Accepted for publication in Management Decision
Paper 2: Imitate to emerge. Who is not imitating?				
Paris Koumbarakis, Thierry Volery	Inconclusive results regarding the impact of innovative/imitative business ideas on firm emergence. Social identity theory as antecedent of innovative/imitative business ideas	Binary logistic regression, confirmatory factor analysis, 180 nascent entrepreneurs	Emergence success, imitation and innovation as well as social identity theory literature; explaining why imitation can positively impact firm emergence and delineating based on the social identity theory the likelihood of pursuing an imitative or an innovative business idea	Presented at Babson College ET&P Workshop 2019 Submitted to Small Business Economics
Paper 3: Predicting the venture gestation outcome with machine learning methods and artificial neural networks				
Paris Koumbarakis	Lack of knowledge of the applicability of newer machine learning methods including artificial neural networks to predict the venture outcome	Logistic regression, k-nearest neighbors algorithm, decision tree, random forest, XGBoost, artificial neural networks, 1,452 nascent entrepreneurs	Emergence success, prediction and machine learning literature; offering based on a diverse set of machine learning models a decision support system to predict the likelihood of firm birth and firm abandonment	Presented at doctoral colloquium of the European Academy of Management 2019 To be submitted to Decision Support Systems

3. The impact of founder's regulatory focus on firm birth and firm abandonment

3.1 Abstract

The purpose of this paper is to show how self-regulation influences the relationship between nascent entrepreneurial exploitation activities, firm birth, and firm abandonment. This study draws from a unique longitudinal dataset of 181 nascent entrepreneurs from Switzerland who have been interviewed by phone in 2015 and 2016. It uses moderated binary logistic regression to test the hypotheses. This study provides evidence that discrepancies in promotion orientation can explain the different ways exploitation can lead to an increased likelihood of firm birth and a decreased likelihood of firm abandonment, while respectively increasing persistence. Findings suggest that this is attributed to the regulatory fit between a promotion orientation and exploitation activities. For scholars, our findings provide insights into the reasons for entrepreneurial persistence, as well as how firm birth can be achieved with different levels of exploitation activities. This study provides entrepreneurs with information on how to increase their persistence as well as the likelihood of firm birth while considering their regulatory focus. Based on regulatory focus theory, this paper highlights different paths to firm birth with varying quantity of exploitation activities. We contribute to a greater understanding of firm emergence by accounting for the impact of regulatory foci.

3.2 Introduction

What are the steps leading to a successful business launch and how do entrepreneurs steer themselves towards this goal? This question is at the center of the gestation process in entrepreneurship. A large stream of literature in gestation has examined the start-up activities leading to business launch (Delmar & Shane, 2003, 2004; Lichtenstein et al., 2006, 2007; Manolova et al., 2012). Most of the empirical studies in this field draw on longitudinal studies such as the Panel Studies of Entrepreneurial Dynamics (PSED), which examine the actions of entrepreneurs at the time they are starting their business (Gartner & Liao, 2012; Reynolds & Curtin, 2009). Extant literature suggests that entrepreneurs who conduct a higher number of start-up activities are more likely to achieve firm birth (Brush et al., 2008; Carter et al., 1996; Reynolds, 2016), with discovery activities typically conducted at an early stage and exploitation activities at a later stage (Gordon, 2012). However, recent research from Arenius et al. (2017) and Shim and Davidsson (2018) indicate that the gestation duration might be shorter than

generally expected and that completing a small number of start-up activities may suffice for venture emergence. Nevertheless, the literature on gestation activities has largely ignored how entrepreneurs go about trying to accomplish their tasks and objectives. The cognitive and motivational aspects driving start-up activities remain relatively underdeveloped (Gartner & Shaver, 2012). In this context, self-regulation, which is widely seen as a process that enables people to guide their goal-related activities over time, is a fitting theoretical anchor to shed some light on the entrepreneurs' actions during start-up. Self-regulation implies "modulation of thought, affect, behavior, or attention via deliberate or automated use of specific mechanisms and skills" (Karoly, 1993, p. 25). To navigate successfully through the challenges of the start-up process, entrepreneurs need to adjust their goals and actions to the specific circumstances they face (Kröper et al., 2011). In so doing, they primarily rely on the enactment of their own cognitive resources and motivation.

Amongst self-regulation theories, regulatory focus theory (RFT) (Higgins, 1998) is particularly relevant for entrepreneurship because it draws attention to the motivational and strategic inclinations that entrepreneurs employ during business launch (Brockner et al., 2004). In a nutshell, the theory posits that in their efforts to pursue particular goals, people self-regulate the energy, means, and strategies they deploy in ways that maximizes the "fit" with their goal-pursuit values and beliefs, or in Higgins' terminology, their regulatory focus orientations. In recent years, a growing number of scholars have drawn on RFT processes and principles to investigate various aspects of entrepreneurs' effort and behavior, including opportunity recognition and exploitation (Tumasjan & Braun, 2012), effort in entrepreneurial task (Trevelyan, 2011), innovativeness (Cowden & Bendickson, 2018; Spanjol et al., 2011), and time allocation (Burmeister-Lamp et al., 2012).

In this study, we analyze how differences in the entrepreneurial regulatory foci influence the relationship between conducted start-up activities and the outcome of the entrepreneurial process over time. Specifically, we address the following research question: How does self-regulation impact the relationship between the number of exploitation activities and the venture outcome? We use moderated binary logistic regression to analyze the effect of regulatory focus on the relationship between exploitation activities and the possible outcomes of the gestation process: firm birth, ongoing gestation and firm abandonment. Following the Panel Study of Entrepreneurial Dynamics (PSED) design, we draw on a longitudinal dataset of nascent entrepreneurs in Switzerland (n=181).

We contribute to the developing literature on self-regulation within the gestation process in several ways. Firstly, we closely align RFT with the discovery and exploitation framework.By doing so, we provide further insight into the emergence of new organizations (Davidsson, 2016). Specifically, we contribute to the field of regulatory focus studies and examine the role of regulatory focus in the initial entrepreneurial action phase (Johnson et al., 2015). Secondly, we contribute to the literature on highly persistent entrepreneurs by aligning regulatory focus with the persistence decision (e.g., Baum & Locke, 2004; DeTienne et al., 2008). Finally, we provide a possible explanation for why a small number of activities may lead to the birth of a new firm (Arenius et al., 2017).

The remainder of the article is organized as follows. First, we conduct a review of the literature about discovery and exploitation activities during the start-up process and we theorize how the relationship between exploitation and firm birth as well as firm abandonment may differ in light of differences in regulatory focus. Next, we provide a description of the data and methodology. After that, we detail the results, including descriptive statistics, the results of the moderated binary logistic regression, and an analysis of the hypotheses. Finally, we provide a discussion and outline the contribution and limitations of the study.

3.3 Theory and hypotheses development

3.3.1 Discovery and exploitation framework and the venture outcome

The firm creation process is often characterized by two distinct, yet sometimes overlapping phases: the discovery and the exploitation of a business idea (Davidsson, 2016; Gordon, 2012; Shane & Venkataraman, 2000). Discovery refers to the conceptual aspect of the venture development and encompasses a series of activities from ideation to business concept development. Specifically, these actions deal with planning and projecting future information, as well as engaging with and gathering information from potential customers or competitors. On the other hand, exploitation reflects the attempt to execute the business idea via activities such as resource procurement and market entry (Davidsson, 2016). Wiklund et al., (2018) describes the exploitation phase “as the process of engaging in concrete actions (e.g., developing products, assembling resources, talking to potential customers, registering the new business) to bring a new business into existence (p. 388).” Similarly, Kuckertz et al., (2017) conceptualize exploitative behavior with activities such as acquiring funds, putting together an entrepreneurial team, and setting up an organization.

Next to the conceptualization of different start-up activities within the discovery and exploitation framework, Gordon (2012) analyzed the sequence in which firm birth and firm abandonment occurs. While being distinct, discovery and exploitation often comprise of an implied temporal order which delineates the sequence in which the process enfolds, that is a dominance of discovery activities in an early stage and a prevalence of exploitation activities in a later stage. Furthermore, Gordon (2012) suggests that discovery activities do not guarantee successful exploitation and that business ideas cannot be exploited if they have not first been discovered. Once the business idea is discovered and formulated, the entrepreneurship process outcome (birth, ongoing gestation, or abandonment) depends primarily on exploitation activities, which are at the core of our interest in this study.

Building upon these arguments, we posit that entrepreneurs who are conducting exploitation activities are more likely to become cash positive and less likely to terminate their venturing effort (Gordon, 2012). This proposition is in line with the extant literature, which suggests that entrepreneurs with a higher number of start-up activities are more likely to achieve firm birth and less likely to pull out during gestation (e.g., Brush et al., 2008; Carter et al., 1996; Manolova et al., 2012; Reynolds, 2016). Thus, our baseline hypotheses are the following:

H1a: The number of conducted exploitation activities has a positive effect on firm birth.
H1b: The number of conducted exploitation activities has a negative effect on firm abandonment.

3.3.2 Regulatory focus, exploitation activities and its relation to firm birth

In the past decade, an increasing number of scholars have suggested that entrepreneurs' regulatory focus plays a decisive role in understanding the underlying cognitive mechanisms of the entrepreneurial process (Baron et al., 2016; Brockner et al., 2004; Markovits, 2012). Regulatory focus theory "delineates how people engage in self-regulation, the process of bringing oneself into alignment with one's standards and goals" (Brockner et al., 2004, p. 203).

Two different foci exist to engage in self-regulation: a promotion and a prevention focus. A promotion focus describes preferences towards growth, development and advancement needs. It is often associated with an eagerness strategy by focusing on ensuring and maximizing "hits" (Crowe & Higgins, 1997). This eagerness strategy relates to the enthusiasm to attain advancements, as well as gains and emphasizes speed and achieving maximal levels of performance (Förster et al., 2003). In contrast, a prevention focus refers to preferences towards

security and safety (Higgins et al., 2001). It is associated with a vigilance strategy and its strategic means is focused on avoiding errors by ensuring "correct rejections" (Crowe & Higgins, 1997), thus being vigilant. If the adopted approach fits the respective regulatory orientation of the individual, an increase in motivation as well as performance has been shown (Higgins, 2000). This fit can alter different relationships, such as the impact of feedback signs (positive or negative) on motivation (Van-Dijk & Kluger, 2004, 2011).

Differences in self-regulation can influence the conduct of start-up activities and venture success (Brockner et al., 2004). For example, a promotion orientation might be beneficial for certain activities during the rollout process, such as motivating people to work for the venture. On the other hand, a prevention orientation focus might be relatively more advantageous for activities that entail assessing the feasibility of the venture, such as screening the idea. As such, a possible fit between the individual's orientation and the conducted task may positively impact motivation and performance (Shah et al., 1998). Drawing on this fit perspective, we argue that not only single activities, but in all probability the exploitation phase as a whole is affected by the regulatory focus. This assumption is supported by Fischer et al. (2018), who outline that during the rollout process, entrepreneurs increasingly engage in a promotion-focused self-regulatory process. This suggests that the rollout during the exploitation stage may be prone to a promotion orientation. Accordingly, entrepreneurs with high levels in promotion orientation may feel more comfortable with activities in the exploitation phase. This, in turn, is likely to affect the outcome of firm gestation and the birth of a firm. Therefore, we argue that, based on the fit between exploitation activities and the promotion orientation, entrepreneurs with a promotion focus may conduct exploitation activities differently from those with a prevention orientation. Such fit not only positively increases the enjoyment the actions provide (Freitas & Higgins, 2002) but also the engagement as well as the effort in respective tasks (Trevelyan, 2011). Given this difference in action intensity in relation to the proposed fit, promotion-oriented individuals may differ in how they conduct exploitation activities to achieve firm birth.

Moreover, there is evidence that entrepreneurs with a strong promotion focus tend to allocate more of their time to exploitation activities (e.g., hiring employees, register business, rent office places etc.) which encompass higher risk and cost compared to conceptual, discovery activities (e.g., conducting competitive analysis, writing a business plan) (Burmeister-Lamp et al., 2012). Therefore, aligning this proposed fit of the exploitation phase and the entrepreneur's orientation could alter the relationship between the number of exploitation activities and firm birth. In other words, if two entrepreneurs conduct the same number of exploitation activities but differ in their regulatory focus, the entrepreneur with a strong promotion orientation might

have a higher likelihood of firm birth because he or she conducts exploitation activities with more enthusiasm and a higher intensity. Hence, we hypothesize:

H2a: A high level in promotion focus strengthens the positive relationship between the number of conducted exploitation activities and firm birth.

Prevention-focused individuals are vigilant and cautious about exploiting potential business ideas and obtaining resources to exploit them (Gamache et al., 2015). They plan carefully and prefer formalized activities (Brockner et al., 2004). Such an approach fits with activities related to the discovery phase rather than the exploitation phase. Entrepreneurs with a prevention orientation, therefore, may be less inclined to show the same effort, enjoyment and engagement intensity to exploit the business idea in order to achieve firm birth compared with promotion-oriented entrepreneurs.

Moreover, if each additional hour allocated to the venture yields additional risks, prevention-focused entrepreneurs appear to allocate fewer hours to the venture than entrepreneurs with a weaker prevention focus (Burmeister-Lamp et al., 2012). Presuming that the exploitation phase is riskier than the discovery phase, this could further alter the positive relationship between the number of exploitation activities and firm birth. A high prevention orientation can induce the entrepreneur to be less motivated and over-cautious in translating respective activities into the creation of a firm. Hence, the relationship between the number of exploitative activities and firm birth weakens under the condition of high levels of prevention focus. The entrepreneur feels less enjoyment, allocates fewer hours into respective risky activities at a time where probably more effort is necessary, and lacks the persistence needed to launch the firm. Therefore, we formulate the following hypothesis:

H2b: A high level in prevention focus attenuates the positive relationship between the number of conducted exploitation activities and firm birth.

3.3.3 Regulatory focus, exploitation activities and its relation to firm abandonment

As outlined in hypothesis H1b, the numbers of start-up activities, especially exploitation activities, is associated with a decrease in venture abandonment (Gordon, 2012; Reynolds, 2016). Building on this argumentation, we believe that this relationship is also altered by differences in regulatory focus. High levels in promotion focus can drive entrepreneurs to persevere in face of difficulties and setbacks and help them remain committed to their tasks

(Crowe & Higgins, 1997). Promotion-oriented individuals show higher persistence on tasks that are framed in line with their regulatory orientation (Shah et al., 1998).

In relation to the proposed fit between the exploitation phase and the entrepreneurial orientation, promotion-oriented entrepreneurs may persevere in the face of adversity after having conducted a certain number of exploitative activities. Hence, the outcome of the gestation process (i.e., to persist rather than to abandon) for a given number of conducted exploitation activities is likely to differ between promotion-oriented and prevention-oriented entrepreneurs. Thus, we hypothesize:

H3a: A high level in promotion focus attenuates the negative relationship between the number of conducted exploitation activities and firm abandonment.

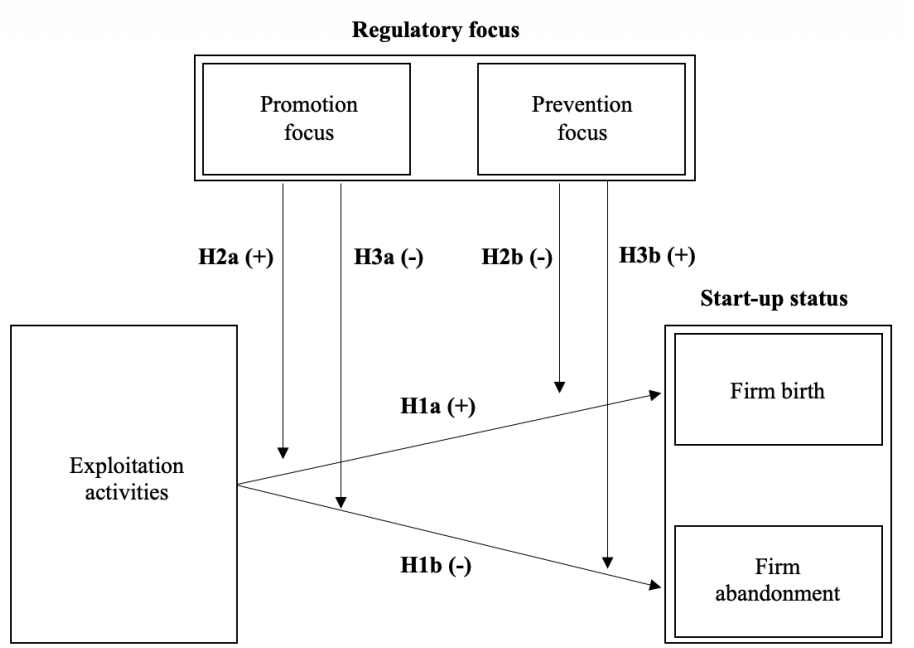
Conversely, we argue that high levels in prevention orientation alter the relationship between exploitation activities and firm abandonment in a way that a high number of exploitation activities increases the likelihood of firm abandonment. Given the assumed missing fit between high levels in prevention orientation and the exploitation phase, the positive effect on the persistence level (Shah et al., 1998) may be less pronounced. Moreover, prevention-oriented individuals are susceptible to become "over-cautious" and evoke a pessimistic view in light of information relevant to failure, leaning towards worst-case scenario planning (Higgins & Tykocinski, 1992; McMullen & Kier, 2016). Based on their attitude of ensuring correct rejections, they aim to avoid making mistake, which might lead them to "pull the plug" sooner rather than later. This vigilance approach may, in a difficult situation, motivate entrepreneurs to pull out in order to avoid committing an error (Crowe & Higgins, 1997).

Highly prevention-oriented individuals may therefore abandon their venture project rather than to persist after having conducted a certain number of exploitation activities without having achieved the desired goal. In other words, comparing two entrepreneurs with an equal number of conducted exploitation activities, a highly prevention-oriented entrepreneur is more likely to quit rather than to push forward. Thus, the negative relationship between the number of conducted exploitation activities and firm abandonment becomes stronger. Therefore, we hypothesize:

H3b: A high level in prevention focus strengthens the negative relationship between the number of conducted exploitation activities and firm abandonment.

The proposed model of this study is presented in Figure 5.

Figure 5. Hypothesized model



3.4 Research methodology

3.4.1 Data and sampling procedure

To test our hypotheses, we utilized data from the Swiss Panel of Emerging Businesses (SwissPEB), a longitudinal study of Swiss nascent entrepreneurs which follows the research design of the Panel Study of Entrepreneurial Dynamics (PSED) (Reynolds et al., 2016a). Nascent entrepreneurs were defined and identified as individuals who are currently actively trying to start a new business, and which is still in the pre-profit stage. In 2015, a total of 28,599 individuals in Switzerland were screened in a telephone survey by a market and social research firm. During this screening, a total of 680 nascent entrepreneurs were identified, of which 298

participated in the first detailed phone interview (wave 1). After twelve months, 292 of these nascent entrepreneurs were contacted again, and 67 did not participate, leaving a sample of 225 in the second wave. The rate of attrition of 23% in the Swiss sample is similar to previous longitudinal studies of nascent entrepreneurs (e.g., Brush et al., 2008; Manolova et al., 2012). Due to missing values, the analysis of this paper is based on a subsample of 181 observations.

The questionnaire draws on the protocol of the Panel Study of Entrepreneurial Dynamics (PSED) (Reynolds et al., 2016a) which increases validity and comparability with other PSED studies. The questionnaire included questions on nascent entrepreneurs' characteristics, firm characteristics, and discovery, as well as exploitation activities conducted thus far. This study used moderated binary logistic regression to analyze the effect of the regulatory focus on the relationship between exploitation activities on firm birth and firm abandonment.

3.4.2 Common method bias

A variety of steps was taken in order to mitigate the potential effects of common method bias. Firstly, we measured our dependent and independent variables one year apart, in different waves of our survey and used different response formats (Podsakoff et al., 2003). Secondly, strict confidentiality as well as the protection of the respondent anonymity was assured in order to decrease respondents' motivation to answer the survey in a socially desirable way (Podsakoff et al., 2003). Thirdly, the questions related to the constructs were embedded in a more comprehensive survey making it improbable the respondents could draw any conclusions about the research goal. All these measures ensure that common method bias was minimized in our study.

3.5 Measures

3.5.1 Dependent variables

Firm birth and firm abandonment were measured by a self-reported categorical variable as either being still nascent ("ongoing gestation"), having abandoned the venture, or having achieved initial profits. Following previous PSED studies (Reynolds et al., 2016a), achieving initial profits was used as proxy for firm birth. Of all the 181 people who had been nascent entrepreneurs in wave 1, after a period of 12 months (wave 2), 27 were profitable, 122 were still actively trying (ongoing gestation) and 32 abandoned the venture. The outcome and as such the two dependent variables, firm birth (0=ongoing gestation; 1=firm birth; abandoned ventures were excluded from the analysis) and abandonment (0=ongoing gestation, 1=firm

abandonment; profitable ventures were excluded from the analysis) were coded in a binary fashion. For the two models, ongoing gestation (n=122) is the reference variable, and it is compared to firm birth for the first model (n=27) and firm abandonment for the second model (n=32); thus the sample size is 149 in the first model and 154 in the second model. While the dependent variable was measured in the second wave of the data collection, all the other variables are taken from the first wave.

3.5.2 Independent variable

Exploitation activities related to the behaviors taken to establish the venture were measured using an eleven item questionnaire following the categorization of Gordon (2012). Specifically, the entrepreneurs were asked if they had already conducted either of the following activities: *Sought outside funding; Applied for IP protection; Hired employees; Created business website; Began marketing; Established legal form; Invested own money; Initial sales; Acquired or leased facility; Acquired or leased equipment; Participated in a support program.* For each activity, the answers were coded as binary variable (1=yes, 0=no). We computed the sum of all conducted exploitation activities as our main independent variable.

3.5.3 Moderator variable

Regulatory focus was measured using a scale of nine items. This scale was developed by von Grone et al. (2016) to capture regulatory focus in an entrepreneurial setting and it was adapted from Higgins's et al. (2001) original measurement. Promotion focus included four items: "*If you are facing a challenge in the start-up phase of your company, how important is it for you... a) ...to apply creative solutions to difficult situations? b) ...to focus on what you could gain? c) ...to make quick progress? d) ...to achieve great things?"* Prevention focus included five items: *"If you are facing a challenge in the start-up phase of your company, how important is it for you... a) ...to weigh the risks before taking a decision? b) ...to consider the long-term consequences of your actions? c) ...to weigh up the pros and cons? d) ...to go into detail? e) ...to protect yourself/the company from possible hazards?".* The items were measured using a Likert-type scale (1 = strongly disagree, 10 = strongly agree). Maximum likelihood estimation procedure was used to examine all scales in confirmatory factor analysis. CFA measures, based on Satorra-Bentler Scaled Chi-Squared Test due to non-normality of the items (Satorra & Bentler, 1994), suggest a good-model fit (SB_RMSEA=0.020, CFI=0.990, TLI=0.986).

3.5.4 Control variables

We controlled for several variables of the individual (gender, age, education, start-up experience), organizational (venture age, i.e., time since first conducted gestation activity, venture team, hi-tech, number of conducted discovery activities) and contextual level (industry type, perceived competitive environment) that previous studies associated with firm birth, firm abandonment, start-up activities, as well as self-regulation (e.g., Lichtenstein et al., 2007; Manolova et al., 2012, Reynolds, 2016). We note here that controlling for the number of discovery activities is particularly important as it is a proxy for the development stage of the venture. Entrepreneurs may realize that their venture is not feasible during the discovery phase and therefore may not even delve into an exploitation phase and abandon at an early stage.

3.6 Analysis and results

3.6.1 Descriptive statistics and correlations

Table 3 shows the mean, standard deviation, and correlation of the used variables. With regards to the variables under research, the table reveals three findings. Firstly, exploitation activities and firm abandonment are negatively correlated ($R = -0.39$, $p < 0.01$). Secondly, as anticipated, promotion focus and exploitation activities are positively correlated ($R = 0{,}17$, $p < 0.05$). Thirdly, even though not significant, exploitation activities appear to negatively correlate with firm birth ($R = -0.15$, p=n.s.). Moreover, we note that the venture age correlates with exploitation activities ($R = 0{,}29$, $p < 0.01$), as well as firm birth ($R = -0{,}30$ $p < 0.01$), which is in line with insights from Gordon (2012).

3.6.2 Hypotheses testing results

Table 4 presents the results of the logistic regressions. For the hypothesis tests, analyses were conducted in three steps for both dependent variables (firm birth and firm abandonment). In the first step, the control variables only are used (Models 1a, 2a) for both dependent variables. In the second step, exploitation activities as well as promotion and prevention focus were entered (Models 1b, 2b). Finally, the moderating effect of each regulatory focus and exploitation activities on both dependent variables was tested (Model 1c, 1d, 2c, 2d). Given the relatively small sample size and to account for possible variance in the distribution assumption we further applied bootstrapping (B=1000). The results of this re-sampling procedure did not differ, thus underlining the reliability of the reported statistical values. Our data did not suffer from multicollinearity, as all variance inflation factors (VIFs) were below 2, and hence, were below

the threshold of 10 (Hair et al., 2010). The -2 log likelihood ratio as well as the Nagelkerke R^2 are measures of how well the model fits the data.

This study provided some interesting insights regarding the tested hypotheses. First of all, our findings did not support *H1a*, but did support *H1b*: a high number of conducted exploitation activities appears to stave off firm abandonment but does not positively correlate with firm birth.

Table 3. Descriptive statistics and correlations

		Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.	Firm birth (compared to ongoing gestation)	0,18	0,39	1															
2.	Firm abandonment (compared to ongoing gestation)	0,21	0,41		1														
3.	Exploitation activities	4,41	2,15	-0,15	-,39**	1													
4.	Discovery activities	3,91	1,19	-,23**	-0,03	,24**	1												
5.	Promotion focus (PRO)	7,62	1,34	-0,04	-0,09	,17*	0,14	1											
6.	Prevention focus (PRE)	8,01	1,39	0,06	0,11	-0,11	0,02	,35**	1										
7.	Venture age [Duration in month]	36,11	22,63	-,30**	-0,15	,29**	0,11	0,08	-0,01	1									
8.	Venture team[a]	0,50	0,50	-0,02	0,15	-0,01	0,08	-0,07	0,03	-0,14	1								
9.	Perceived as a hi-tech venture[a]	0,09	0,28	-0,08	0,05	0,08	,15*	0,10	0,01	0,09	0,12	1							
10.	Service industry[a]	0,74	0,44	-0,05	-0,07	-0,06	-0,03	-0,04	-0,08	0,03	-,19*	-0,13	1						
11.	Manufacturing industry[a]	0,15	0,36	0,15	0,08	0,10	0,03	0,02	0,12	-0,02	,15*	,19*	-,72**	1					
12.	Nascent entrepreneur age [in years]	41,69	10,56	-0,01	-0,08	0,10	0,07	0,10	0,10	0,14	-,15*	0,05	0,11	0,02	1				
13.	Nascent entrepreneur gender[b]	0,59	0,49	0,14	,18*	-0,01	,23**	-0,03	-0,12	-0,10	,30**	,18*	-,29**	,23**	-,15*	1			
14.	Tertiary education[a]	0,50	0,50	-0,07	0,08	-0,04	,16*	-0,13	-0,11	,18*	0,01	0,08	0,14	-0,12	,17*	0,09	1		
15.	Previous start-up experience[a]	0,40	0,49	0,06	0,11	-0,06	0,09	-0,02	-0,03	-0,10	,23**	0,10	-0,03	0,05	0,09	,20**	0,12	1	
16.	Perceived competitive environment	3,43	0,99	-0,04	-,19*	0,03	-0,04	-0,09	-0,04	0,01	-0,10	0,01	,17*	-,18*	0,07	-,22**	0,12	-0,02	1

*Notes. N = 181, a: 1=yes 0=no, b: 1=male 0=female, c: 5=low 1=high, ** p<0.01, * p<0.05, Significance levels are two-tailed.*

However, by including the moderation effect of the regulatory focus, more specifically the promotion focus, the relationship between the number of conducted exploitation activities on firm birth changes. This indicates that the relationship between exploitation activities and firm birth may be more complex. Figure 6 shows the interaction effect in graphic form: under low levels of promotion focus the relationship between exploitation activities and firm birth is in line with the hypothesis *H1a* and the results of Gordon (2012). Under a high level of promotion focus, the effect of exploitation activities on firm birth turns negative, indicating that highly promotion-oriented entrepreneurs have a higher likelihood of achieving firm birth with few exploitation activities. Hence, our results suggest that a promotion focus influences the direction of the relationship between the number of conducted exploitation activities and firm birth, thus partially supporting hypothesis *H2a*. Our results also suggest that under a high level of promotion orientation and a high number of conducted exploitation activities the likelihood of firm birth decreases. This unexpected finding indicates that a high level in promotion orientation in combination with a high number of conducted exploitation activities is not necessarily as positive as we hypothesized in *H2a*. We discuss possible reasons for this finding below.

In relation to the prevention focus, we did not find any support for *H2b*. A prevention focus does not significantly influence the relationship between the number of exploitation activities and firm birth.

In the context of firm abandonment, the negative relationship between the number of exploitation activities and firm abandonment is altered by high levels in promotion focus, thus supporting hypothesis *H3a*. The interaction plot is shown in figure 7. Hence, the results from Gordon (2012) are supported insofar as the entrepreneur shows high levels in promotion focus. However, this effect changes with low levels in promotion focus and an increased number of conducted exploitation activities does not reduce the likelihood of firm abandonment. No significant relationship was identified for prevention focus. Thus, we did not find any support for *H3b*.

In summary, both full models containing all predictors were significant at the $p<0.001$ level, with Nagelkerke R^2 of 0.41. The two-way interaction is statistically significant at the $p<0.05$ level and the b coefficients are -0.33 for the Model 1c and -0.33 for the Model 2c.

Table 4. Results logistic regression

Variables	Firm birth (N= 149)								Ongoing gestation – Firm abandonment (N= 154)							
	Model 1a		*Model 1b*		*Model 1c*		*Model 1d*		*Model 2a*		*Model 2b*		*Model 2c*		*Model 2d*	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Constant	-1,84	1,93	-2,62	2,79	-13,85*	5,96	-4,56	4,89	0,30	1,36	0,21	2,39	-8,40*	4,11	-1,13	3,73
Control variables																
Venture level																
Venture age [in month]	-0,07**	0,02	-0,06**	0,02	-0,07**	0,02	-0,06**	0,02	-0,02†	0,01	0,00	0,01	-0,01	0,01	0,00	0,01
Venture team	-0,84	0,53	-1,00†	0,57	-1,05†	0,59	-0,96†	0,57	0,39	0,49	0,22	0,54	0,23	0,56	0,22	0,54
Perceived as a hi-tech venture	-1,43	1,21	-1,41	1,30	-1,59	1,42	-1,4	1,31	0,11	0,7	0,31	0,75	0,33	0,78	0,35	0,76
Service industry	0,98	1,14	1,12	1,19	1,38	1,28	1,11	1,19	0,1	0,67	0,00	0,74	0,03	0,77	0,02	0,75
Manufacturing industry	2,09	1,24	2,04	1,30	2,05	1,39	2,01	1,31	0,42	0,79	0,84	0,9	0,65	0,94	0,83	0,9
Individual level																
Nascent entrepreneur age	0,01	0,02	0,01	0,03	0,01	0,03	0,01	0,03	-0,01	0,02	0,00	0,02	0,00	0,02	-0,01	0,02
Nascent entrepreneur gender	1,12*	0,56	1,73**	0,66	2,12**	0,71	1,74**	0,66	0,47	0,52	0,66	0,59	0,94	0,62	0,68	0,60
Nascent entrepreneur education	-0,03	0,52	0,14	0,56	-0,03	0,58	0,12	0,57	0,54	0,46	0,28	0,51	0,16	0,52	0,25	0,51
Start-up experience	0,26	0,51	0,32	0,53	0,23	0,55	0,26	0,55	0,26	0,45	0,30	0,50	0,13	0,51	0,28	0,50
Market level																
Perceived competitive environment	0,17	0,26	0,23	0,28	0,15	0,3	0,22	0,28	-0,46*	0,22	-,44†	0,24	-0,5†	0,24	-0,45†	0,24
Main effects																
Discovery - Exploitation Framework																
Discovery activities			-0,60*	0,24	-0,76**	0,28	-0,60*	0,24			0,08	0,24	0,08	0,24	0,07	0,24
Exploitation activities			-0,05	0,15	2,51*	1,16	0,38	0,88			-0,52**	0,13	1,93*	0,91	-0,14	0,80
Regulatory focus																
Promotion focus (PRO)			0,26	0,26	1,76*	0,76	0,26	0,26			-0,08	0,18	0,95*	0,44	-0,07	0,19
Prevention focus (PRE)			0,03	0,21	0,07	0,22	0,27	0,54			0,19	0,2	0,32	0,22	0,35	0,40
Interactions																
PRO x Exploitation activities					-0,33*	0,15							-0,33*	0,13		
PRE x Exploitation activities							-0,05	0,11							-0,05	0,10
Regression function																
LR chi-squared	29,11**		37,06***		42,65***		35,90**		16,37†		39,17***		46,89***		39,39***	
d.f.	10		14		15		15		10		14		15		15	
-2 Log likelihood	111.91		103,96		98,37		105,18		141,06		118,22		110,50		118,00	
Nagelkerke R Square	0,29		0,36		0,41		0,35		0,16		0,35		0,41		0,35	

*Notes. N = 181 (B=1000) *** p<0.001, ** p<0.01, * p<0.05, †p<0.1, Significance levels are two-tailed.*

Figure 6. Interaction effect on firm birth (compared to ongoing gestation)

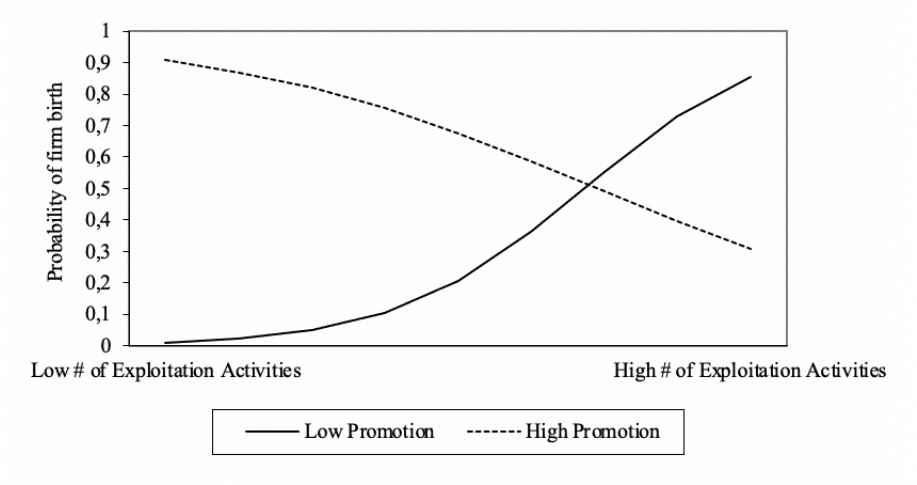
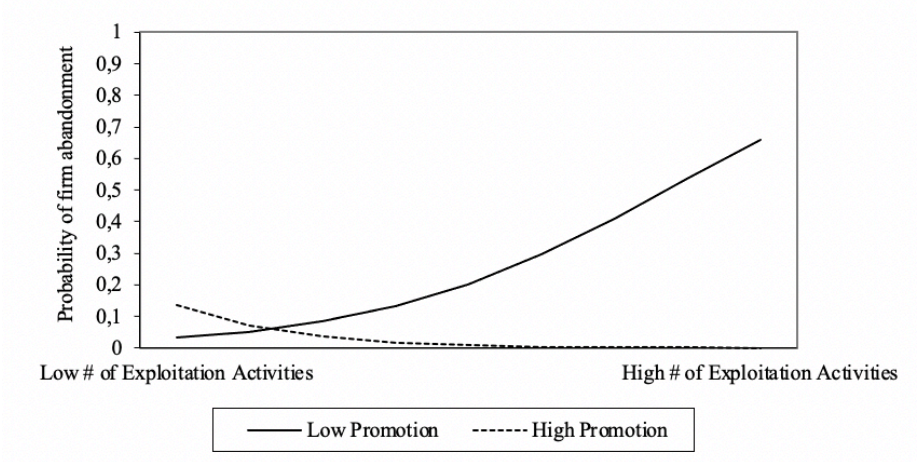


Figure 7. Interaction effect on firm abandonment (compared to ongoing gestation)



3.7 Discussion and contribution

One important aspect of the research on firm gestation is understanding differences in the actions entrepreneurs take in the lead up to a successful venture launch (Autio et al., 2013). The extant literature suggests that entrepreneurs who conduct a higher number of start-up activities are more likely to achieve firm birth and less likely to abandon the gestation process (Brush et al., 2008; Carter et al., 1996; Reynolds, 2016). Drawing on regulatory focus theory, we provide a somewhat different view on this relationship. In line with recent studies (Arenius et al., 2017; Shim & Davidsson, 2018), our results indicate that this relationship is more complex. Indeed, the completion of a small number of start-up activities may suffice for venture emergence. Our study provides an explanation for this phenomenon and thus contributes to a better understanding of the gestation process.

We find that the impact of the number of exploitation activities on firm birth is altered by the level of promotion orientation of the entrepreneur. In general, the number of start-up activities positively relates to firm emergence (Reynolds, 2016). However, a higher level in promotion orientation increases the likelihood of a firm to become profitable after the entrepreneur has conducted a small number of start-up activities. We explain this effect based on the fit between exploitation activities and a promotion orientation. Such fit is suggested to increase the decision makers' confidence, effort and engagement (Avnet & Higgins, 2006). By highlighting such a possible fit between individuals regulatory focus orientation and task demands (Spiegel et al., 2004), we contribute to research on the phenomenon of regulatory fit in a specific context, the gestation process. Further, we highlight that a promotion focus may not only influence a single task but potentially the entire exploitation phase in the entrepreneurship process. In showing the positive effect of the fit between the exploitation phase and the promotion orientation, we enlarge our understanding of how regulatory focus may influence the entrepreneurship process. Specifically, we show that for the same low number of exploitation activities, entrepreneurs with a high level in promotion orientation are more likely to achieve firm birth.

In context of a higher number of conducted exploitation activities, this effect is however less pronounced as illustrated in the downward slope within the moderation plot. An explanation for this effect could be the choice of the firm and/or the industry type. Entrepreneurs with high levels of promotion orientation may not choose venture ideas that allow for quick wins with a small number of exploitations activities but instead may opt for more ambitious projects that need several years of development and higher investments. These aspects may be

more prevalent for hi-tech ventures and/or in industries requiring significant investment to launch the venture (e.g., manufacturing). By including an industry (manufacturing and service industry) and a technology related control variable (hi-tech vs. low tech venture), we show that the effect of achieving firm birth despite a high number of start-up activities under a high level of promotion orientation is even more pronounced. Thus, the underlying level of technology (hi-tech firm) can explain to a certain extent the downwards slope. Further, it would reflect the effect that entrepreneurs with a higher promotion focus tend to pursue more innovative new ventures (Cowden & Bendickson, 2018).

In relation to firm abandonment, this study revealed that for a high number of exploitation activities, entrepreneurs with high levels of promotion orientation are less likely to abandon the firm, whereas entrepreneurs with low levels of promotion orientation are more likely to abandon the venture. A possible explanation for this result could be the susceptibility of high promotion-oriented individuals towards escalation of commitment (Kier et al., 2014). In light of an increase in time investment during the risky exploitation phase (Burmeister-Lamp et al., 2012), as well as the delineated perseverance of highly promotion-oriented individuals, there might be a tendency towards escalation of commitment. This suggestion aligns with the empirical results from Barsky and Zyphur (2016), who show that high promotion focus exacerbates escalation behavior as a project nears completion. These results provide an interesting insight by suggesting that this decision-making process (i.e., to abandon or not) might be influenced by different levels of promotion focus. We note that this persistence per se may be deemed neither positive or negative, however the persistence decision itself is an important part of the entrepreneurial process (Holland & Shepherd, 2013). Next to contributing to a better understanding of highly persistent entrepreneurs, these findings extend previous entrepreneurship studies in the context of firm abandonment, and provide evidence of another factor mitigating the impact of preventing nascent entrepreneurs from disengaging (Khan et al., 2014).

Finally, we did not find any moderating effect of the prevention focus. One explanation could be the association between high levels of prevention focus with formalization activities, like for example writing a business plan, collect competitor information or make financial projections. One might propose that entrepreneurs with high levels of prevention orientation stay longer within the discovery phase completing more formalized activities. Then, delving into the exploitation phase, they need longer to either achieve firm birth or decide to abandon the venture. Another explanation is that the application of regulatory focus theory in entrepreneurship tends to be unidimensional as most empirical studies analyze regulatory focus-

outcome relationships that are more inclined to reveal the positive effects of promotion and the negative effects of prevention (McMullen & Kier, 2016). Intuitively, this makes sense as entrepreneurs launching are likely to see their goals as creating paths to gains or advancement and concentrate on the rewards that will accrue when they achieve those.

In summary, our findings go beyond prior research by broadening the focus onto the influence of self-regulation on the gestation phase, as we show how firm birth as well as ongoing gestation can be fostered through a promotion focus. This extension is important, as nascent entrepreneurs often face various challenges during their journey and understanding the impact of motivation and mindset may help in overcoming various hurdles. Thus, by including regulatory focus within the discovery-exploitation framework, we provide evidence of the relationship between activities and motivational theory, via regulatory focus, and further contribute to a theory on the micro-foundations of entrepreneurial action (Shepherd, 2015).

3.7.1 Practical implications

Our study has several practical implications for nascent entrepreneurs with different levels of promotion orientation. First, this study shows that high levels in promotion orientation may increase the likelihood of achieving firm birth with few exploitation activities. However, entrepreneurs with high levels of promotion orientation who already have conducted a high number of exploitation activities without having achieved firm birth may reconsider their business idea, particularly in light of a possible escalation of commitment trap. Following the argumentation of McMullen and Kier (2016), an implication for nascent entrepreneurs to pursue their ideas with a pinch of skepticism and, after having conducted a vast number of exploitation activities without having achieved firm birth, they should reassess the feasibility and profitability of their business idea.

Moreover, in order to avoid falling into an escalation of commitment trap, entrepreneurs could adopt a stage gate approach with the option to pull out at predetermined different stages of the venture development, with an affordable loss. These tactics would allow revisiting and challenging the status quo and learning from past behavior, including already experienced negative emotions related to such trap. As Ku (2008) remarked, the imagination of such post-escalation regrets would already allow for de-escalation. Thus, for highly promotion-oriented entrepreneurs who have already conducted a high number of exploitation activities without having achieved firm birth, raising the awareness and identifying exit strategies would allow for a decrease in costs that arise with an escalation towards a commitment. This is a valuable

insight for incubators and support programs as it outlines the need for an early feedback round with a possible pivot of the idea. In doing so, the knowledge of the entrepreneurial regulatory focus can further guide an optimal framing of the feedback. For example, for promotion-oriented individuals this feedback may be verbally labeled towards gains to foster the motivation (Jarzebowski & Palermo, 2012).

3.7.2 Limitations and future research

This study has a number of limitations which also offer further avenues for research. Firstly, despite the validity and goodness of fit of the statistics used to capture regulatory focus in an entrepreneurial setting, this study is still of an exploratory nature. Further research applying the scale in the entrepreneurial context would be fruitful in order to account for differences outlined by von Grone et al. (2016). Moreover, even though the regulatory focus theory provided valuable insights regarding the venture process, other cognitive parameters may also influence the relationship between exploitation activities, firm birth and firm abandonment. For example, social identity theory has recently attracted the interest of the entrepreneurial research community and appears to play a central aspect in the behavioral process that drives venture creation (Murnieks et al., 2014; Powell & Baker, 2017; Shepherd et al., 2015). Thus, further analyses combining other cognitive theories might provide additional insights into how a gestation process progresses toward a successful outcome.

Secondly, even though the conceptualization of the entrepreneurial process as a dualistic discovery and exploitation phase is well established and related to specific activities (e.g., Kuckertz et al., 2017; Shane, 2012; Shane & Venkataraman, 2000), the classification of activities within each phase is still open to interpretation. The creation of a business plan for example might occur in a later stage in combination with acquiring funding where a firm is already established. Therefore, for future studies that research the effect of the regulatory focus on a specific activity, we recommend a further breakdown of the exact type of tasks as well as the purpose for such activities.

Furthermore, one can argue that the entrepreneurship process does not follow a sequential order, moving from discovery to exploitation, and that the exploitation of an idea can occur before or even without conducting discovery activities (Davidsson, 2003). In this context, the results of Gordon (2012, p. 197), suggest that the general flow of the venture creation process is a directional sequence from discovery toward exploitation and that "a majority of individual events are a symbiosis of discovery and exploitation action" where the entrepreneurial process

"begins and ends with distinct discovery and distinct exploitation". However, half-way through the venture creation these processes may overlap. This is the case when simultaneous discovery and exploitation actions are pursued (Gordon, 2012). In order to account for the effect of such simultaneous activities, we controlled for the respective number of conducted discovery activities and conceptualized the sequences underpinning the gestation process. Nevertheless, further research combining regulatory focus with complexity dynamics (Lichtenstein et al., 2007) could provide interesting insights.

Moreover, business creation can be seen as an idiosyncratic process (Newbert, 2005), and depending on the type of business created, not all exploitation activities may be relevant to enable firm birth. Some business ideas may be rather easy to implement while others need longer development periods as well as more exploitation activities. Thus, future research could look into differences of industry specific activities that influence the venture outcome.

3.8 Conclusion

With a growing relevance of start-up activities in our economy, it becomes important to better understand how cognition drives the entrepreneurial process and, subsequently, the successful birth of a firm. This study extends the knowledge of regulatory focus and its moderating effect on start-up activities and firm birth, as well as firm abandonment. Our findings confirm the impact different levels of promotion orientation can have on the gestation process leading to firm birth as well as firm abandonment.

4. Imitate to emerge – who is not imitating?

4.1 Abstract

The impact of innovativeness on a firm's emergence and survival remains theoretically as well as empirically inconclusive. In this paper, we seek to shed some light on the impact of imitative business ideas on the likelihood of a firm's birth and try to understand who the imitators and innovators are. Drawing on social identity theory, this study further delineates how different types of founder identities impact the likelihood of pursuing an imitative or innovative business idea. Using a unique longitudinal dataset of Swiss nascent entrepreneurs, our results show that nascent entrepreneurs who imitate existing business ideas are during the gestation process more likely to achieve firm birth compared to more innovative entrepreneurs. Moreover, based on social identity theory, our results suggest that "Missionaries" are less likely to imitate and more likely to create new, truly innovative products and services. Thus, our results suggest that from a policy perspective in order to foster innovation and social change it is worthwhile considering different social identity types.

4.2 Introduction

"It is better to fail in originality than to succeed in imitation."
(Herman Melville, 1850)

There is a wide agreement in the entrepreneurship literature that for new firms the degree of innovation amongst new business ventures varies greatly (Abernathy & Clark, 1985; Aldrich & Ruef, 2006; Cliff et al., 2006; Shepherd et al., 2000). In the past decade, a vast amount of research focused on understanding antecedents and outcomes of such innovations (e.g., Adomako, 2017; Amir, 2015; Schott & Sedaghat, 2014). However, the impact of innovation on firm emergence as well as the survival rate of a firm remains theoretically as well as empirically inconclusive.

Innovativeness was found to be related to positive outcomes such as competitive advantages (Porter, 1980), production cost reduction (Cohen & Klepper, 1996) or enhanced firm market power (Schumpeter, 1934). Such traits are underpinned by a variety of empirical analyses which delineate the positive effect of innovativeness on the firms survival rate (e.g., Cefis & Marsili, 2005, 2006, 2012; Colombelli et al., 2013; Helmers and Rogers, 2010; Wagner & Cockburn, 2010). However, innovative ventures may face a greater liability of novelty (Amason et al.,

2006), longer and uncertain payback times (Boyer & Blazy, 2013) and a more risky and complicated start-up process (Samuelsson & Davidsson, 2009). These outcomes in turn have been negatively associated with the subsequent survival rate (Hyytinen et al., 2015). Such insights are underlined by an emerging empirical strand that highlights a context-dependency in respect of innovativeness and firm survival (Boyer & Blazy, 2013; Buddelmeyer et al., 2009; Hyytinen et al., 2015; Reid & Smith, 2000).

Conversely, the imitation of an established product or successful business model can provide initial legitimacy for new ventures (Aldrich & Fiol, 1994; Dimaggio & Powell, 1983; Fu & Tiez, 2015, 2017), as underlined by numerous prominent examples: when it comes to specific smartphones (e.g., Mi phones), China's Xiaomi is often the target of some people who believe it is good at copying Apple; entrepreneurs considering crowdfunding find it hard to distinguish the innovativeness of Kickstarter (launched in 2009) compared to Indiegogo (launched in 2007); and, Rocket Internet, Europe's most notorious incubator, makes millions of dollars from ruthlessly copying Silicon Valley start-ups. Despite the number of such imitative business, our knowledge of their gestation process remains limited (Koellinger, 2008). Is it truly better to fail in originality than to succeed in imitation as outlined in the initial quote? Essentially, we argue that it is context dependent. It may be true in the long term, but we outline that in the early venture process imitation, in contrast to innovation, may foster the emergence of a new firm and may even be the first step towards originality later on.

Given the importance of different entrepreneurial attributes on innovativeness (Dyer et al., 2008), we further aim to identify who at the early venture stage the imitators are, and who they are not. Drawing on social identity theory, we delineate the likelihood of pursuing an imitative rather than an innovative business idea. Social identity proved to offer valuable insights in answering questions such as *why* difference in behavior, action and performance among entrepreneurs exists (Bryant, 2007; Farmer et al., 2011; Fauchart & Gruber, 2011; Pollack et al., 2015; Powell & Baker, 2017; Tumasjan & Braun, 2012). These identities influence entrepreneurial decisions differently (Shepherd et al., 2015) and have recently attracted the attention within the entrepreneurial field (e.g., Brändle et al., 2018; Murnieks et al., 2014; Powell & Baker, 2014, 2017; Sieger et al., 2016; Xu, 2014). Thus, we suggest that different identity types influence the likelihood of pursuing an imitative vs. innovative business idea.

Building upon a longitudinal dataset of n=180 Swiss nascent entrepreneurs, our results suggest that nascent entrepreneurs who imitate are more likely to achieve the successful birth of a firm compared to innovative entrepreneurs. Moreover, drawing on social identity theory,

our results indicate that Missionaries are less likely to imitate and more likely to create new, innovative product and services.

This article contributes to the literature on nascent entrepreneurship (e.g., Davidsson, 2016; Davidsson & Gordon, 2012), entrepreneurial innovation (e.g., Audretsch et al., 2008; Koellinger, 2008; Shane & Venkataraman, 2000), and social identity theory (e.g., Brändle et al., 2018; Powell & Baker, 2014). Specifically, we extend the literature on nascent entrepreneurship by providing empirical evidence about the positive relationship between imitation and firm birth. By measuring imitation, respectively innovation ex ante, we follow Hyytinen et al. (2015) in their call to tackle selection and survival bias which have plagued past studies. We further contribute to the emerging strand of entrepreneurial innovation that suggests a context-dependent analysis of innovation. In the context of the nascent start-up phase, we shed new light on the theoretical tension outlined above and propose, based on possible liabilities of new firms (Stinchcombe, 1965), a positive relationship between imitation and firm birth. Thus, we contribute to the emerging, though not less important, research stream; that is the benefits that imitative business ideas may bear during the early entrepreneurial process. Lastly, we contribute to the growing literature on social identity theory by advancing research linking identity types with imitative and innovative ventures (Brändle et al., 2018; Fauchart & Gruber, 2011; Hoang & Gimeno, 2010). We show that founder identity types, especially the Missionary identity, influence the likelihood of an emergence of an innovative business venture.

The remainder of the paper is organized as follows. Section 2 outlines how imitation compared to innovation can increase the likelihood of firm birth. Then, we draw on social identity theory in order to further narrow down the profile of imitators and innovators. In section 3, we describe the database used in this study and present the statistical procedures that are used to test the hypotheses. In section 4, we present our empirical results, and, in the last section, we discuss the implications for research and practice before concluding with limitations and directions for future research.

4.3 Theoretical background and hypotheses development

4.3.1 Innovation and imitation in the firm creation process

The transition from merely a progressing nascent venture towards the establishment of a firm is considered one of the most important research topics for entrepreneurship (Davidsson, 2016). A crucial factor influencing this venture creation process is the business idea and with it the

possible innovation that may provide the entrepreneur with a competitive advantage (Cefis & Marsili, 2012; Dahlqvist & Wiklund, 2012; Schumpeter, 1934). However, this business idea does not always have to be innovative in order to drive the entrepreneurial process (Davidsson, 2016). Various practical examples (Indiegogo vs. Kickstarter, Alando vs. Ebay) show the success and impact imitative businesses can have on the founder, the firm and the wider economy.

Building on Koellinger (2008, p. 23), we define imitative nascent entrepreneurs "as individuals trying to start a business in an established population whose routines, competencies, and offers vary only minimally, if at all, from those of existing organizations." In contrast to innovative entrepreneurs, their venture ideas only differ marginally from those of existing ventures (Aldrich & Martinez, 2001; Samuelsson & Davidsson, 2009) and do not reflect any innovativeness on a technological, market or customer level (Koellinger, 2008). The impact of such imitative businesses is manifold, making them an essential part of the entrepreneurship concept. Next to economic and employment related aspects, these businesses provide not only consumers with additional choices but may alter also competitive behaviors of incumbent firms, thus fostering further innovation (Davidsson, 2016).

In the context of nascent entrepreneurship, we know that the process of new venture creation differs between innovative and imitative ventures (Koellinger, 2008; Samuelsson & Davidsson, 2009). However, the performance effects of imitation are much less studied (Ethiraj & Zhu, 2008) and in a nascent entrepreneurial context few attempts have been made to distinguish between innovative and imitative venture ideas (Dahlqvist, 2007; Samuelsson & Davidsson, 2009). In an organizational context, the prevailing view in the empirical literature outlines a positive association between innovativeness and firm performance as well as venture survival rate (Hyytinen et al., 2015). However, recent research also indicates that innovativeness appears to be very much context dependent (Boyer & Blazy, 2013; Hyytinen et al., 2015; Rosenbusch et al., 2011) and that factors such as firm size, firm age, innovation type and cultural difference impact the innovation-performance relationship (Rosenbusch et al., 2011). In the context of micro companies, innovative firms may even face a lower survival rate compared to non-innovative firms because of a variety of factors such as professional experience or higher uncertainty about the entrepreneur's perception of his survival propensity (Boyer & Blazy, 2013). This negative effect of innovation on startups may be explained through liability of newness (Stinchcombe, 1965), liability of smallness (Freeman et al., 1983) or/and the risk attitude of an entrepreneur (Hyytinen et al., 2015). Liability of newness refers to the lower survival probability of a new market participant due to less experience and less legitimacy in

the eyes of potential financiers. Liability of smallness refers to problem of the size of the company and thus of not being able to attract human capital or other types of resources. Innovative ventures appear to follow a less linear start-up process and are in need of more efforts to legitimize the emerging venture (Samuelsson & Davidsson, 2009). They must contend with being both new and different (Amason et al., 2006) and face a customer reaction that is highly uncertain (Utterback, 1996). Therefore, they face harder market conditions, making the venture creation, especially in an early stage, more difficult. Moreover, innovativeness impacts the risk profile of the venture in such a way, that revenue streams are more variable and skewed in distribution (Scherer & Harhoff, 2000) and are frequently associated with long and uncertain paybacks, which may complicate the emergence of a new firm.

These liabilities and risks are less acute for imitative business ventures. Compared to innovation, imitation builds on an established market with a known demand for specific products (Koellinger, 2008). Imitators can observe the performance of competitors and the behavior of the customers and can strive faster through this knowledge. Imitative entrepreneurs face especially in the early stage lower financial risks given the lower investment needs in research and development and therefore making the possible outcome less uncertain. Thus, we hypothesize:

H1: Nascent entrepreneurs with an imitative approach are more likely to achieve firm birth.

Despite the outlined positive aspects, we underline that imitation itself may indeed not be as simple as described. Imitation follows a different creation process (Samuelsson & Davidsson, 2009) that may require a different set of skills. Imitative entrepreneurs face different degrees of financial and technological risk (Koellinger, 2008). For example, once entered the market, competitors may react on the imitator, influencing their respective performance. The key distinction we want to make here is that especially in an early stage, until the birth of a firm, imitative business ideas may face less risks compared to novel and innovative ideas.

4.3.2 Social identity theory and imitation / innovation

Besides the research on personality traits (Rauch & Frese, 2007) or social networks (e.g., Muller & Peres, 2018; Schott & Sedaghat, 2014), in recent years, scholars have started to focus on cognition theories in order to explain variations in entrepreneurial behavior and action (Grégoire et al., 2011; Randolph-Seng et al., 2015; Sassetti et al., 2018). Thorgren and Wincent (2015) found that serial entrepreneurs experience a higher level of passion for entrepreneurial

activities. In a similar vein, Murnieks et al. (2014) examined the impact social identity has on entrepreneurial behavior, outlining the role of passion. Overall, social identity provides valuable insights in answering questions such as *why* differences in action, duration and performance among entrepreneurs exist (Bryant, 2007; Farmer et al., 2011; Fauchart & Gruber, 2011; Pollack et al., 2015; Powell & Baker, 2017; Tumasjan & Braun, 2012). To add to this stream of research, we relate social identity theory to the innovative or imitative approach of nascent entrepreneurs.

Identity theory aims to explain the human "self" and behavior in terms of peoples' identities (Powell & Baker, 2017). Founder identities shape the goals and strategies around which the firms were initially founded and play a central aspect in the behavioral process that drives venture creation (Cardon et al., 2009). Fauchart and Gruber (2011) proposed a categorization into three founder identity types: "Darwinian" (focus on the financial success), "Communitarian" (contribution to a community as central aspect) and "Missionary" (focus on understanding the firm as an agent of change). We propose that these identity types have a significant influence on the likelihood of following an innovative or imitative business idea.

Consider Darwinians and the example of the Samwer brothers and their company Rocket Internet, a leading European incubator that builds start-ups and is often considered a copycat factory (Rest, 2013). This approach emphasizes a generic perspective of firm creation. In an published online interview (Heeg et al., 2016), Oliver Samwer, one of the co-founders, stated: "We are not the Einstein of the internet. But at the end of the day no one asks who invented something, only who is the market leader." This quote exemplifies the fact that these entrepreneurs see their reference point in the competitive environment and the financial benefits of being market leader rather than adopting an innovative behavior. Such competitive driven behavior fits the Darwinian profile and as shown by the Samwer brothers and Rocket Internet, may increase the likelihood to start an imitative venture by identifying a well-established market. Furthermore, Rocket Internet and their founders, have often attracted negative publicity because of their harsh work environment (Hüsing, 2011). Such behavior towards employees may also be linked to the attitude of Darwinians, who often feel responsible only for themselves and experience less likely negative emotions (Sieger et al., 2016). This trait may further influence the attitude in feeling less guilt when it comes to imitating a competitor as long as the financial return is acceptable. Hence, we posit that Darwinians are less likely to invent. Instead, they build upon existing solutions and try to generate a profit by tweaking existing product and business models to common needs, thus rather implementing imitative business ideas. Therefore, we hypothesize:

H2: An entrepreneur with a Darwinian identity type is more likely to follow an imitative venture approach.

Conversely, Communitarians employ their prior knowledge to solve a problem they identified within their community. As an example, consider Airbnb. The idea came up as the founders Brian Chesky and Joe Gebbia saw that for a conference in San Francisco, participants could not find a place to stay as all hotels were booked out. As at that time they could not afford the rent for their loft, they put an air mattress in their living room to start generating extra revenue by turning their apartment into a bed and breakfast. Building upon the problem for the "conference community" of finding a place to stay, as well as their own issue of basically "surviving" financially, they started to build an innovative platform which could offer this service to a wider community (McCann & Chris, 2018). The Airbnb founders reflect to a certain extent typical attributes of a Communitarian type. Communitarians appear to relate to innovation-related problem solving (von Hippel, 1994) and improve the quality of the community they serve and to whom they feel responsible (Fauchart & Gruber, 2011). To a large extent, this captures the attitude of the Airbnb founders who, besides their own financial dilemma, considered the community of conference attendees. As exemplified by Airbnb, Communitarians produce first for themselves and then realize the existence of a market, after demand from their community gains momentum (Fauchart & Gruber, 2011). By solving such problem for themselves, they primarily seek a creative way of solving a wider problem at hand. They may rely on existing solutions, like the concept of "bed and breakfast" in the Airbnb example, however, they apply them in a different context, innovate from the community perspective and create an entire new industry. Hence, we hypothesize:

H3: An entrepreneur with a Communitarian identity type is less likely to follow an imitative venture approach.

Missionary entrepreneurs are driven by the desire to improve the world and the well-being of the society as a whole (Fauchart & Gruber, 2011). The goals Missionaries want to achieve are of a much larger scale than the other types of entrepreneurs and they require a considerable amount of effort in overcoming the challenges ahead. Consider Ethan Brown, founder of Beyond Meat, as a possible example of this identity type. Beyond Meat produces a variety of protein-based food products with the goal to move away from animal meat (Beyond Meat Inc., n.d.). In an interview, Ethan Brown mentioned four main reasons for building the company: first, there is a global problem with the supply of livestock; second, there are health issues with

eating red and processed meat; third, there is the climate effect due to resource demand and fourth, there is the animal welfare issue (Park, 2019). Solving these huge problems resemble the attitude of a Missionary, who as outlined above considers the firm as an agent of change. Moreover, such visions are often linked to changes in the consumption pattern of consumers (Fauchart & Gruber, 2011) and thus may involve a new and innovative perspective in solving an identified problem or addressing a new opportunity. This reflects in our opinion well the way, how Beyond Meat tries to disrupt the market, enticing consumers to move away from animal meat. Missionaries provide solutions used for "better practices" that have an inspirational and highly impactful character (Fauchart & Gruber, 2011). These traits are largely congruent with the concept of innovation, following the aspirations towards newness from a customer, market and technological perspective (Koellinger, 2008). Ethan Brown tries precisely to innovate on all these areas with his "meatless" burger. For customers, Beyond Meat changes consumption patterns towards a positive future, within the market it alters the way competitors think about their products, and from a technological level, this entrepreneur introduces a new and innovative product that may improve society as a whole. Thus, we hypothesize:

H4: An entrepreneur with a Missionary identity type is less likely to follow an imitative venture approach.

4.4 Methodology

4.4.1 Sample and data collection

We draw on data from the Swiss Panel of Emerging Businesses (SwissPEB), a longitudinal study of Swiss nascent entrepreneurs which follows the research design of the Panel Study of Entrepreneurial Dynamics (PSED) (Reynolds, 2016). The study covers three waves of interviews. The wave 1 interview was conducted in 2015, wave 2 in 2016, and wave 3 in 2017. Nascent entrepreneurs were defined and identified as individuals who are currently actively trying to start a new business, which they will own or co-own and which is still in the pre-profit stage. In 2015, a total of 28,599 individuals in Switzerland were screened in a telephone interview by a professional market and social research firm. Of the 28,599 people screened, only 680 qualified as nascent entrepreneurs (2.4%), of which 298 participated in the first detailed phone interview. After twelve months, 292 of these nascent entrepreneurs were contacted again. 67 did not participate in the second round, leaving a sample of 225 in wave 2. The rate of attrition of roughly 23% in the Swiss sample is similar to previous longitudinal

studies of nascent entrepreneurs (e.g., Brush et al., 2008; Manolova et al., 2012). Due to missing values, the analysis of this paper is based on the insights from the outcome of a subsample of 180 observations for wave 2 and 158 for wave 3. The data that support the findings of this study are available from the corresponding author upon reasonable request. This study uses a series of binary logistic regressions to analyze on the one hand the effect of imitation on firm birth and on the other hand the effect of different founder identities types on the likelihood of pursuing an imitative or innovative business idea.

4.4.2 Common method bias

Following Hyytinen et al. (2015) our results use an ex post measure for imitation/innovation. As imitativeness is measured as a binary variable (1 = imitative and 0 = innovative), our model defines imitation as being neither novel from a technological nor from a customer perspective, while there are a few or many competitors already providing the same service or product. Given the longitudinal design of the study, we could track the outcome of the nascent entrepreneurs over a period of 2 years. We additionally tested the impact of the imitative variable on the outcome of year two and derive at the same results ($OR=2.78$, $p<0.05$). Henceforth, we believe that common method bias with regards to the outcome variable firm birth is very unlikely to be a problem in our study, most important, because we measure our dependent and independent variable at different points in time.

Looking at the relationship between imitation and the different identity types a variety of ex-ante steps as well as ex-post analysis were taken. First, given that in self-report survey research, social desirability bias is always a potential concern, strict confidentiality as well as the protection of the respondent anonymity was assured in order to decrease respondents' motivation to answer the survey in a socially desirable way (Podsakoff et al., 2003). Second, the question related to the constructs were embedded in a more comprehensive survey making it impossible for the respondent to draw any conclusions about the research goal. Third, we conducted a SEM (CFA) for our social identity variables (see Chapter 4.5.3) which provided acceptable goodness of fit statistics.

4.5 Measures

4.5.1 Firm birth

We measured the birth of a firm as taking place over a period of two years by a self-reported categorical variable as either being at the time of the follow-up interviews still nascent

("ongoing gestation"), having abandoned the venture or having achieved initial profits. Following the PSED design, achieving initials profits is used as proxy for firm birth (Reynolds, 2017). In order to identify the impact on firm birth we used initial profits as the reference variable and followed Gartner and Liao (2012) coding it binary for 1 having achieved initial profits and 0 otherwise (either abandon or ongoing gestation). We measured the outcome in the second and third interview round, while the independent and control variables were measured in the first interview round. Table 5 summarizes the outcome for each wave.

Table 5. Start-up status

	t_{12}	t_{24}
Profitable (firm birth)	27	39
Ongoing gestation or abandoned	153	119

4.5.2 Imitation / innovation

The degree of imitativeness and innovativeness is based on five items that have also been used within the Global Entrepreneurship Monitor (Koellinger, 2008; Reynolds et al., 2005). Table 6 details the questions and possible answers. We consider in our analysis the cases with the perception of few as well as many business competitors, no customer newness, low technological newness, no R&D focus and not being considered as a hi-tech firm as imitators. This conceptualization is mainly based on Koellinger (2008) with the exception of also having considered imitation within a market with few competitors as imitators. The variable was coded binary, 1 for being an imitator and 0 otherwise. In total, we categorized 42 ventures as pure imitators and 138 ventures as otherwise, thus being to a certain extant innovative from either a market, customer or technological perspective.

Table 6. SwissPEB survey questions on innovation and imitation

Survey question	Answer categories
C - "Is the product or service for all, some, or none of your potential customers a real innovation?"	C1 - All C2 - Some C3 - None
M - "Right now, are there many, few, or no other businesses offering the same products or service to your potential customers?"	M1 - Many business competitors M2 - Few business competitors M3 - No business competitors
TY - "Were the technologies or procedures required for the products or services of your company generally available for less than a year, between one and five years or more than five years ago?"	TY1 - Less than a year TY2 - From one to five years TY3 - More than five years
TH - "Would you consider to be a hi-tech company?"	TH1 - Yes TH2 - No
TRD - "Will spending on research and development be a major priority for your company?"	TRD1 - Yes TRD2 - No
Imitation / innovation measurement	
Pure imitation	C3, M1, M2, TY2, TY3, TH2, TRD2
Innovation (any kind)	Every other combination

4.5.3 Social identity theory

For each founder identity type we used three items based on the theory of Fauchart and Gruber (2011) and build on an early adaption of the developed items based of the paper of Sieger et al. (2016). To derive at the identity types the mean of the three types was calculated. The evaluations were captured through a Likert-type scale ranging from 1 to 5 points (where 1 = Strongly disagree and 5 = Strongly agree). We conducted a confirmatory factor analysis. Goodness of fit statistics (Hooper et al. 2008) for the constructs provide an acceptable model fit (RMSEA= 0.058, CFI= 0.966, TLI= 0.944, GFI=0.979). The list of items is provided in Table 7.

Table 7. Founder identity scale

For the following questions, please reply on a scale of 1 to 5. 1 means that the statement does not apply at all and 5 means that it fully applies.
Darwinian
a) I am creating my firm in order to make money and become rich.
b) As a firm founder, it is very important to me to operate my firm on the basis of solid management practices.
c) When managing my firm, it will be very important to me to perform better than the competitors.
Communitarian
a) I am creating my firm in order to solve a special problem of a group of people that I strongly identify with.
b) As a firm founder, it is very important to me to provide a product/service that is useful to a group of people that I strongly identify with.
c) When managing my firm, it will be very important to me to have a strong focus on the group of people that I strongly identify with.
Missionary
a) I am creating my firm in order to play a proactive role in changing how the world operates.
b) As a firm founder, it is very important to me to contribute to make the world a "better place".
c) When managing my firm, it will be very important to me to focus on what the firm is able to achieve for society-at-large.

4.5.4 Control variables

Personal and market related aspects influence the imitativeness vs. innovativeness of a start-up (Koellinger, 2008). We controlled for a variety of variables at the individual level (gender, age, education, full-time or part-time employed, and previous work as well as start-up experience), venture (venture age, venture team, service and manufacturing industry) and market level (perceived competitive environment).

4.6 Analysis and results

4.6.1 Descriptive statistics

Table 8 shows the mean, standard deviation, and correlation of the variables under research. The correlations indicate that imitation is positively correlated to firm birth over both periods (t_{12} and t_{24}). Moreover, there is a negative relationship between venture age and imitation which may indicate that the longer the gestation process, the lower the likelihood of perusing an imitative business idea. Looking at the identity types, there exists a significant positive correlation between Darwinians and the venture team as well as the gender, outlining a positive relationship between males and Darwinian and a positive relationship between females and Missionaries. Finally, there exists a significant negative correlation between Communitarians and Missionaries on imitativeness. Our data did not suffer from multicollinearity, as all variance inflation factors (VIFs) were below 3.2. This was well below the acceptable threshold of 10 (Hair et al., 2010).

4.6.2 Hypothesis testing

We tested the hypotheses using a series of binary logistic regression models (Table 9). Model 1 shows the model including the control variables only for *H1*. Model 2 shows the full model including the main effect of imitativeness on the likelihood of firm birth (relating to *H1*). Model 3 shows the model including control variables only while using imitativeness as dependent variable. Model 4 uses the full model including the three different social identities (for testing *H2-H4*).

The -2 log likelihood ratio as well as the Nagelkerke R^2 provide an indication of how well the model fits the data. Imitativeness is able to increase the model fit ($p<0.01$) as well as the models' explanatory character (Nagelkerke R^2 increase of roughly 4%). The Hosmer–Lemeshow test is a statistical test for goodness of fit for logistic regression models. It shows a $p>0.05$ in all models and thus does accurately predict the likelihood of firm birth or imitation. Moreover, we report the odds ratios (OR) rather than the coefficients b as they are more intuitively interpretable in terms of the increase/decrease in the odds of achieving firm birth or not and pursuing an imitative idea or not. The results are presented in Table 9.

Table 8. Descriptive statistics and correlations

		Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	Firm birth (t_{12})	0,15	0,36	1																
2	Firm birth (t_{24})	0,25	0,43	,79**	1															
3	Imitation	0,23	0,42	,25**	,27**	1														
4	Darwinian	3,23	0,86	0,07	-0,06	-0,02	1													
5	Communitarian	3,82	1,1	0,03	0,04	-,16*	0,08	1												
6	Missionary	3,22	1,28	-0,02	-0,01	-,25**	0,11	,36**	1											
7	Venture age[a]	35,23	21,67	-,23**	-,27**	-,20**	0,04	0	0,09	1										
8	Venture team[b]	0,48	0,5	-0,09	-0,06	0,02	,19**	-0,04	-0,06	-0,12	1									
9	Service industry[b]	0,75	0,43	0,03	0,05	-0,05	-0,08	0,07	0,06	0,05	-,19*	1								
10	Manufacturing industry[b]	0,13	0,34	0,06	0	0,09	0,01	-0,14	-0,01	-0,04	0,11	-,68**	1							
11	Age[c]	41,57	10,78	0,03	0,06	0,12	-0,04	0,11	0,05	0,13	-,20**	,17*	-0,02	1						
12	Gender[d]	0,59	0,49	0,03	0,11	0,03	,32**	-0,12	-,16*	-0,09	,32**	-,27**	,22**	-0,14	1					
13	Tertiary education[b]	0,52	0,5	-0,1	-0,14	-0,08	0,1	0,13	0,07	,19*	-0,01	,17*	-,15*	,22**	0,09	1				
14	Full-time or part-time employed[b]	0,81	0,4	0,09	0,13	0,11	0,03	0,07	0,02	,16*	0	0,01	-0,01	0,12	0,14	0,06	1			
15	Work experience[c]	10,73	11,01	0,09	0,15	0,11	-0,05	0,13	0,03	-0,02	-0,14	,16*	0,02	,51**	-0,02	,17*	0,05	1		
16	Start-up experience[b]	0,4	0,49	0,04	0,06	0,01	0,07	0,12	-0,02	-0,11	,19*	0,03	0,01	0,07	,17*	0,12	0	,19**	1	
17	Perceived competition[e]	2,57	0,98	0,01	-0,02	0,01	0,06	-0,12	-0,14	0	0,1	-,18*	,18*	-0,09	,23**	-0,13	0,12	0,05	-0,01	1

*Notes. N=180 (t_{12}), N=158 (t_{24}), a: in months, b: 1 = yes, 0 = no, c: in years, d: 1 = Male 0 = Female, e: Likert scale from 1-5 (1=high level, 5=low level), **p < 0.01, *p < 0.05, Significance levels are two-tailed.*

Model 2 confirms hypothesis 1 that nascent entrepreneurs with an imitative strategy achieve firm birth with a higher likelihood compared to entrepreneurs pursuing innovative ideas (OR = 3.06, $p<0.05$). The probability (3.06/(1+3.06)) indicates that imitative entrepreneurs are 75.3% more likely to achieve firm birth. Model 4 tested hypotheses 2, 3 and 4. We did not find a significant association between Darwinian (OR = 1.12; $p=0.67$), respectively Communitarian (OR = 0.75; $p=0.12$) on the likelihood of an imitative strategy. Hence, we could not find any support for *H2* and *H3*. Even though not significant, it should be noted the OR for Darwinians is above one, which suggests an increased likelihood of being involved with an imitative business idea while for Communitarians the OR is below one, which suggests a decreased likelihood of being involved with an imitative business idea.

We find support for *H4* that Missionaries are less likely to pursue an imitative venture (OR=0.63; $p<0.05$). This indicates that an increase in a Missionary identity type decreases the odds of imitating a business by 39%.

Finally, with regards to the *H2-H4*, we ran the models for each single identity type as well as the combination of just two identity types. The results for the Missionary and Darwinian type did not change, neither at the single nor at the joint level. Yet, including only the Communitarians, the results show a significant value (OR = 0.64; $p<0.05$), thus confirming to a certain extent *H3*. However, it appears that the correlation between Communitarian and Missionary renders this relationship insignificant in the full model. A possible explanation is that identity types can also occur in a mixed form (Fauchart & Gruber, 2011). Nevertheless, the coefficients are negative for both the Communitarians and the Missionaries in the full as well as the single models, indicating a relationship as proposed in *H3* and *H4*.

Table 9. Results binary logistic regression

	Model 1		Model 2		Model 3		Model 4	
	Firm birth$_{t12}$				Imitation			
Variables	OR	SE	OR	SE	OR	SE	OR	SE
Constant	0.18	1,66	0,24	1,7	0,14	1,22	0,17	1,56
Control variables								
Venture level								
Venture age	0,95**	0,02	0,96**	0,02	0,97**	0,01	0,97**	0,01
Venture team	0,41†	0,52	0,37†	0,53	1,08	0,41	1,11	0,44
Service industry	3,3	1,11	4,22	1,17	0,97	0,64	0,98	0,68
Producing industry	3,93	1,21	3,92	1,27	1,63	0,76	1,9	0,81
Individual level								
Age	1	0,03	0,99	0,03	1,03	0,02	1,03	0,02
Gender	1,49	0,55	1,62	0,56	1,03	0,44	0,76	0,49
Tertiary education	0,59	0,5	0,58	0,51	0,73	0,41	0,8	0,44
Full-time or part-time employed	2,73	0,68	2,39	0,69	2,42	0,54	3,02†	0,58
Industry experience	1,01	0,03	1,01	0,03	1,01	0,02	1,02	0,02
Start-up experience	1,25	0,5	1,46	0,52	0,82	0,41	0,87	0,44
Market level								
Competitive environment	0,89	0,24	0,88	0,25	0,92	0,21	0,83	0,22
Main effects								
Imitation			3,06*	0,5				
Social identity								
Darwinian							1,12	0,26
Communitarian							0,75	0,18
Missionary							0,63**	0,17
Model fit								
LR chi-squared (df)	23,08* (11)		31,01** (12)		18,26† (11)		32.77** (14)	
-2 Log likelihood	129,01		124,20		177,32		162.81	
Nagelkerke R^2	0,21		0,25		0,15		0.25	
Hosmor and Lemeshow Test	9,08 (p=0,34)		5,92 (p=0,66)		5,67 (p=0,68)		9,09 (p=0,34)	

*Notes. N = 180, ** p<0.01, * p<0.05, †<0.1, Significance levels are two-tailed.*

4.7 Discussion and conclusion

4.7.1 Summary and discussion

Innovation and imitation are two strategic positioning opposites, impacting the performance as well as the survival rate of a firm (Buddelmeyer et al., 2009; Cefis & Marsili, 2006). Even though innovation is an essential part for a firm and an economy to thrive, the possible positive effect appear to be context-dependent and may even turn negative (Boyer & Blazy, 2013; Hyytinen et al., 2015). Building up on this context dependency, this paper outlines that in a nascent phase an imitative behavior tends to increase the likelihood of firm emergence. Imitative ventures may be less laden with liabilities of newness and smallness. They face less uncertainty given the already established market and demand for their products and services, thereby increasing the possibility of achieving profits and firm birth.

Our results are in line with previous studies who look at possible drawbacks of entrepreneurial innovation. This strand of literature identified that firms exploiting new market knowledge are less likely to attain substantial growth compared to firms exploiting existing market knowledge (Saemundsson & Dahlstrand, 2005). It adds to the evidence of an increased survival rate, especially at the earliest venture stage, of non-innovative young firms (Boyer & Blazy, 2014; Hyytinen et al., 2015).

A second major insight of this study is a characterization of the impact of an entrepreneur's social identity, more specifically a Missionary identity type, on the likelihood of innovating, respectively not imitating a business idea. Our results indicate that Missionaries are less likely to purely imitate and thus are more likely to innovate on either a customer, market or technological level. We explain this by the problems Missionaries try to tackle. Their urge to change current consumption behaviors and to "make the world a better place" (Brändle et al., 2018, p. 19) drives them towards the exploration of innovative solutions. This motivation towards the implementation of new social practices (Fauchart & Gruber, 2011) does not fit an imitative strategy and thus possibly rather drives such entrepreneurs towards an innovative behavior. In a similar vein, looking at the Communitarians detached from the other two identity types, they rather innovate to create value for a certain community than purely imitate. Darwinians on the other hand appear not to relate to an imitative/innovative behavior. This non-finding could be possibly explained by their strong pursue of private self-interest (Sieger et al., 2016) that could, given different situations and venture types, either drive them

towards innovating or imitating. We suggest further research to empirically assesses this relationship.

4.7.2 Theoretical and practical implications

Our study makes several theoretical and practical contributions. From a theoretical point of view, we contribute to the emerging strand of innovation literature (Boyer & Blazy, 2013; Buddelmeyer et al., 2009; Hyytinen et al., 2015; Reid & Smith, 2000) that calls for a more context dependent analysis. We provide evidence that in a nascent phase, imitative business ideas are possibly more likely to lead to firm birth than the innovative ones. Our results contribute to this literature stream by showing how ex ante measured imitativeness influences the outcome of a venture over time, thus rendering the survivorship bias of ideas very small. In addition, while recent studies outlined the impact of innovation/imitation on the survival of a venture, our study examines the impact of innovation/imitation on firm birth, an essential first step in achieving business profitability and growth (Reynolds, 2016). Moreover, by laying the focus on imitation, we further take into account the importance that not only innovation but also imitation (Koellinger, 2008) plays within the venture creation process.

Furthermore, we contribute to the growing strand of social identity theory within the entrepreneurial literature. For example, Hoang and Gimeno (2010) delineate that a founder's identity influences the transition from exiting the current job into undertaking founding activities. Pan et al. (2019), Wry and York (2017) or York et al. (2016) outline how a 'pro-social identity' as consistent with the "Missionaries" increases the engagement in social entrepreneurial ventures such as environmental entrepreneurship. Thus, by linking the Missionaries to innovative behavior, we empirically tested another attribute that further explains differences in entrepreneurial behavior.

From a practical perspective our results suggest that entrepreneurs following an imitative strategy in an early stage may increase their likelihood of firm birth. We want to note here that even though an imitative strategy may appear beneficial, innovative behavior still may be in a later stage of the entrepreneurial endeavor valuable in order to overcome the competitive market forces, thus call for the benefits of innovations (Cefis & Marsili, 2005, 2006, 2012; Helmers & Rogers, 2010; Wagner & Cockburn, 2010). Nevertheless, our aim was to illustrate that it is important to consider not only the role of innovative but also imitative entrepreneurs, who can make a significant contribution towards economic growth. Moreover, our findings have implications from a wider public policy perspective. Given the public goal to support and

subsidize startups' innovativeness, an inclusion of social identity types could be used as an additional selection criterion in directing respective investments. Specifically, Missionaries appear to be drivers of innovation (e.g., market, consumer or technological level) and thus may foster desired societal change, a widespread policy goal. Finally, public support may sometimes be driven by the notion that innovative firms are more likely to survive and to create stable jobs (Hyytinen et al., 2015). Our analysis casts some doubt on this premise, indicating that, in the early start-up stage, innovative ventures are probably more in need of support in mastering the first milestone, that is the birth of the firm.

4.7.3 Limitations and future research

The particular strength of the data is its longitudinal character as well as the focus on entrepreneurs in a nascent phase. We measured imitation/innovation ex ante allowing for a clearer analysis of the outcome of firm birth. However, the outcome variable as well as the measurements used for imitation/innovation are based on the judgments of individuals. Thus, we believe that further research drawing on a more objective and unifying measure for either imitation or innovation is desirable.

Moreover, even though we controlled for service and manufacturing industry types and included in our measurement variable market as well as technological facets, we are limited regarding differences of imitativeness and innovativeness related to further industry and venture characteristics. As such one can argue that a consulting firm or even a hairdresser is way faster in achieving firm birth compared to an agricultural venture type. Both may be considered imitators, however given the higher resource intensity, the former may be faster in achieving firm birth. Thus, further analyses of imitative approaches with firm birth as an outcome within different industry and venture types could be insightful. Furthermore, the sample of imitative entrepreneurs is relatively small compared to the innovative ones, thus a bigger sample size also covering a longer research period may provide additional insights.

Finally, even though our measurements for the different social identity types show acceptable goodness of fit statistics, the scale is still of exploratory character. Therefore, further empirical analyses of the social identity based on a clear and uniform scale (cf. Sieger et al., 2016) is eligible. Given the impact different social identity types can have on the venture process (Brändle et al., 2018), we believe further research would be fruitful in order to nurture a better understanding of the entrepreneurial behavior and action.

5. Predicting the venture gestation outcome with machine learning methods and artificial neural networks

5.1 Abstract

In light that most ventures do not get past the earliest stage of the firm creation process within the first six years, the related costs for all involved parties are enormous. This study explores a diverse set of machine learning methods, as well as the applicability of neural networks to forecast the likelihood of firm birth as well as firm abandonment after five years in a sequential order. The aim of this study is to provide at the early stage of the venture creation process, where uncertainty prevails, a decision support system that allows involved parties such as investors, advisors and support programs to identify entrepreneurs who are successful in achieving firm birth and less likely to abandon the venture. Thus, the proposed systems allow for optimization of resource allocation in start-ups. In doing so, this study further compares the predictability of artificial neural networks with other approaches such as logistic regression, k-nearest neighbors, decision tree, random forest, XGBoost and support vector machines. The results show that for the applied dataset neural networks can provide a comparable performance in predicting the likelihood of firm birth and firm abandonment.

5.2 Introduction

Launching a new business venture requires a lot of time and effort. According to one estimate, the amount of money invested by angel investors into start-ups in the US equaled more than $26 billion in 2018 (WBAF, 2020) and the time entrepreneurs devote to starting new firms amounted to around 2.7 percent of total paid work in 2005 (Reynolds & Curtin, 2007). The related sunk costs for the stakeholders in start-ups and for the economy appear to be correspondingly high considering that within the first six years of a new business creation the majority of entrepreneurs give up their business idea (Reynolds & Curtin, 2009). Business angels, venture capitalists and incubator managers continuously face the decision to put extra time, money and effort into entrepreneurs and their fledgling business ventures, but how do they know whether entrepreneurs are likely to launch a profitable business venture, abandon their business project, or possibly spent years pottering about a business idea that will never come true?

A possible misallocation of resources does not only impact the direct loss of the related party but also the loss of opportunity of further supporting ventures who for example directly

contribute via job creation to economic growth. As agency theory and resulting information asymmetry suggest, stakeholders in entrepreneurship face particularly informational risk and should be cautious about how they allocate their resources (Epure & Guasch, 2020). As they operate at an early stage of the entrepreneurial process, they have to deal with uncertainty (L. Huang & Pearce, 2015) and where a prediction could possibly support and optimize the resource allocation process. Thus, predicting the likelihood of firm birth and firm abandonment of a nascent entrepreneur, the person undertaking activities to create a business, during different times of the gestation – conception to birth – process (Reynolds & Miller, 1992) benefits a variety of involved parties, including investors, advisors as well as private and public support programs.

Albeit such relevance, predicting at the earliest stage of the entrepreneurial journey the outcome of a venture is an extremely complex topic given the missing financial information and the breath of influencing factors on a market, industry, venture and individual (e.g., cognitive and motivational) level. Over and above, the entrepreneurial journey itself is a complex and dynamic process (Lichtenstein et al., 2007) which evolves over time, while peaking at a first stage (Reynolds, 2017) in one of three possible outcomes, that is firm birth, firm abandonment and ongoing gestation. Given the relationship of a variety of factors between the birth of a firm, the persistence to stick with the venture (termed in this paper as ongoing gestation) or to abandon the firm, an overall understanding of the influencing parameters is essential.

Taking into account these aspects, this paper provides in the context of the gestation process an overview of empirically up-to-date identified influence factors and tries to predict the likelihood that a nascent entrepreneur achieves firm birth or abandons the venture. Such prediction can support and guide the decision-making process as well as reduce future arising costs for the involved parties during various stages of the gestation phase. Thus, the main contributions of this work are:

- Summarize and provide an overview of key success and failure factors within the gestation process based on recent empirical results related to the venture outcome;
- provide a decision support system via predicting the likelihood of firm birth and firm abandonment after a period of 5 years, while taking into account a diverse number of variables measured in a yearly sequential order. In the literature, the proposed existing models dealing with this prediction problem usually either account for one period, only

include a limited number of factors that influence the profitability of a venture or are mainly applied on established businesses;

- evaluate the performance of artificial neural networks to predict the likelihood of firm birth and firm abandonment and compare artificial neural networks to statistical models such as the logistic regression (LR) and five other data mining techniques (KNN, DT, RF, XGBoost, SVM) to evaluate the performance of more advanced methods in an entrepreneurial context;
- evaluate the feature importance of the firm birth and firm abandonment prediction models over the different sequences.

The remainder of the paper is organized as follows. The next session provides a review of i) previously identified success and failure factors; ii) used data analysis techniques for predicting the venture outcome. Section 2 describes the dataset, applied methodologies with the conceptual framework and the evaluation methods. Section 3 presents and discusses the results, while the last sections 4 and 5 outline the implications as well as limitations for research and practice before concluding with directions for future research.

5.3 Related work

5.3.1 Meta-analysis of the empirical literature on factors influencing the venture outcome during the gestation phase

What factors exist and can be used to predict the venture outcome? This is one of the key questions that entrepreneurial research tries to solve (Davidsson & Honig, 2003; Pompe & Bilderbeek, 2005; van Gelderen et al., 2005; Westhead et al., 2001). Given that in the early stage of the start-up process financial information and ratios are scarce and are, in light of the focus on the individual level, not well suited to predict the venture outcome, the entrepreneurial literature focused rather on non-financial elements within the early stage of the gestation process (Chrisman, 1999; Reynolds, 2016). In light of the venture outcome, firm birth, ongoing gestation and firm abandonment share directly or inversely a lot of common traits. Thus, in order to derive at a firm birth and firm abandonment prediction model, it is important to understand the factors that possibly influence corresponding outcomes.

The empirical research within the gestation phase largely focused on the US Panel Study of Entrepreneurial Dynamics (US PSED I and II) and related studies such as the Comprehensive Australian Study of Entrepreneurial Emergence (CAUSEE 2007–2013), the Swedish Panel

Study of Entrepreneurial Dynamics, the Norwegian PSED study and the Chinese Panel Study of Entrepreneurial Dynamics. As representative surveys of entrepreneurial activities, these studies portray individuals during their business creation process over 72 months.

The studies that build upon these surveys analyzed a variety of elements such as human and social capital, types (e.g., business planning, marketing activities etc.) as well as occurrence and sequence of gestation activities that impact the likelihood of firm birth and firm abandonment. Table 11 in the Appendix provides an overview of articles published in peer reviewed journals that deal in one way or another with the gestation phase, one of its three outcome variables and uses among others one of the in the previous paragraph described datasets (as at 08.10.2019). Even though the results regarding the factors that impact the venture outcome are not yet completely conclusive, following aspects can be summarized.

The type as well as the amount of gestation activities appear to impact the venture outcome (Newbert, 2005; Tornikoski & Newbert, 2007; Tornikoski & Puhakka, 2009). Next to specific activity types, three factors (time, rate and concentration) have been studied that look at the dynamics of the gestation process (Lichtenstein et al., 2007). On the one hand, results show that the timing, quantified whether the bulk of the organizing activities is accomplished earlier or later during the start-up process, has a positive impact on firm emergence. On the other hand, the rate, defined as the total number of start-up activities undertaken by the particular nascent entrepreneur divided by the duration of the gestation process of the new business, as well as the concentration, quantified as how closely entrepreneurial activities are undertaken in relationship to other start-up activities, on firm emergence remains inconclusive (Hopp & Sonderegger, 2015; Lichtenstein et al., 2007) and thus are possibly in need of further research.

Next to the previously described three complexity dynamics, this paper proposes that the effort the nascent entrepreneur puts into the venture can further indicate the likelihood of firm birth and firm abandonment. This hypothesis originates, among others, from the educational sciences, where the notion of the positive effect of study effort on college success is supported by decades of research (Pu et al., 2020). Transferring the effort effect into an entrepreneurial context, it is proposed that effort can play a crucial role in predicting the likelihood of firm birth and firm abandonment.

Moreover, results regarding the impact of specific activities (e.g., business planning) are inconsistent (Chwolka & Raith, 2012; Daft & Weick, 1984; Delmar & Shane, 2003; Honig & Karlsson, 2004; Honig & Samuelsson, 2012; Liao & Gartner, 2007) and even though some level

of activity is needed, apparently, no particular gestation activities are necessary to achieve firm birth (Arenius et al., 2017; Shim & Davidsson, 2018).

Taking a closer look at single start-up activities, business planning and funding have been closely related to the likelihood of firm birth. Despite the inconclusive empirical evidence (Davidsson, 2015; Delmar, 2015; Delmar & Shane, 2003; Honig & Samuelsson, 2012, 2014) business planning appears to play, if not directly, a role in relation to the venture outcome (Honig & Karlsson, 2004; Liao & Gartner, 2006, 2007)

When it comes to start-up capital (van Gelderen et al., 2005) and funding, external equity as well as the development of the percentage of ownership shares (Hechavarria et al., 2012) influence the firm outcome. For example, it appears that financial capital significantly decreases the odds of discontinuance (Liao et al., 2009). Furthermore, networks and activities that connect the nascent entrepreneur with others appear to positively impact firm birth (Diochon et al., 2005a; Newbert et al., 2013; Newbert & Tornikoski, 2012).

Personal backgrounds (e.g., socio-demographic, work, and career backgrounds) appear not to have an impact on firm emergence (Diochon et al., 2005a). However, they may impact another firm outcome that is entrepreneurial persistence (Freeland & Keister, 2016). Furthermore, individual aspects such as opportunity confidence, industry experience (Dimov, 2010; Hopp, 2012), wealth levels (Frid et al., 2016), age (Liao et al., 2009), household income (Nuñez, 2015a), personal investments (Frid et al., 2015), entrepreneurial experience (Rotefoss & Kolvereid, 2005; van Gelderen et al., 2005) and intentions affect the conducted start-up activities and the respective progress (Hopp & Sonderegger, 2015). Thus influencing the firm outcome, such as by staving off firm abandonment (Delmar & Shane, 2004). Moreover, motivational antecedents such as self-efficacy appear to influence the firm outcome, especially firm abandonment (Hechavarria et al., 2012; Khan et al., 2014)

With regards to the team, initial team size (Chandler et al., 2005), team resource heterogeneity (Muñoz-Bullon et al., 2015) as well as balanced founding team experience (Delmar & Shane, 2006; Thiess et al., 2016) especially with regards to the long term impact (Steffens et al., 2012) positively influence the firm outcome (Nuñez, 2015b). In the context of homogenous teams, also with regards to homogenous start-up experience, there exists a negative relationship towards firm performance in the long term (Steffens et al., 2012). Even though certain controversial results subsist (Tornikoski, 2008) a possible impact of the team on the venture outcome thus may exist.

Additionally, the industry in which the venture acts appears to have an impact on the firm outcome. For example, ventures within the service or manufacturing industry (Steffens et al., 2012; van Gelderen et al., 2005) are likely to outperform ventures started in other industries in terms of being operational or profitable (Steffens et al., 2012).

Finally, the entrepreneurial journey is also influenced by macro-economic factors. Crises appear to have a corresponding impact on the activity as well as the performance, for example through a drastic drop in demand for goods and services (Giotopoulos et al., 2017; Vegetti & Adăscăliţei, 2017).

5.3.2 Literature review of empirical entrepreneurial success and failure prediction models

Predicting the success or failure of a start-up is nothing new. However, even though prediction models in the context of entrepreneurship exist and certain factors appear to allow for a prediction of firm success and failure across different countries (Baidoun et al., 2018; Cooper et al., 1991; Hyder & Lussier, 2016; Lussier, 1995; Lussier & Halabi, 2010; Marom & Lussier, 2014; Reynolds, 1987), these models show a variety of limitations. First, they relate to the later stage of the start-up process, that is the phase from firm birth to becoming an operating business (Reynolds, 2017), and not on the gestation phase, the topic under research within this paper. These studies build up on a firms level with a questionnaire sent out to owners of already established as well as liquidated, Chapter 11 companies (Lussier, 1995)[2]. In that context, different variables (see Chapter 5.3.1) may play a role in the gestation phase compared to either an already liquidated or successfully operating company.

Second, most of these studies use data collected at a single point in time with no information included about the time, duration and event development since the businesses became operational. As the venture process is considered to be a dynamic process (Packard et al., 2017) the application of a longitudinal dataset, as used within this study, better allows to grasp impact factors that affect the likelihood of firm birth and firm abandonment. As such, it could be that based on current success and failure models (e.g., Lussier & Halabi, 2010) a company is deemed successful in one year whereas after 3 years that company may have had to apply for chapter 11, thus impacting the originally identified success factor.

[2] Note that this model (Lussier, 1995) describes success that "the business must make at least industry-average profits. A young firm is defined as being between 0 to 10 years old. (p. 9)" Moreover, failure is defined as "being involved in court proceedings or voluntary action which result in loss to creditors (p. 9)", as such delineating rather an already emerged firm.

Third, around half of the used variables of the above-mentioned prediction models are based on subjective measures and given the ex-post analysis of the business owners, their answers - considering for the possibility of a hindsight bias - could turn out different, thus altering corresponding results. Moreover, as outlined in the literature review, despite the impact of a variety of other factors, only a few prediction models (e.g., Baidoun et al., 2018) include control variables to account for the impact of the venture age or industry differences.

Finally, regarding the methodology, the most common method used in the entrepreneurial literature to predict respective firm outcome in the context of an operating business (e.g., Baidoun et al., 2018; Lussier, 1995) as well as the nascent entrepreneurial gestation literature (see Table 11 in the Appendix) are logistic regression models. Despite the popularity of such statistical methods, with recent development in technology, researchers in various fields have started to use machine learning techniques to predict firm performance (J. Chen et al., 2012; Galbraith et al., 2007; Lee et al., 2017; Loureiro et al., 2018; Moro et al., 2014; Ribeiro & Lopes, 2011). Even though the predictive power, especially with classification systems, seems to relate to the data characteristics (Kiang, 2003), such recent technological models appear to outperform classical models (e.g., logistic or linear regression models) in different ways. Various scholars provide for example evidence of a higher prediction accuracy (Alharthi, 2019; Loureiro et al., 2018; Moro et al., 2014; Ravisankar et al., 2011; Sadat-Hashemi et al., 2005; Williams, 2017) as well as outline the benefit of recognizing the appearance of non-linearity and interactions in input data (Gerasimovic & Bugaric, 2018).

More recently, new technologies, namely artificial and deep neural networks, have provided even higher levels of prediction accuracy (Lecun et al., 2015) and thus have been successfully applied in various contexts such as speech recognition (Dahl et al., 2012), image classification (Chaudhuri & Bose, 2020; Hu et al., 2018) as well as demand (W. Huang et al., 2014), fraud (Bian et al., 2018), sales (Loureiro et al., 2018) and performance forecasting (Lee et al., 2017). Despite these benefits, studies that involve artificial neural networks within an entrepreneurial context are incipient (Chuang & Chao, 2013; Kim et al., 2020).

To foster our knowledge in the field of entrepreneurship, this study combines recent insights of early stage venture outcome factors and applies a diverse set of machine learning methods and specifically, artificial neural networks to forecast the likelihood of firm birth and firm abandonment in the gestation process. Henceforth, this work intends to contribute to the entrepreneurial literature by evaluating the efficiency of machine learning techniques and artificial neural networks to predict the likelihood of firm birth and firm abandonment on the

venture process. Moreover, considering simultaneously aspects of different natures, namely: start-up activities and their evolvement over time, firm specific attributes such as industry types, technological level etc. and founder attributes such as motivation and human capital.

5.4 Data and methodology

This section outlines the dataset as well as the methodology and variables used to predict the venture outcome. Furthermore, this section provides a brief description of the applied models.

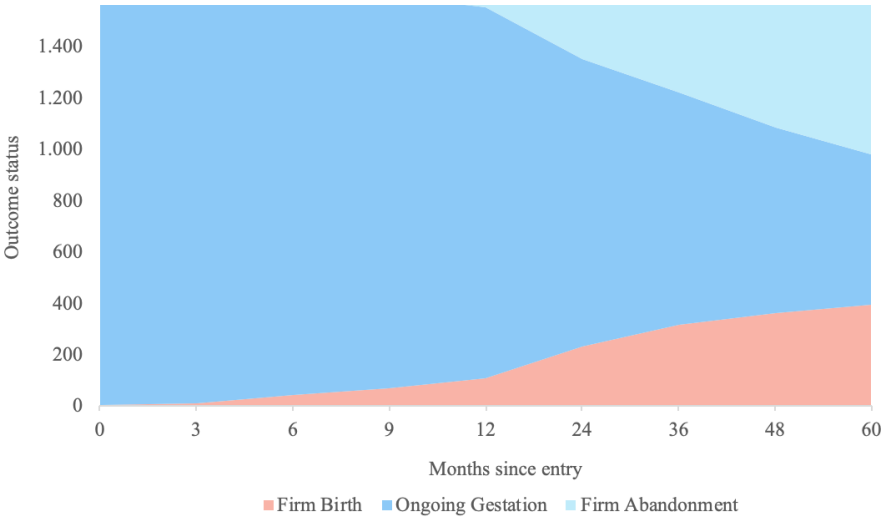
5.4.1 Dataset and variables

This study draws on a dataset of five longitudinal studies of four different countries: the US Panel Study of Entrepreneurial Dynamics (US PSED I with data from 1998-2004, N= 779; US PSED II with data from 2005-2008, N = 1071), the Swedish Panel Study of Entrepreneurial Dynamics (1998–2000, N = 577), the Comprehensive Australian Study of Entrepreneurial Emergence (CAUSEE study conducted in 2007–2013, N = 605) and the Chinese Panel Study of Entrepreneurial Dynamics (N = 505). These datasets have been harmonized into one dataset (Reynolds et al., 2016b), consisting of 3,537 active nascent entrepreneurs (for more details see Reynolds et al., 2016a). After data cleansing, the final dataset consisted of n= 1,452 cases that have been tracked over a period of 72 months. Characteristics of the venture, the nascent entrepreneurs and the conducted start-up activities were tracked. This paper looks at a period of 60 months for two reasons. First, this period appears to be especially crucial since half of start-ups discontinue their operation in their first five years of existence (SBA, 2012). Second, in the 72-month period a lot of missing cases exist, thus, using 60 months allows for the inclusion of more cases. The definition of the venture outcome (firm birth, firm abandonment and ongoing gestation) builds upon the PSED. Firm birth is defined as the presence of monthly profits that cover expenses and owner salaries, firm abandonment is defined based on having quit or stopped working on the start-up venture, and ongoing gestation relates to the entrepreneurs who are still in the process and have neither become a profitable new firm nor quit the venture (Reynolds et al., 2016a).

The definition of being considered as a nascent entrepreneur, thus being defined as entry point, is based on a variety of aspects, such as the amount of conducted start-up activities within a 12-month period (for more information see Reynolds et al., 2016a). Based on this entry point date, the nascent entrepreneurs provided a variety of information such as the outcome of the venture (i.e., firm birth, firm abandonment and ongoing gestation) in a three-month interval,

after different time periods. A summary of the development of the venture outcomes is graphically represented in figure 8.

Figure 8. Outcome status by month



Besides the two dependent variables firm birth and firm abandonment, a variety of other variables were measured which are summarized in Table 10. The dataset consists of numerical (i.e., team size, rate, concentration, timing and effort of the conducted start-up activities) as well as categorical (e.g., education, motivation etc.) variables. The measured start-up activities (e.g., writing a business plan) are coded binary (0=not conducted; 1=conducted). Moreover, this research looks also at firm abandonment and given that the research period of a part of the data covers a pre- as well as a post-dotcom and financial crises sequence (i.e., PSED I from 1998-2004; CAUSEE started from 2007-2013), a binary variable for such crises years is included. Pre-crises covers a period of three years prior to the outbreak date (this date and threshold is defined by various events such as the Lehman Brothers collapse, bailout programs etc.) and a two year crises period after the outbreak date (Giotopoulos et al., 2017). For example, for the financial crises this non-crisis period included the years 2005–2008, while the crisis period covers the years 2009–2011. For the dotcom crises the years 1998-1999 are considered non-crises years, while 2000-2002 are considered as crises years.

Table 10. List of predictor variables

Variables	Type of variable	Type of variable
Gender	PC	Categorical
Age	PC	Categorical
Education	PC	Categorical
Born in the country	PC	Categorical
Start-up experience	PC	Categorical
Industry experience	PC	Categorical
Management experience	PC	Categorical
Motivation to start a business	ME	Categorical
Growth preference	ME	Categorical
Team size	VC	Numerical
Ownership (minority, equal, majority)	VC	Categorical
Industry type (manufacturing, professional services, information technology, construction, transportation and agriculture)	VC	Categorical
Hi-tech venture	VC	Categorical
Availability of technology 5 years ago	VC	Categorical
R&D focus of the venture	VC	Categorical
A phone listing for the business acquired	SA	Categorical
Outside funding from institutions or external people received	SA	Categorical
Financial institutions or other people asked for funding	SA	Categorical
Credit from suppliers established	SA	Categorical
Own money has been invested into the business	SA	Categorical
Patent, trademark or copyright applied	SA	Categorical
Employees or managers hired	SA	Categorical
Major items, equipment, facilities or property purchased, leased or rented	SA	Categorical
Material, supplies, inventory purchased	SA	Categorical

New firm registered	SA	Categorical
Devoted full time to business	SA	Categorical
Marketing or promotion activities started	SA	Categorical
Effort made to define the market opportunity	SA	Categorical
Startup team organized	SA	Categorical
Financial projections prepared	SA	Categorical
A business plan prepared	SA	Categorical
Worked on a model or prototype for delivery of the product or service	SA	Categorical
Reporting first sales	SA	Categorical
Sum of conducted activities for each period (t_1-t_4)	SA	Numerical
Rate	CM	Numerical
Concentration	CM	Numerical
Timing	CM	Numerical
Effort	CM	Numerical
Crises	MA	Categorical

Notes. PC = personal characteristics, ME = motivational elements, VC = venture related characteristics, SA = start-up activities, CM = complexity measure, MA = market related measure

Finally, to complement the short summary of each variable presented in Table 10, the following four variables are further explained: rate, concentration, timing, and effort. Subsequent elaboration of the rate, concentration and the timing follows previous research (Hopp & Sonderegger, 2015; Lichtenstein et al., 2007), while the calculation of the effort is new and builds upon the hypothesized relationship outlined in Chapter 5.3.1.

Rate is defined as "the total number of start-up activities undertaken by the nascent entrepreneur divided by the duration of the gestation process of the new business" (Hopp & Sonderegger, 2015, p. 1084). An example would be, if the nascent entrepreneur has conducted in the following months (1, 3, 6, 6, 12) organizing activities. This sequence would equal five activities conducted within a time span of 12 months, resulting in a rate of 0.42 and reflecting the average pace of organizing activities across the gestation process.

Concentration reflects "how closely entrepreneurial activities are undertaken in relation to other start-up activities" (Hopp & Sonderegger, 2015, p. 1085). It is operationalized in terms of the variance of monthly activity times. High values reflect a high dispersion of organized activities whereas low values indicate that more of the start-up activities are bunched together (e.g., variance of zero if all activities are conducted in one month). For example, cases with the event histories of {1, 1, 1, 1} and {1, 3, 6, 6} have a concentration of 0 and 6 respectively.

Timing indicates "whether the bulk of start-up activities is accomplished earlier or later during the start-up process" (Hopp & Sonderegger, 2015, p. 1085). It is measured by taking the average event time divided by the duration of the gestation process. For example, the average event time related to the start-up activities {1, 3, 3, 6, 12, 12} with a duration of 12 months is 6.12. This figure is divided by the duration of 12 months, resulting in a timing of 0.51. Values close to one indicate that start-up activities occurred rather at the end of the gestation process whereas values close to 0 mean that organizing activities rather occurred in the first month of the gestation process.

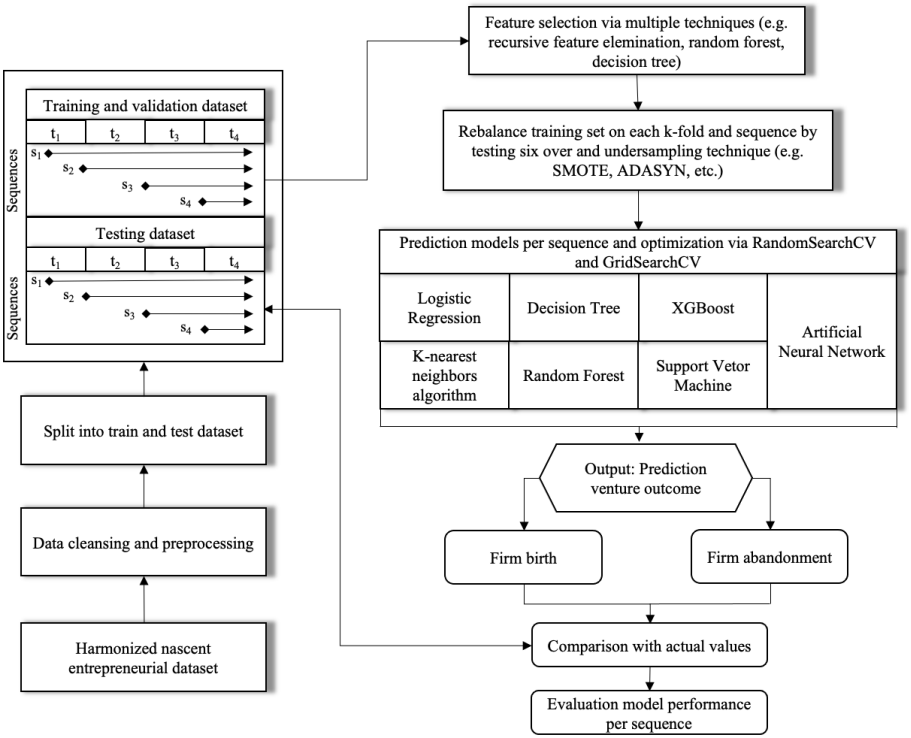
Effort describes the development of the start-up activities over the period. It is calculated by first looking at the difference between the total amount of conducted activities of two different periods. For example, if the entrepreneur conducted 2 activities at t_1, which is the minimum number of activities to be considered nascent, and 5 activities twelve months later at t_2, the effort invested over this period is calculated by subtracting (t_2 -t_1) the amount of activities divided by the duration, that is 12 months. The higher the value, the higher the effort in the time period.

5.4.2 General analytical approach

The goal of this paper is to predict, based on independent variables after 12, 24, 36 and 48 months (t_1-t_4), the likelihood of firm birth and firm abandonment after 60 months, while aiming to use different data mining techniques and explore the potential of artificial neural networks in the context of a smaller dataset. Building upon previous research (Delmar & Shane, 2004; Gartner & Liao, 2012) the dependent variable was coded for the analysis on firm birth as 1=firm birth and 0=otherwise, while for firm abandonment the coding was 1=firm abandonment and 0=otherwise. Data of different levels (see Table 12) of the gestation process are employed as independent variables to predict the respective venture outcome. The independent variables are measured in a yearly period. For example, in order to predict the outcome after 60 months, the recorded independent variables after 12, 24, 36 and 48 months are tested.

Overall, the methodological process includes three steps: first data preprocessing, second an optimization step, third an evaluation step. The optimization phase included selecting the optimal set of independent variables as well as conducting hyperparameter tuning for each applied technique using either a grid search (GridSearchCV) or a random search (RandomSearchCV) approach with k-fold cross-validation (k=10). Both approaches are common techniques to optimize the models hyperparameters and to derive at an optimized model (Bergstra & Bengio, 2012; Ching et al., 2018; Vo et al., 2019). The choice of the approach related to the needed computational power (i.e. random search for k-nearest neighbors, decision tree, random forest, XGBoost, SVM, ANN and grid search for the logistic regression). The final and third step included the testing process with the model comparison. The methodology followed in this study is outlined in Figure 9 and is presented in the next paragraphs.

Figure 9. Schematic outline of the analytical approach



After the data preprocessing, the initial step was to split the data into a training, validation, and testing datasets. Training and validation included 75% of the observation (n=1,089) while 25% of the dataset are used as a test and hold-out set (n=363) to validate the performance of the models. Given the imbalanced dataset (see figure 8), the data split included stratification based on the outcome variable.

To further optimize the models, different variable combinations are tested which lead to better performing models. This feature selection procedure was done by applying eight different techniques, including pearson correlation coefficient, chi-2, random forest, decision tree, XGBoost and recursive feature elimination (RFE) using three different classifiers (logistic regression, k-nearest neighbors, support vector machine). RFE can shortly be described as

follows: it starts with the full training dataset and selects features by recursively considering smaller and smaller sets of features. Based on these results, the feature identified the most within each approach is chosen for training and validating the model.

A feature was considered important if it was selected in seven out of the nine applied techniques. As one of the goals in this study is to analyze the differences in feature importance between the single sequences (t_1-t_4), this procedure was repeated four times for each set of independent variables.

After this feature selection, the train and validation procedure builds on a k-fold cross-validation step (k=10) to calculate the highest F1-score for each model and sequence. This cross-validation procedure ensures that any sampling biases can be eliminated from the training process (Topuz et al., 2018). To provide reassurance that the selected features improve the model performance, a comparison of the mean values from each k-fold of the full as well as the feature selected model was conducted.

Moreover, as already outlined, the dataset consists of fewer firm birth and firm abandonment cases. From a decision maker perspective, the aim is to identify entrepreneurs who are likely to either achieve firm birth or abandon the venture. Hence, the minority classes of firm birth and firm abandonment need to be accurate. Thus, during the training process of the models, various class balancing techniques were applied in order to put more weight on the firm birth and abandonment predictions. Different over- and under- sampling techniques exist and their effect on the model performance varies (Burnaev et al., 2015). In the literature, next to classical random over-sampling and random under-sampling, synthetic minority oversampling technique (SMOTE) or its extension, the adaptive synthetic sampling (ADASYN), are popular oversampling methods (Blagus & Lusa, 2013) that performed well in various other areas (Han et al., 2005). Next to these techniques, this study further tested other, though less applied, sampling approaches such as the Neighborhood Cleaning Rule (Laurikkala, 2002) and the Edited Nearest Neighbour (Raniszewski, 2010). Depending on the model, the technique that increased the F1-score and provided reasonable metrics for the area under the receiver operating curve (ROC AUC) and accuracy, while keeping the recall or precision metrics on a relative acceptable level, was applied. Recall refers to the percentage of total relevant results correctly classified by the applied model, while precision provides information about the fraction of correctly classified cases out of all the positive classified cases. The F1-score is then defined as the harmonic mean of precision and recall. The ROC AUC shows the

performance of the model and can take a maximum value of one. The closer the value to one, the better the model.

The over- and under-sampling procedure was performed inside the cross-validation loop for each k-fold. This procedure was applied after the feature selection (Blagus & Lusa, 2013). In the last step and as outlined above, a k-fold cross-validation (k=10) with either random search (using thirty iterations) or grid search was applied for tuning the hyperparameters.

In order to evaluate the models, three widely used evaluation categories were applied (Topuz et al., 2018; Veganzones & Séverin, 2018): (i) accuracy for the probability of correct classifications, (ii) precision as well as recall while using a confusion matrix for recognizing the nascent entrepreneur of a certain group (i.e., firm birth or firm abandonment), and (iii) ROC AUC and F1-score for evaluating the overall performance. These measures are calculated for each k-fold (k=10), averaged and evaluated for each of the data mining techniques and sequences to obtain an overall estimate of the performance measures for the learning algorithms. Finally, all the trained models were evaluated on the hold-out, test dataset.

5.5 Classification methods

This paper tackles a typical classification problem using an imbalanced dataset. A variety of classification methods exist and accordingly, besides classical statistical methods such as logistic regression, six data mining techniques (k-nearest neighbors, decision trees, random forest, XGBoost, support vector machine, artificial neural network) are applied to predict the likelihood of firm birth and firm abandonment. In the following, these methods are shortly described.

5.5.1 K-nearest neighbors algorithm

Popular in pattern recognition and to solve classification problems, the KNN algorithm is considered as an efficient and a relative simple low-computation complexity method (Wang et al., 2013). The algorithm is based on the concept that data points of the same class should be closer in the feature space. The distance can be defined via the number of samples closest in distance to the new, prediction point (k-nearest neighbor learning). Due to its simplicity and effectiveness, this algorithm has been applied in variety of settings and provided good results (J. Li & Wang, 2015).

Using RandomSearchCV, configuration included the number of neighbors ranging from 2-70 in steps of two, four different algorithm types (auto, ball tree, k-d tree and brute force),

leaf size in steps of 5 ranging from 10-40, as well as weights (uniform or distance) and a power parameter (1 or 2).

5.5.2 Decision tree

Decision trees are a powerful and efficient non-parametric method for classifications. Decision trees consist of nodes where attributes are tested and where greedy search methods, which involve growing and pruning decision tree structures, are typically employed to explore the exponential space of possible models (Mehmed, 2011). Even though the performance of decision tree models appears to be lower compared to other methods, decision trees are, among other reasons, popular because of their ease of use and interpretability. This study applies the CART (classification and regression trees) algorithm in Python which is similar to the C4.5 algorithm that has been applied in other studies as well (L. Li et al., 2014).

Nodes, split leaf, depth and features of the tree were tuned. For the depth, values ranging from 2 to 30 are tuned, while the values for the split and leaf ranged from 1-20 in steps of 1.

5.5.3 Random forest

This technique consists of an ensemble of decision trees and allows for the ability to overcome certain weaknesses of simple decision trees, such as high sensitivity to small variations in data (Loureiro et al., 2018). Each tree is grown by using a random subset of the input variables and at each split a random sample of predictors is examined. The tree is then allowed to grow fully. Thus, no pruning techniques are required. Random forest models have proven to provide good performance (Veganzones & Séverin, 2018).

The number of trees to grow are tested for the values 10, 50, 100, 150, 200, 250, 500, 1000, 1500, 2000. Values of depth ranged from 2-50 in steps of 2, while the splits ranged from 10-200 in steps of 25.

5.5.4 XGBoost

Standing for extreme gradient boosting framework, this technique builds on previous ideas in gradient boosting (Friedman, 2001). Simplified, while boosting refers to modifying weak learners (e.g., decision trees) to strong learners, the objective of gradient boosting is to minimize weaknesses based on a gradient decent approach. XGBoost is an improved and more efficient model that introduces a regularized model formalization, thus reducing overfitting and

increasing the predictive performance. Among others, XGBoost has for example been applied to anticipate failure or success of banks (Climent et al., 2019).

A variety of parameters are tuned. A "gbtree" booster is applied and the parameter tuned to include the number of trees (500, 1000), depth (3,5,7,9), the learning rate (0.01,0.1,0.2,0.3), gamma (0-0.4) and subsampling with values ranging from 0.5 to 0.9 with steps of 0.1.

5.5.5 Support vector machine

Support Vector Machine (SVM) is a traditional machine learning technique (Kraus & Feuerriegel, 2017) that has been widely applied in different research fields - including the entrepreneurship literature (Blanco-Oliver et al., 2014; Tu et al., 2019; Zekić-Sušac et al., 2015) - to predict a variety of outcomes such as credit rating (Z. Huang et al., 2004), time series (Lu et al., 2009), fraud detection (Ravisankar et al., 2011) as well as financial distress (Blanco-Oliver et al., 2014). Support vector machines can also be used for smaller sample sizes, provided good predictive results (Tu et al., 2019) and appear to outperform other classical statistical models (Chaudhuri & Bose, 2020), thus building an essential model for comparison reason. For more details towards the SVM algorithm see Mehmed (2011).

Using RandomSearchCV, the influence of two different kernel types (linear, rbf) as well as the cost and gamma parameters required for each kernel were studied. For the parameter cost, values ranging from 0.1-2 with a step of 0.1 were considered, for gamma next to auto and scale, the boundary test values ranged from 0.5-5 with steps of 0.5.

5.5.6 Artificial neural networks

Artificial Neural Networks (ANN) are an abstract computational model of the human brain. ANN consists of a massive parallel distributed processor, made up of simple processing units (neurons) that are connected with each other and where the ability to learn experiential knowledge is expressed through the interunit connection strength (Mehmed, 2011). In other words, the neuros in these networks are organized in layers where the neurons receive first an input and then provide as an output a nonlinear weighted sum to the next connected layer. This weighted connection can be described as the learning process of the model to understand the relationships between variables. A simple ANN consists of an input, a hidden as well as an output layer with a corresponding activation function. In this study, a multilayer perceptron (MLP) model with two hidden layers was applied using backpropagation learning methods. Backpropagation refers to the train and learning process of a neural network and is currently

one of the most widely used neural network algorithms (Z. Huang et al., 2004; LeCun et al., 2015; Lee et al., 2017). In general, ANN have been previously applied in similar contexts such as bankruptcy prediction models (Veganzones & Séverin, 2018) and offer several useful properties and capabilities such as nonlinearity, learning from examples, adaptivity and fault tolerance (Mehmed, 2011).

For the MLP model, the configuration of the parameters was based on a random search k-fold (k=10) approach. Configuration included, the number of hidden neurons ranging between 25 and 200 with a step of 25 for each hidden layer, different activation functions (relu, tanh, sigmoid, hard sigmoid, swish), different solvers (sgd, adam, nadam, adagrad, adadelta, rmsprop and lfbgs), iterations ranging from 25-300 with steps of 25, batch sizes ranging from 10-150 with steps of 10 and three different learning rates (constant, adaptive and in-scaling).

5.6 Results and discussion

Table 10 summarizes the results of each of the applied techniques for each sequence (t_1-t_4) on the testing dataset. The results show that most of the proposed models performed better than chance to predict firm birth as well as firm abandonment for each sequence under observation. Even though none of the considered techniques extremely outperformed another, certain distinctions can be made.

It appears that the XGBoost model performed the best in predicting firm birth in the sequences t_1-t_4. Taking for example the predictors after 36 months, the model achieved an accuracy of 78.79% while having an F1-score of 62.44%. The ANN achieved for the same sequence an accuracy of 72.45% with an F1-Score of 57.63%. While being around 6% worse, the neural network model performed better than the decision tree, random forest or k-nearest neighbors model.

For predicting the likelihood of firm abandonment, different models performed better for different sequences. Taking for example the sequence after 12 months, XGBoost achieved an accuracy of 52% with an F1-score of 55%. After 24 months, KNN provided, given the problem under research, reasonably acceptable results with an accuracy of 60.06% and a F1-score of 55.11%, while the support vector machine achieved an accuracy of 57.30% and a F1-Score of 57.06% for the predictors after 36 months. Finally, the logistic regression model achieved with the predictors of month 48 a comparably good predictability with an accuracy of 71% and an F1-score of 62%. The performance of the ANN model for predicting firm abandonment was in the mid-range compared to the other models over the different sequences.

Translating these numbers into a practical context, different aspects need to be considered. This study aimed for balanced precision and recall metrics, thus tuning the models in order to increase the overall F1-score. Thus, the evaluation of the models is based on a reasonable balance between F1-score, accuracy and ROC AUC, which is a measure that is completely independent from any misclassification costs (du Jardin, 2018). However, from a practical perspective, this tuning decision could vary in a sense that a higher precision score for firm birth and a higher recall score to identify firm abandonment may be preferred.

Table 11. Performance evaluation of classification models on test dataset

	Model Classifier	Logistic Regression (LogR)			
Firm birth	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	54,27%	66,12%	77,69%	72,73%
	ROC AUC	62,48%	73,60%	79,44%	82,71%
	Precision	28,57%	38,27%	51,67%	45,39%
	Recall	63,53%	72,94%	72,94%	81,18%
	F1 Score	39,42%	50,20%	60,49%	58,23%
	TP / TN / FP / FN**	54 / 143 / 135 / 31	62 / 178 / 100 / 23	62 / 220 / 58 / 23	69 / 195 / 83 / 16
Optimization	No. Features	27	25	17	14
	Sampling Technique*	RUS	ROS	EDN	EDN
	Hyperparameters	C=4, penalty='l2', solver='lbfgs'	C=0.2, penalty='l1', solver='saga'	C=0.1, penalty='l1', solver='saga'	C=0.1, penalty='l1', solver='liblinear'
Firm abandonment	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	49,59%	60,61%	61,16%	71,07%
	ROC AUC	62,33%	62,66%	64,13%	72,68%
	Precision	40,01%	46,01%	46,10%	58,39%
	Recall	82,31%	57,69%	50,00%	66,92%
	F1 Score	53,90%	51,19%	47,97%	62,37%
	TP / TN / FP / FN**	107 / 73 / 160 / 23	75 / 145 / 88 / 55	65 / 157 / 76 / 65	87 / 171 / 62 / 43
Optimization	No. Features	31	27	25	62
	Sampling Technique*	EDN	ADA	ADA	ADA
	Hyperparameters	C=2, penalty='l1', solver='liblinear'	C=0.1, penalty='l2', solver='saga'	C=0.2, penalty='l2', solver='saga'	C=0.3, penalty='l1', solver='liblinear'

*ROS = RandomOverSampler; RUS = RandomUnderSampler; EDN = EditedNearestNeighbours; NCR = NeighbourhoodCleaningRule; SMT = SMOTE; ADA = ADASYN

** TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

Table 11. (Continued)

	Model Classifier	K-nearest neighbors (KNN)			
Firm birth	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	55,65%	73,00%	73,55%	77,96%
	ROC AUC	69,19%	74,54%	79,28%	82,67%
	Precision	30,41%	44,25%	45,80%	52,10%
	Recall	69,41%	58,82%	70,59%	72,94%
	F1 Score	42,29%	50,51%	55,56%	60,78%
	TP / TN / FP / FN**	59 / 143 / 135 / 26	50 / 215 / 63 / 35	60 / 207 / 71 / 25	62 / 221 / 57 / 23
Optimization	No. Features	27	25	17	14
	Sampling Technique*	ADA	ROS	RUS	NCR
	Hyperparameters	n_neighbors=48, weights='uniform', algorithm='brute', leaf_size=15, p=2	n_neighbors=68, weights='uniform', algorithm='brute', leaf_size=35, p=1	n_neighbors=44, weights='uniform', algorithm='brute', leaf_size=25, p=1	n_neighbors=68, weights='uniform', algorithm='brute', leaf_size=25, p=1
Firm abandonment	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	45,45%	60,06%	58,13%	61,43%
	ROC AUC	57,92%	63,26%	63,25%	68,46%
	Precision	37,86%	46,11%	44,44%	47,47%
	Recall	81,54%	68,46%	67,69%	72,31%
	F1 Score	51,71%	55,11%	53,66%	57,32%
	TP / TN / FP / FN**	106 / 59 / 174 / 24	89 / 129 / 104 / 41	88 / 123 / 110 / 42	94 / 129 / 104 / 36
Optimization	No. Features	31	27	25	22
	Sampling Technique*	EDN	RUS	RUS	SMT
	Hyperparameters	n_neighbors=6, weights='distance', algorithm='kd_tree', leaf_size=25, p=2	n_neighbors=64, weights='distance', algorithm='brute', leaf_size=25, p=1	n_neighbors=68, weights='uniform', algorithm='brute', leaf_size=35, p=1	n_neighbors=68, weights='uniform', algorithm='brute', leaf_size=35, p=1

*ROS = RandomOverSampler; RUS = RandomUnderSampler; EDN = EditedNearestNeighbours; NCR = NeighbourhoodCleaningRule; SMT = SMOTE; ADA = ADASYN

** TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

Table 11. (Continued)

	Model Classifier	Decision Tree (DT)			
Firm birth	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	57,30%	74,10%	77,69%	72,73%
	ROC AUC	59,15%	72,72%	78,53%	78,49%
	Precision	29,17%	45,16%	52,00%	45,00%
	Recall	57,65%	49,41%	61,18%	72,12%
	F1 Score	38,74%	47,19%	56,22%	56,00%
	TP / TN / FP / FN**	49 / 159 / 119 / 36	42 / 227 / 51 / 43	52 / 230 / 48 / 33	63 / 201 / 77 / 22
Optimization	No. Features	27	25	17	14
	Sampling Technique*	ROS	EDN	EDN	ADA
	Hyperparameters	min_samples_split = 16, min_samples_leaf = 19, max_depth=2, max_features='auto'	min_samples_split = 20, min_samples_leaf= 8, max_depth=29, max_features='log2'	min_samples_split =4, min_samples_leaf = 16, max_depth=28, max_features='log2'	min_samples_split =8, min_samples_leaf =5, max_depth=3, max_features='log2'
Firm abandonment	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	42,42%	39,67%	47,66%	54,27%
	ROC AUC	56,74%	55,34%	57,12%	65,27%
	Precision	36,33%	35,69%	39,44%	42,86%
	Recall	80,77%	85,38%	86,15%	83,08%
	F1 Score	50,12%	50,34%	54,11%	56,54%
	TP / TN / FP / FN**	105 / 49 / 184 / 25	111 / 33 / 200 / 19	112 / 61 / 172 / 18	108 / 89 / 144 / 22
Optimization	No. Features	31	27	25	22
	Sampling Technique*	EDN	EDN	EDN	EDN
	Hyperparameters	min_samples_split = 13, min_samples_leaf = 25, max_depth=23, max_features='log2'	min_samples_split =8, min_samples_leaf = 5, max_depth=3, max_features='log2'	min_samples_split = 3, min_samples_leaf =25, max_depth=29, max_features='log2'	min_samples_split = 17, min_samples_leaf =18, max_depth=20, max_features='sqrt'

*ROS = RandomOverSampler; RUS = RandomUnderSampler; EDN = EditedNearestNeighbours; NCR = NeighbourhoodCleaningRule; SMT = SMOTE; ADA = ADASYN

** TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

Table 11. (Continued)

	Model Classifier	Random Forest (RF)			
Firm birth	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	56,20%	46,01%	59,78%	58,95%
	ROC AUC	59,40%	59,11%	62,84%	70,24%
	Precision	42,49%	39,29%	45,56%	45,85%
	Recall	63,08%	93,08%	63,08%	80,77%
	F1 Score	50,77%	55,25%	52,90%	58,50%
	TP / TN / FP / FN**	82 / 122 / 111 / 48	121 / 46 / 187 / 9	82 / 135 / 98 / 48	105 / 109 / 124 / 25
Optimization	No. Features	27	25	17	14
	Sampling Technique*	ADA	EDN	ROS	NCR
	Hyperparameters	n_estimators = 2000, max_depth=2, min_samples_split=50, max_features='sqrt'	n_estimators = 100, max_depth=18, min_samples_split=150, max_features='sqrt'	n_estimators = 200, max_depth=46, min_samples_split=150, max_features='auto'	n_estimators = 200, max_depth=30, min_samples_split=100, max_features='sqrt'
Firm abandonment	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	44,63%	61,71%	57,85%	61,71%
	ROC AUC	62,93%	63,05%	62,24%	66,34%
	Precision	38,66%	46,94%	43,58%	47,49%
	Recall	93,08%	53,08%	60,00%	65,38%
	F1 Score	54,63%	49,82%	50,49%	55,02%
	TP / TN / FP / FN**	121 / 41 / 192 / 9	69 / 155 / 78 / 61	78 / 132 / 101 / 52	85 / 139 / 94 / 45
Optimization	No. Features	31	27	25	22
	Sampling Technique*	EDN	ROS	ROS	RUS
	Hyperparameters	n_estimators = 150, max_depth=34, min_samples_split=10, max_features='auto'	n_estimators = 50, max_depth=36, min_samples_split=175, max_features='auto'	n_estimators = 50, max_depth=6, min_samples_split=175, max_features='auto'	n_estimators = 100, max_depth=18, min_samples_split=10, max_features='sqrt

*ROS = RandomOverSampler; RUS = RandomUnderSampler; EDN = EditedNearestNeighbours; NCR = NeighbourhoodCleaningRule; SMT = SMOTE; ADA = ADASYN

** TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

Table 11. (Continued)

	Model Classifier	eXtreme Gradient Boosting (XGBoost)			
Firm birth	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	72,18%	74,38%	78,79%	75,48%
	ROC AUC	76,71%	77,91%	83,39%	83,08%
	Precision	43,75%	46,61%	53,33%	48,48%
	Recall	65,88%	64,71%	75,29%	75,29%
	F1 Score	52,58%	54,19%	62,44%	58,99%
	TP / TN / FP / FN**	56 / 206 / 72 / 29	55 / 215 / 63 / 30	64 / 222 / 56 / 21	85 / 210 / 68 / 21
Optimization	No. Features	27	25	17	14
	Sampling Technique*	EDN	EDN	NCR	EDN
	Hyperparameters	learning_rate =0.2, n_estimators=400, max_depth=3, min_child_weight=2, gamma=0.5, subsample=0.7, colsample_bytree=0.2	learning_rate =0.3, n_estimators=100, max_depth=7, min_child_weight=0.5, gamma=0.9, subsample=0.8, colsample_bytree=0.1	learning_rate =0.01, n_estimators=200, max_depth=7, min_child_weight=1, gamma=0.6, subsample=0.5, colsample_bytree=0.8	learning_rate =0.01, n_estimators=400, max_depth=8, min_child_weight=3, gamma=0.8, subsample=0.9, colsample_bytree=0.2
Firm abandonment	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	51,79%	60,33%	52,89%	65,56%
	ROC AUC	61,41%	60,73%	62,98%	68,31%
	Precision	41,31%	44,93%	41,63%	51,82%
	Recall	82,31%	47,69%	78,46%	54,62%
	F1 Score	55,01%	46,27%	54,40%	53,18%
	TP / TN / FP / FN**	107 / 81 / 152 / 23	62 / 157 / 76 / 68	102 / 90 / 143 / 28	71 / 167 / 66 / 59
Optimization	No. Features	31	27	25	22
	Sampling Technique*	EDN	ROS	EDN	SMT
	Hyperparameters	learning_rate=0.2, n_estimators=600, max_depth=8, min_child_weight=0.5, gamma=0.7, subsample=0.5, colsample_bytree=0.8	learning_rate =0.1, n_estimators=200, max_depth=2, min_child_weight=0.5, gamma=0.7, subsample=0.6, colsample_bytree=0.9	learning_rate =0.1, n_estimators=200, max_depth=2, min_child_weight=4, gamma=0.7, subsample=0.5, colsample_bytree=0.4	learning_rate =0.1, n_estimators=200, max_depth=2, min_child_weight=0.5, gamma=0.7, subsample=0.6, colsample_bytree=0.9

*ROS = RandomOverSampler; RUS = RandomUnderSampler; EDN = EditedNearestNeighbours; NCR = NeighbourhoodCleaningRule; SMT = SMOTE; ADA = ADASYN

** TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

Table 11. (Continued)

	Model Classifier	Support Vector Machine (SVM)			
Firm birth	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	61,43%	74,66%	77,69%	74,38%
	ROC AUC	63,20%	73,97%	80,18%	82,63%
	Precision	30,77%	46,79%	51,64%	47,18%
	Recall	51,76%	60,00%	74,12%	78,82%
	F1 Score	38,60%	52,58%	60,87%	59,03%
	TP / TN / FP / FN**	44 / 179 / 99 / 41	51 / 220 / 58 / 34	63 / 219 / 59 / 22	67 / 203 / 75 / 18
Optimization	No. Features	27	25	17	14
	Sampling Technique*	ROS	EDN	NCR	EDN
	Hyperparameters	kernel = 'linear', shrinking=True, gamma='auto', probability=True, C=0.5, decision_function_shape='ovo'	kernel = 'linear', shrinking=False, gamma=2, probability=True, C=0.2, decision_function_shape='ovr'	kernel = 'linear', shrinking=False, gamma=3, probability=True, C=1.6, decision_function_shape='ovo'	kernel = 'linear', shrinking=True, gamma=3, probability=True, C=1.4, decision_function_shape='ovo'
Firm abandonment	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	47,66%	52,62%	57,30%	63,91%
	ROC AUC	61,40%	62,97%	67,41%	69,14%
	Precision	39,05%	41,18%	45,58%	49,66%
	Recall	82,31%	75,38%	76,30%	55,38%
	F1 Score	52,97%	53,26%	57,06%	52,36%
	TP / TN / FP / FN**	107 / 66 / 167 / 23	98 / 93 / 140 / 32	103 / 105 / 123 / 23	72 / 160 / 73 / 58
Optimization	No. Features	31	27	25	22
	Sampling Technique*	EDN	EDN	RUS	SMT
	Hyperparameters	kernel = 'linear', shrinking=False, gamma=3, probability=True, C=1.1, decision_function_shape='ovr'	kernel = 'linear', shrinking=False, gamma=3, probability=True, C=1.6, decision_function_shape='ovo'	kernel = 'linear', shrinking=False, gamma=2, probability=True, C=1.7, decision_function_shape='ovo'	kernel = 'linear', shrinking=True, gamma=3, probability=True, C=1.4, decision_function_shape='ovo'

*ROS = RandomOverSampler; RUS = RandomUnderSampler; EDN = EditedNearestNeighbours; NCR = NeighbourhoodCleaningRule; SMT = SMOTE; ADA = ADASYN

** TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

Table 11. (Continued)

	Model Classifier	Artificial Neural Networks (ANN)			
Firm birth	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	70,25%	65,29%	72,45%	71,63%
	ROC AUC	69,59%	76,09%	80,43%	82,23%
	Precision	39,45%	37,87%	45,03%	44,30%
	Recall	50,59%	75,29%	80,00%	82,35%
	F1 Score	44,33%	50,39%	57,63%	57,61%
	TP / TN / FP / FN**	43 / 212 / 66 / 42	64 / 173 / 105 / 21	68 / 195 / 83 / 17	70 / 190 / 88 / 15
Optimization	No. Features	27	25	17	14
	Sampling Technique*	ADA	ROS	NCR	EDN
	Hyperparameters	hidden_layer_sizes=(100,75), max_iter=150, batch_size=70, activation='relu', solver='sgd', alpha= 0.0001, learning_rate='constant'	hidden_layer_sizes=(100,75), max_iter=150, batch_size=70, activation='relu', solver='sgd', alpha= 0.0001, learning_rate='constant'	hidden_layer_sizes=(125,100), max_iter=150, batch_size=140, activation='relu', solver='sgd', alpha= 0.4, learning_rate='adaptive'	hidden_layer_sizes=(100,75), max_iter=150, batch_size=70, activation='relu', solver='sgd', alpha= 0.0001, learning_rate='constant'
Firm abandonment	Time	t_{12} - t_{60}	t_{24} - t_{60}	t_{36} - t_{60}	t_{48} - t_{60}
Evaluation metric	Accuracy	47,11%	63,64%	57,02%	61,16%
	ROC AUC	56,96%	63,86%	64,40%	69,78%
	Precision	38,69%	49,31%	43,93%	47,51%
	Recall	81,54%	54,62%	72,31%	80,77%
	F1 Score	52,48%	51,82%	54,65%	59,83%
	TP / TN / FP / FN**	106 / 65 / 168 / 24	115 / 68 / 165 / 15	94 / 113 / 120 / 36	105 / 117 / 116 / 25
Optimization	No. Features	31	27	25	22
	Sampling Technique*	EDN	ROS	NCR	NCR
	Hyperparameters	hidden_layer_sizes=(75,75), max_iter=250, activation='relu', batch_size=150, solver='adam', alpha= 0.2, learning_rate='invscaling'	hidden_layer_sizes=(125,100),max_iter=150,batch_size=140, activation='relu', solver='sgd', alpha= 0.4, learning_rate='adaptive'	hidden_layer_sizes=(100, 75), max_iter=150, batch_size=70, activation='relu', solver='sgd', alpha= 0.0001, learning_rate='constant'	hidden_layer_sizes=(100,75), max_iter=150, batch_size=70 activation='relu', solver='sgd', alpha= 0.0001, learning_rate='constant'

*ROS = RandomOverSampler; RUS = RandomUnderSampler; EDN = EditedNearestNeighbours; NCR = NeighbourhoodCleaningRule; SMT = SMOTE; ADA = ADASYN

** TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative

Taking for example firm birth: on the one hand, if the model accidentally predicts that an investment into a profitable venture is bad (false negative), we miss the chance to invest. This would involve opportunity costs but otherwise no effective monetary loss. If the model predicts that the entrepreneur achieves firm birth, but actually he does not (false positive) and abandons the venture, the costs will directly relate to the monetary and non-monetary investment size. Thus, in the case of identifying a successful nascent entrepreneur, the model with the higher precision should be preferred.

Looking for example at the XGBoost model with a precision of 53,33% based on the predictors of month 36, the model identified 120 entrepreneurs as profitable. Out of these 120 profitable entrepreneurs, 64 entrepreneurs are correctly identified as profitable while 56 entrepreneurs are wrongly classified as profitable, thus they have either abandoned or are still in the gestation process after 60 months. At the same time, the neural network achieved a precision of 45.03% and identified 151 entrepreneurs as profitable, out of which 68 entrepreneurs are identified correctly and 83 entrepreneurs are wrongly classified as profitable. While more profitable cases are identified in the neural network model, the amount of wrongly classified entrepreneurs is higher. Thus, the XGBoost was able to achieve a better precision score. It is important to note that the models can be optimized for precision, however one must be aware that this optimization will be at the expense of lower recall values. Nevertheless, the reported values can already be considered an improvement compared to randomly guessing, especially considering that two third of nascent entrepreneurs do not achieve firm birth after a period of 5 years (Reynolds & Curtin, 2009).

On the other hand, the model with the higher recall should be preferred if the goal is to identify entrepreneurs who do not abandon the venture. That means, it is acceptable to have more false positives (i.e., nascent entrepreneurs who are not abandoning are considered to have abandoned the venture) than false negatives (i.e., nascent entrepreneurs who abandon the venture are not identified as such). Taking again for example the results of the XGBoost with the predictors after 12 months: the model predicted that 104 entrepreneurs did not abandon the venture after 60 months. Out of this number, 81 entrepreneurs are correctly identified as such while 23 are wrongly classified as abandoning the venture. This equals a recall value of 82.31%. In other words, if an investor invests in these 104 cases, “only” 17.69% of the resources invested are allocate to entrepreneurs who abandon the venture after a period of five years, thus are definitely lost. The neural network model achieved a recall value of 81.54% and identified 89 entrepreneurs who did not abandon the venture. Out of these 89 cases, 65 are correctly classified

and 24 entrepreneurs are wrongly classified. Less cases are identified as not having abandoned the venture and thus, the XGBoost is able to achieve a better recall score.

Overall, comparing the XGBoost and the ANN, such performance difference may be based on various reasons. One relates to the dataset, which consists of a rather small and structured data set, where XGBoost often achieves state-of-the-art results and outperforms artificial neural networks (T. Chen & Guestrin, 2016; Climent et al., 2019; Joharestani et al., 2019). Second, neural networks include a complex set of hyperparameters that can be tuned and given the applied random search approach, additional optimization to achieve even better results cannot be ruled out. Third, neural networks often achieve better results if trained on larger data sets and further optimization could be done using deeper network structures (D'souza et al., 2020). Even though the neural network performed worse than the XGBoost model, the results are still promising given that at such an early stage of the start-up process (e.g., 12 month) most of the identified cases that do not abandon the venture are correctly classified as such.

In summary, the combination of all the proposed firm birth and firm abandonment prediction models could provide a valuable system to foster an optimal allocation of resources in successful entrepreneurs, while reducing respective resource misallocation in entrepreneurs who abandon the venture.

Furthermore, as outlined in Chapter 5.4.2, different feature selection methods have been applied to identify for each sequence the most important prediction factors. Figure 10 and 11 summarize the ten most important features for predicting firm birth and firm abandonment. Even though a variety of feature selection methods have been applied (see Chapter 5.4.2), the reflected feature importance in the figures is based on the results of the random forest model as the random forest (RF) method to select the key features in comparable problems has proven to be very effective (Bahl et al., 2019; Yue Wu et al., 2019).

For firm birth, the number of conducted start-up activities, the complexity dynamics, and a set of specific activities are crucial in predicting firm birth. Comparing the different sequences, it appears that for example financial projections, hiring employees or having formed a start-up team play at an early stage a bigger role in predicting firm birth, while financial aspects (i.e., asked for supplier credit) are at a later stage (i.e., 36 and 48 months) more important in order to predict firm birth. Moreover, among the complexity dynamic variables, the hypothesized and newly introduced effort variable appears to be an additional crucial factor in predicting firm birth. This result underpins, in an entrepreneurial context, the importance of the impact of

different level efforts on the successful firm creation process. Henceforth, scholars as well as practitioners are encouraged to include such dynamic variable in corresponding models.

For firm abandonment, activities related to the implementation of the venture (e.g., money invested, initiated phone lines, having achieved sales, revenues or income), the educational level, experience, industry type and the total number of conducted start-up activities are important indicators for predicting firm abandonment. Referring to the sequence, having achieved (or not) sales, revenues, incomes or having implemented a phone line are more important predictors during the later gestation process (i.e., 36 and 48 month). Otherwise, the distribution of the importance of the features to predict the likelihood of firm abandonment over the sequences appears overall quite stable.

Figure 10. Feature importance firm birth

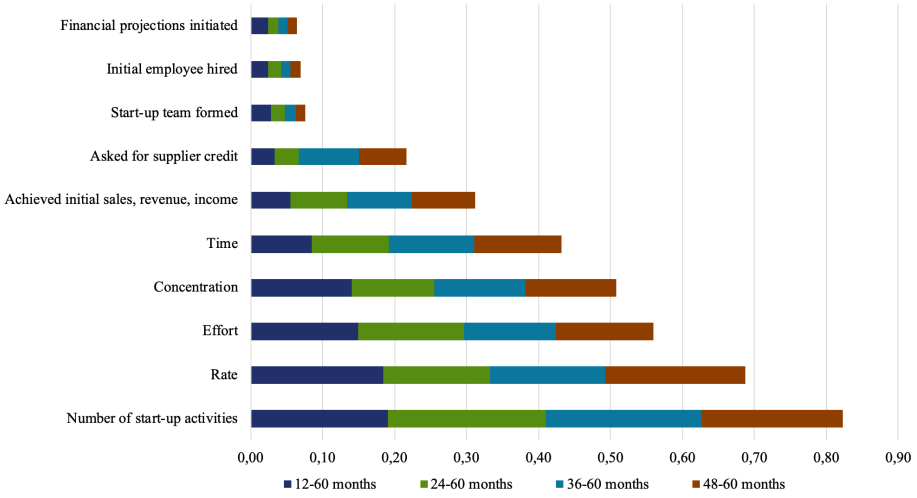
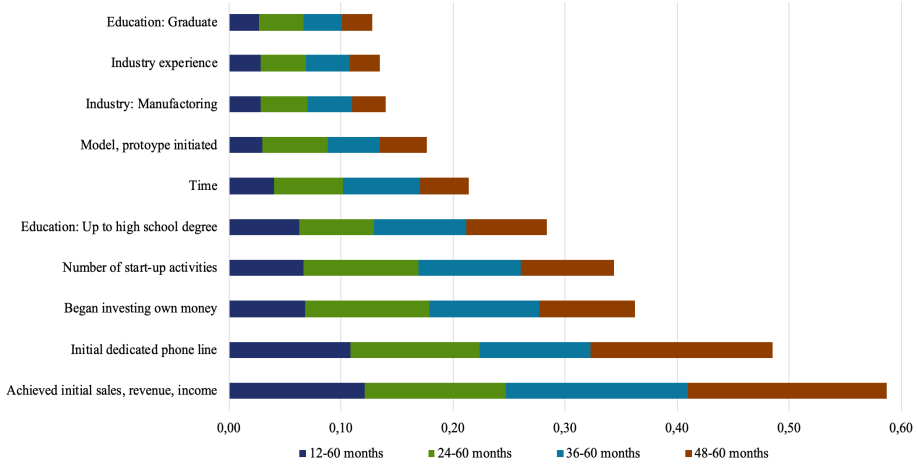


Figure 11. Feature importance firm abandonment



5.7 Practical implications and limitations

In light of the complexity of predicting the venture outcome, most of the proposed models achieve a better performance than simply guessing. Thus, they provide a valid possibility to predict the likelihood of firm birth as well as firm abandonment. The models can be used as an initial step to identify at an early stage, possible successful entrepreneurs as well as entrepreneurs who are likely to abandon the venture during the gestation process. In relation to firm birth, the involved parties such as investors, advisors and support programs can profit from such models by optimally allocating their monetary and non-monetary resources towards promising individuals and ventures, thus supporting them in their investment decision process.

On the other hand, the identification of entrepreneurs who possibly abandon the venture could also help for example support programs to seek exchange with the entrepreneurs before respective disengagement decision arises. This would allow in an early stage to point at crucial factors, may that be based on motivational matters or other concerns. Moreover, a combination of both models would allow for the ability to narrow down entrepreneurs who are possibly successful while at the same time identify entrepreneurs, whom may possibly abandon the venture after a period of 5 years. Even though these implications are worthwhile, a few particularities associated with the use of these models exist.

First, the variables used are self-reported and thus should not be applied as a sole predictor for success or venture disengagement. The models provide a possibility to improve decision making but the role of the expert remains extremely important in order to provide among other aspects, well-founded advises. Taking for example investors, their knowledge and sensitivity of an entrepreneurial character as well as the customer and the market related changes, remains extremely valuable in providing guidance. Following this line of argumentation, the firm emergence process itself is extremely complex and thus further variables may play an essential role. In that context, given the focus on the entrepreneur in such an early stage, cognition appears to have a profound effect on the likelihood of succeeding (Higgins & Spiegel, 2004; Hmielseki & Baron, 2008). Therefore, future research is encouraged to include such variables into the proposed models. Moreover, as outlined within the sequential analysis, the involved parties using such models must be aware of the changes over time. Henceforth, the models need to be retrained at least on a yearly basis in order to provide the necessary information.

Finally, the involved parties should be aware of the complexities associated with the different models, also with regard to a possible implementation. As data-mining techniques

such as neural networks and support vector machines are often considered a black box for themselves (Cortez & Embrechts, 2013), models with better visualization possibilities such as decision trees and to a certain extent random forest models, may provide a more cost-effective alternative.

5.8 Conclusions and future work

Predicting the likelihood of firm birth and firm abandonment is a complex problem. In light that only one third of the entrepreneurs appear to successfully navigate to the earliest stages of the firm creation process, the time and monetary costs are high for all involved parties. More accurate forecasting tools can serve as a support and an early alert system that can guide the involved parties to identify promising and successful entrepreneurs, while at the same time provide them over the gestation process with information to address critical situations before their occurrence.

Besides providing a literature review on nascent entrepreneurial success and failure factors, the goal of this paper was to develop such systems and establish as well as evaluate forecasting models to estimate the likelihood of firm birth and firm abandonment after 5 years in a sequential manner. Using the publicly available harmonized PSED dataset, this paper further applied a neural network learning model to assess its possible applicability. The results showed that the application of a neural network as well as other data mining techniques for predicting the likelihood of firm birth and firm abandonment is very promising. In that sense, the proposed models provide the involved parties such as investors, advisors and support programs with a valuable decision support tool to optimally allocate investment resources into entrepreneurs and their ventures. While XGBoost is very promising, neural networks also provide reasonable performance metrics, thus undergird their applicability not only for big data but also for smaller datasets, as used within this study. Even though their applicability is more complex, compared to other models such as decision trees or random forest, they can provide additional performance validity.

Moreover, a variety of over- and under-sampling techniques were applied to derive at the optimal predictions. In doing so, the results underline previous insights (Burnaev et al., 2015) who showed that there is no method that would always outperform the others, also in relation to the different sequences.

Finally, by identifying key features to predict firm birth and firm abandonment, valuable insights into the relationship between different attributes within the sequence could be

generated. Especially the sales activity and complexity dynamics such as the rate, time, concentration and the effort are key elements in predicting the venture outcome. Moreover, looking at the whole process, a variety of parameters are more significant at an early stage (e.g., forming a start-up team), while others are more important at a later stage of the gestation process (e.g., ask for supplier credits). This dynamism underpins the complexity of the gestation process and thus further underlines the need of an in-depth knowledge of the external parties about the entrepreneur, their progress and the market environment.

To improve the predictability of the proposed models, future research would profit from including further predictive variables such as cognitive factors, competitor data, further start-up activities, information related to the business idea or perceived product market-fit. Other possibilities of model improvements may be achieved by including additional layers in the neural network model. Lastly, combining text mining models of business plans, who have already been applied in an initial coin offering (ICO) context (Bian et al., 2018), with the proposed models could possibly further increase the predictive power.

5.9 Appendix

Table 12. Previous empirical research on venture outcome

Article	Sample	Firm birth / Firm abandonment / Ongoing gestation measure	Analytical model	Identified success factor
Arenius et al. (2017)	1,673 nascent entrepreneurs in the Harmonized PSED (Panel Study of Entrepreneurial Dynamics)	"Operationalized the criteria for the emergence of a new firm as reaching initial profit" (p. 89).	Necessary Condition Analysis (NCA)	"None of the individual gestation activities is a necessary condition for having profit at 24 months. Some level of activity is required for firm emergence" (p. 89).
Brush et al. (2008)	646 nascent entrepreneurs (PSED I)	Dichotomized self-reported categorical variable (operating business, still an active start-up, an inactive start-up, or no longer being worked by anyone, or something else). The fourth category (no longer being worked on by anyone) was coded as nascent venture disbanding (or failure). The other four categories were treated as continuation of the organizing effort (or non-failure).	Event history approach, Cox proportional hazard model, Cox regression survival procedure with implementing the Breslow method	Intentionality, resources, boundary and exchange are "necessary for firm survival in the short-term and those firms that organize more slowly are more likely to continue to organize" (p. 547).
Chandler et al. (2005)	124 nascent entrepreneurs (Sweden)	Self-reported as having reached profitability at any time during the study (6, 12, 18, or 24-month intervals).	Binary logistic regression	"Larger initial team size provides an advantage for new organizations, and that the benefits of adding and dropping team members are contingent on the stage of development of the organization and the dynamism of the environment" (p. 705).
De Carolis et al. (2009)	269 data points of nascent entrepreneurs (alumni from a major U.S. Eastern university)	Venture progress measured via development of start-up activities. The higher the number over time, the farther along the venture has progressed.	Multiple hierarchical regression	"Social networks and relational capital enhance levels of illusion of control, which is directly related to the progress of new venture creation" (p. 527).
Delmar & Shane (2003)	223 nascent entrepreneurs (Sweden)	Self-reported measurements for venture abandonment, product development and venture organizing activities. Abandoned was coded as dichotomous variable (1=yes, 0=no).	Fixed effect regression model and piece-wise exponential event history models	"Business planning reduces the likelihood of venture disbanding and accelerates product development and venture organizing activity" (p. 1165).
Delmar & Shane (2004)	223 nascent entrepreneurs (Sweden)	Self-reported measure for venture disengagement (dichotomous with 1 for abandoned and 0 otherwise). "New ventures that were not terminated by the end of the 30-month observation period were treated as right censored" (p. 395).	Weibull hazard rate models with robust clustering	Compared to beginning marketing activity, venture age, firm founder experience, undertaking activities to generate legitimacy (complete business plan, establish legal entity), asking for funds and obtaining inputs reduces the hazard of venture disbanding. Completing a business plan and establishing a legal entity "increasing the likelihood that the venture will initiate marketing and promotion, obtain inputs, and talk to customers" (p. 405).

Table 12. (Continued)

Delmar & Shane (2006)	223 nascent entrepreneurs (Sweden)	Self-reported measure of venture failure and sales. Coded as 1 if all members of the founding team terminated, otherwise coded as "0". "Those ventures not terminated by the end of a 30-month observation period were treated as censored" (p. 229).	Piece-wise constant exponential model	"Founding team experience (start-up and industry experience) enhances both new venture survival and sales, but that the effects are non-linear, and vary with venture age" (p. 215).
Dimov (2010)	195 nascent entrepreneurs (PSED I)	Self-reported measure as "an operating business (coded as 4), still active (coded as 3), inactive (coded as 2), or no longer worked on (coded as 1)" (p. 1136).	Ordered logit mode	"Opportunity confidence positively affects venture emergence and that, through it, entrepreneurial experience and early planning have only indirect effects on venture emergence. Industry experience has a direct, positive effect on venture emergence" (p. 1123).
Diochon et al. (2005a)	132 Canadian nascent entrepreneurs	Self-reported as either operating, active, inactive, no longer worked on by anyone, or something else. The "variables were dichotomized to facilitate a comparison of engaged (sustained) versus disengaged (discontinued) start-up efforts" (p. 62).	Chi-square analysis	"No significant differences in personal background factors (socio-demographic, work, and career backgrounds) on firm emergence" (p. 53).
Diochon et al. (2005b)	151 nascent entrepreneurs (Canada)	Self-reported measure as dichotomous variable, categorized as either being considered operating or nonoperating.	Discriminant analysis and an independent sample t-test	"Engaging in activities that connect the nascent entrepreneur with others (i.e., asking for funding, organizing a start-up team) are the fulcrum for controlling the resources needed to establish a sustainable venture" (p. 423).
Diochon et al. (2008)	81 nascent entrepreneurs (Canada)	Self-reported measure as either "operating, active, inactive, no longer worked on by anyone, or something else. For this article, only ventures that were operating or abandoned at the time of the fourth-year follow-up (n = 81) were included" (p. 157).	Chi-square analysis	Financial management capability (how the money for the new business would be handled) was found to be associated with firm emergence.
Freeland & Keister (2016)	977 nascent entrepreneurs (PSED II)	Self-reported measure for persistence. Operationalization of persistence as binary variable "but disaggregate new firm creation and disengagement into separate outcome states" (p. 215).	Multinomial logistic regression models	"Blacks were more likely to persist and remain engaged in an immature venture if they did not achieve success after two years in operation, whereas Hispanics were more likely to disengage" (p. 210).
Frid et al. (2016)	1,025 nascent entrepreneurs (PSED II)	Self-reported measure "coded as "1" for nascent entrepreneurs who disengaged during startup gestation and "0" for those who remained in gestation and/or started a new firm" (p. 900). Furthermore, the measurement of revenue as continuous variable was included.	Cox survival regression, binary logistic regression and OLS regression (for revenue variable)	"Higher dropout rates among low-wealth and moderately wealthy nascent entrepreneurs. Wealth appears to have no effect on venture creation for those managing to remain in gestation. Entrepreneurial success is concentrated at the top of the wealth distribution, despite notable evidence of capability for those at the lower end of the wealth distribution" (p. 895).

Table 12. (Continued)

Frid et al. (2015)	1,214 nascent entrepreneurs in the United States (PSED II)	"Coded as "1" for a new firm; "2" for still trying; and "3" for disengagement" (p. 20).	Multinomial logistic regression	"Nascent entrepreneurs that invest larger amounts as a proportion of their net income are more likely to succeed and are less likely to disengage from the process. Personal funds invested as a proportion of net income may therefore be a better measure of future success than the precise amount" (p. 19).
Gartner & Liao (2012)	695 nascent entrepreneurs (PSED I)	Self-reported measure where "operating" was considered as a successful venture formation, which is coded as "1" and the rest were coded as category "0" for "unsuccessful venture formation" (p. 707).	Logistic regression models	Risk-taking propensity, perceptions of environmental uncertainty and venture growth aspirations do not have a significant effect on venture creation success.
Hechavarria et al. (2012)	830 nascent entrepreneurs (PSED I)	Self-reported measure as either new firm, start-up continues, quit initiative. Two dichotomous variables for "new firm" and "quit" with the reference category "start-up continues".	Multinomial logistic regression	Motivational antecedents (self-efficacy as degree of belief in skills and abilities for entrepreneurial tasks) among nascent entrepreneurs significantly influence the likelihood of quitting the start-up process versus continuing nascent entrepreneurial start-up efforts.
Hechavarría et al. (2016)	1,409 nascent entrepreneurs (PSED I and II)	Start-up outcome measured as either new firm outcome, quit outcome, still trying ("persistence").	Competing risk regression hazard model and Cox proportional hazard model	Being financed primarily with external equity does not accelerate time to quitting, it decreases the cumulative incidence of quitting over time. The appreciable impact of external equity decreases as the ownership shares of the founding team increase. Start-ups primarily financed with debt decrease the incidence of quitting over time.
Honig & Karlsson (2004)	396 nascent entrepreneurs (Sweden)	Self-reported measure of survival and profitability. Survival was coded as 1 if the project survived throughout the time horizon and zero if it was abandoned at any point during that time horizon. Profitability was coded as 1 if the project was profitable and zero otherwise.	Hierarchical logistic regression	Planning does not directly affect survival, but instead influences survival only indirectly through its (direct) symbolism, learning, and efficiency outcomes. Writing a business plan has no statistically significant effect on the profitability of the nascent organization.
Honig & Samuelsson (2012)	623 nascent entrepreneurs (Sweden)	"Used the natural log of venture-level performance, measured as self-reported profit or loss, in years two and six as dependent variables" (p. 376).	Ordinary least squares regression (OLS)	"Neither formal planning nor changes in the business plan increased venture-level performance over a six-year study period" (p. 365).
Hopp (2012)	674 nascent entrepreneurs (PSED II)	Self-reported measure. Dichotomous variable categorized as either being operating or nonoperating. Still trying cases were omitted leading to the outcome "perceived emergence" against "venture disbandment" as the dependent variable.	Logistic regression models	Previous nonprofit experience (measured binary as "whether the previous employer was a public sector or government organization, a not-for-profit organization, a stable private sector business, a growing private sector business, or a declining private sector business" (p. 1258)) significantly increases the chances to successfully found a new venture.

Table 12. (Continued)

Hopp & Sonderegger (2015)	874 nascent entrepreneurs (PSED II)	"Occurrence of the first positive cash flow combined with a self-reported measure of being operational (versus still-trying or dis- banding the venture)" (p. 1084).	Probit and instrumental variable regression	"Later temporal concentration of organizing activities enhances the probability of successfully founding" (p. 1076).
Khan et al. (2014)	943 nascent entrepreneurs (PSED II)	Self-reported measure. Disengagement "was recoded as "1", and those considered as new firm or active start-up were consolidated and recoded as "0" (p.45).	Logistic regression models	"Nascent entrepreneurs' self-efficacy beliefs, with respect to their abilities to start a new venture, also prevent them from disengaging from firm creation efforts. Goal commitment to the start-up goal is also a good predictor of disengagement, where high commitment resulted in reduced offs of disengagement" (p. 51).
Liao & Gartner (2006)	276 nascent entrepreneurs (PSED I)	Self-reported measure via four categories – "operating business, active startup, inactive startup and no longer worked on." The first two categories were coded as "continuing startups" (1), while the last two categories were ''discontinued startups'' (0).	Logistic regression models	"Nascent entrepreneurs who completed a business plan were 2.6 times more likely to persist in the process of business emergence than those who did not complete a plan" (p. 23).
Liao & Gartner (2007)	312 nascent entrepreneurs (PSED I)	Self-reported measure as either "in business", "still active" or "inactive/quit". Binary coding with "in business" and "still active" relative to reference category "inactive/quit".	Multinomial logistic regression	"Nascent entrepreneurs who completed a business plan were six times more likely to start a business than individuals who did not complete a business plan. In addition, nascent entrepreneurs who contacted and participated in government-sponsored entrepreneurship programs were five times more likely to start a business than entrepreneurs who did not seek government assistance" (p. 1).
Liao et al. (2009)	830 nascent entrepreneurs (PSED I)	Self-reported measure for discontinuance, coded 1 if discontinued; 0 if survived.	Logistic regression models	Comparison between non- and technology-based start-ups. "With the exception of education and managerial experience, human capital has limited influence on discontinuance. Financial capital significantly decreases the odds of discontinuance" (p.1). Younger and older nascent entrepreneurs may have higher probability of discontinuation.
Lichtenstein et al. (2007)	109 nascent entrepreneurs (PSED I)	Perception of start-up status: (1) An operating business; (2) Still in an active start-up phase (i.e., "still trying"); (3) discontinued. Used a binary outcome of emerged (N=59) and not emerged (N=50) (excluded still trying cases).	Logistic regression models	Organizing activities at a faster rate, lower concentration of activities and an average timing that was later in the process increases the likelihood of firm birth.
Manolova et al. (2012)	525 firm-year observations (Norway sample)	Binary coded self-reported categorical variable (1: new firm founding, 2: quitting, and 3: either new firm or quit event outcome).	Discrete-time survival analysis, using the XTLOGIT procedure in STATA	"Organizations which engage in a greater number of organizing activities are more likely to continue the organizing effort. Intentionality, boundary, and exchange are positively associated with organizational emergence, whereas resources are a necessary, but not sufficient, condition for organizations to continue organizing" (p. 763).

Table 12. (Continued)

Muñoz-Bullon et al. (2015)	287 nascent entrepreneurs (PSED II)	Self-reported measure of established business (1) and discontinued startup (0). Established businesses were measured via owner's salary and the question if the monthly cash flow covers all expenses as well as if the business is established.	Logistic regression models	Team resource heterogeneity, access to financial assistance (e.g., equity, loans or loan guarantees), physical resources (e.g., use of land, space, buildings or equipment) and business assistance (e.g., legal, accounting or clerical assistance) have a positive impact on firm emergence. The positive effect is greater as the team has more experience in the industry in which the new business will compete.
Newbert (2005)	243 nascent entrepreneurs (PSED I)	Self-reported measure of firm emergence. "Operationalized as a dummy variable, with a value of 1 if the nascent entrepreneur indicated having received money, income, or fees from the sale of goods and/or services and 0 otherwise" (p. 62).	Binary logistic regression	Developing models and/or prototypes, purchasing materials, buying and/or renting facilities and/or equipment, investing their own money in the venture, committing full time to the venture, hiring employees, and engaging in promotional efforts are gestation activities common to successful nascent entrepreneurs. Slight differences exist between low-, mid- and high-tech industries.
Newbert & Tornikoski (2012)	402 nascent entrepreneurs (PSED I)	Operationalization of firm emergence according to "the total number of emergence factors that have been achieved divided by the total number of emergence factors that the nascent entrepreneur indicated were relevant to the start-up" (p.150).	Weighted hierarchical ordinary least- squares (OLS) regression	The "governance and content of the supporter network, along with network growth, are important in predicting organizational emergence" (p. 141).
Newbert et al. (2013)	346 nascent entrepreneurs (PSED I)	Self-reported measure of firm emergence. Operationalization of emergence as a dummy variable, with cases assigned a value of one if they had become cash flow positive as of the final wave of data collection and zero otherwise.	Multilevel random coefficient growth curve models	Compared to homogenous network (ties) there appear to be a positive relationship between heterogeneous network (ties) and the likelihood to be more successful at creating new organizations.
Nuñez (2015a)	552 nascent entrepreneurs (PSED II)	"Continuous emergence score that registers the number of founding activities conducted (i.e., how far a firm has "emerged") at each measurement point over the study period" (p. 38).	Multi-level longitudinal model	"Household income does have a significant impact on average firm emergence. Household income is not a significant predictor of start-up activity completion for team" (p. 31).
Nuñez (2015b)	287 nascent entrepreneurs (PSED II)	Firm emergence is determined by a continuous emergence score variable and measured via classifying the startup activities associated with the venture creation process according to the Katz & Gartner model (1988).	Linear mixed-effects model procedure in SAS	"Teams exhibit higher than average emergence, and relatively stable emergence growth" (p. 65).

Table 12. (Continued)

Renko (2013)	193 nascent entrepreneurs (PSED II)	Emergence measured via key resource activities (hires employees, receives external financing, and achieves first sale). Firm emergence was operationalized into a continuous variable (ranging from zero to one) using the total number of milestones (0–3) "achieved divided by the total number of milestones (1–3) that relevant to the start-up (achieving first sale was assumed as constant and relevant to all start-ups)" (p. 1054).	Logistic regression models	"Prosocial motivation decreases the likelihood of firm emergence within a 4-year follow-up period, and even more so when the product offering of the emerging venture is new to the markets" (p. 1045).
Rotefoss & Kolvereid (2005)	197 nascent entrepreneurs (Norway)	Firm emergence measured via a dummy variable (1 = business founder, 0= non-business founder). Business founders (measured if the nascent entrepreneur "(1) invested their own money in the business, (2) received first payment, and (3) registered the business as a legal entity) are defined as being able to establish a firm within a period of 3 years after the first interview" (p. 118).	Logistic regression models	"Entrepreneurial experience is the single most important factor for predicting the outcome of the business start-up process. Even though environmental resources play a role, human resources are generally found to be better predictors of the outcome of the business start-up process" (p. 109).
Shim & Davidsson (2018)	1,673 nascent entrepreneurs in the harmonized PSED	Using variables indicating the "successfully emerged" or "abandoned" outcome. Emergence measured as initial presence of monthly profit, registration and revenue. Abandonment measured via confirmation that the respondent and all other team members were no longer working or having quit on the start-up.	Inverse Probability Weighting (IPW), Exponential (EXP), Weibull (WEI), Log-normal (LN), and Log-logistic (LL) models	Firm emergence chance is "maximized around three months after inception of the process, implying that venture creation in the typical case is faster and simpler than previously believed" (p. 10). A small number of gestation activities is often enough for reaching initial profits at 24 months.
Hopp & Stephan (2012)	590 nascent entrepreneurs (PSED II)	Firm emergence measured via two self-reported variables: "occurrence of the first positive cash flow combined with a self-reported measure of being operational (versus disbanding the venture)" (p. 926). Cases who are still trying were excluded.	Probit model alongside an instrumental variable probit model	Community-level cultural (measured via perceptions of cultural norms in the respondent's community) norms influence "venture emergence indirectly through their impact on nascent entrepreneurs' start-up motivation and entrepreneurial self-efficacy beliefs" (p. 935).
Steffens et al., (2012)	202 new venture teams (Sweden)	Two self-reported measures (operational and profitability). "Operational is a variable indicating whether the venture is an operating business at 6 and 75 months after the initial interview, respectively. Profitability indicates whether the venture has achieved profitability, i.e., revenue covering all costs for 3 months or more, at 6 and 75 months" (p. 733).	Logistic regression models	"Homogeneous teams are less likely to be higher performing in the long term" and "no relationship between homogenous teams and short-term performance" (p. 727). "Greater start-up experience homogeneity leads to poorer long-term performance" (p. 736). "Service industries are likely to outperform ventures started in other industries in terms of being operational and/or being profitable at 75 months" (p. 737).
Thiess et al. (2016)	519 nascent entrepreneurs (PSED II)	Performance was measured via self-reported value of the expected revenues. Operationalized via the logarithmic value of the indicated expected revenues: "what is the total revenue or income expected in the first twelve months of operation?" (p. 61) in wave A of PSED II.	Hierarchical regression models	"Balanced experience types (heterogeneities in management and start-up experience) among nascent venture team members led to increased early-stage performance" (p. 55).

Table 12. (Continued)

Tornikoski (2008)	65 nascent entrepreneurs (French sample)	Firm Emergence is measured of five items related to the three emergence (Katz & Gartner, 1988) properties (boundary, resource and exchange property).	Structural equation modeling / Partial Least Squares path modeling (PLS)	Collective ability of the team, human capital of nascent entrepreneur and technological sophistication seem not to have any relevance in explaining firm emergence. Market conditions seem to be more important determinant of firm emergence than human capital or the characteristics of the team.
Tornikoski (2009)	65 nascent entrepreneurs (French sample)	Two manifest variables for firm emergence: a) new firm markers: range from 1 to 4 (quit - in business) and b) perception of a nascent entrepreneur: range from 0 to 5 via three properties (boundary, resources, exchange) of the Katz & Gartner (1988) framework.	Structural equation modeling / Partial Least Squares path modeling (PLS)	Resource combination has a statistically significant positive, networking a statistically significant negative and improvising has no statistically significant correlation with firm emergence.
Tornikoski & Newbert (2007)	2,490 data points of nascent entrepreneurs (PSED I)	Object and subject measurement variable. The object based on three properties (boundary, resources, exchange) of the Katz & Gartner (1988) framework. The subjective measure was measured as a dichotomous variable (1= operating business, 0= not operational, inactive, no longer being worked on, or something else).	Generalized linear mixed models (GLIMMIX procedure)	Human capital, network behavior, characteristics of the nascent organization, market attractiveness are limited or not sufficient to explain firm emerge. Actions a nascent organization take (began marketing efforts, projected financial statements, opened a bank account, and listed the organization in the phone book and developed a model/prototype and purchased/rented facilities and equipment regarding the later) may be more important than its characteristics in explaining organizational emergence.
Tornikoski & Puhakka (2009)	65 nascent entrepreneurs (French sample)	New firm is seen to be born on five properties (two boundary, two resource and one exchange property). Each time a nascent firm had achieved one of the five properties, it was assigned the value of 1. As such, the dependent variable (firm emergence) varied from zero to five (0–5).	Structural equation modeling / Partial Least Squares path modeling (PLS)	Active behaviors make a statistically significant impact on firm emergence, while initial conditions do not. Firm emerges merely through the active behaviors of entrepreneurs rather than as an outcome of initial conditions.
van Gelderen et al. (2005)	517 nascent entrepreneurs (Dutch sample)	Self-reported measure of being operational (comparison between successful new firm births and abandoned, excluded "still trying" from the analysis).	Logistic regression models	There is a relationship between the perceived risk, start-up capital, engagement (part-time vs. full-time), industry type (i.c. manufacturing), ambition and activity type (e.g., writing a business plan), received support via guidance, previous start-up experience and probability of success in comparison with nascent entrepreneurs in other sectors.
Yusuf (2012)	339 nascent entrepreneurs (PSED I)	Self-reported measure of entrepreneurial disengagement. Operationalized as 0 = no, 1 = yes if the start-up is no longer being worked on by anyone.	Hierarchical cluster analysis using Ward's method	The "results suggest two different disengagement clusters—the intelligent and reactive exit clusters. Cases of intelligent exit are found to be more similar to cases of operating businesses than they are to cases of reactive exit" (p. 783).

6. Concluding chapter

6.1 Contributions

The following chapter summarizes the main contributions made by this cumulative dissertation from theoretical and practical perspectives. A more detailed discussion about the contribution can be found in the corresponding chapters of the papers.

The first academic paper contributes to two streams in the entrepreneurial literature. First, it contributes to the literature on gestation activities by demonstrating how entrepreneurs attempt to accomplish their tasks and objectives via self-regulation. A possible explanation based on proposed regulatory fit is presented for why a small number of activities may lead to the birth of a new firm (Arenius et al., 2017). By doing so, further insight is provided into the emergence of new organizations (Davidsson, 2016). Second, the paper contributes to the literature stream on regulatory focus theory and examines the role of regulatory focus in the initial entrepreneurial action phase (Johnson et al., 2015). The authors highlight that a promotion focus may not only influence a single task but potentially the entire exploitation phase of the gestation process.

The main benefit for practitioners builds upon the understanding of the effect of different regulatory foci, especially the promotion focus, on the venture outcome. Entrepreneurs with elevated levels of promotion orientation who already have conducted a high number of exploitation activities without having achieved firm birth may be prone toward an escalation of commitment trap. To avoid such a trap, various techniques are proposed. For example, entrepreneurs could adopt a stage gate approach with the option to pull out at predetermined different stages of the venture development, with an affordable loss. Revisiting and challenging the status quo and learning from past behavior would further allow for de-escalation. Raising such awareness could also be beneficial for incubators and support programs, as they can identify and better guide the entrepreneur in early feedback rounds. Moreover, knowing the regulatory focus of the entrepreneur can further guide an optimal framing of the feedback and foster the motivation to achieve a desired outcome (Jarzebowski & Palermo, 2012; Van-Dijk & Kluger, 2004).

Paper two contributes to the emerging strand of entrepreneurial innovation that suggests a context-dependent analysis of innovation (Boyer & Blazy, 2013; Buddelmeyer et al., 2009; Hyytinen et al., 2015; Reid & Smith, 2000). More pointedly, the authors highlight possible

benefits that imitative business ideas may bear during the early entrepreneurial process. In doing so, the authors not only nurture further research in the field of imitative business ideas but also contribute to this literature stream by demonstrating how ex-ante measured imitativeness influences the outcome of a venture over time, thus rendering possible survivorship bias very small. Furthermore, the paper extends academic knowledge on the impact of different founder identity types, especially the Missionary types, on the likelihood of pursuing an innovative business idea. The findings are of considerable importance to scholarly understanding and advance the knowledge of who the innovators in society might be (Lawrence et al., 2012).

From a practical perspective, the results suggest that entrepreneurs following an imitative strategy in an early stage may increase their likelihood of firm birth as they avoid possible liabilities of new firms (Stinchcombe, 1965). Moreover, the results provide private and public start-up support programs with an additional possibility to steer monetary and non-monetary investments toward innovative entrepreneurs by further including founder identity type scales, especially the Missionary identity type, as additional selection criteria. Allocation of resources to such entrepreneurs could further nurture the possibility of a desired societal change, a widespread policy goal.

The *third paper* of this thesis contributes to two distinct streams of literature. The first stream relates to the literature on success and failure factors within the gestation process by summarizing insights about the question of what factors impact the venture outcome (Brouder & Eriksson, 2013; Omri & Boujelbene, 2018). Second, the paper contributes to the methodological literature within the field of entrepreneurship. The paper contrasts different machine learning techniques with artificial neural networks and thereby provides insights on the applicability of newer data analysis techniques.

Practitioners such as investors, advisers, and support programs can benefit significantly from these results. The proposed prediction models provide these parties with a decision support system to conduct a first analysis of possible successful entrepreneurs and entrepreneurs who are likely to abandon the venture. Therefore, the proposed systems provide the mentioned parties with a first possibility to save costs by optimally allocating monetary and non-monetary resources in supporting promising and successful nascent entrepreneurs, while at the same time providing them with an alert system to identify entrepreneurs who may be at the brink of abandoning the venture. As a result, interventions are accomplished in a timely and effective manner.

6.2 Limitations and future research

Despite providing valuable insights into the firm emergence process, the three papers are subject to limitations. While detailed specifications of the limitations can be found in the corresponding chapters, in the following, the main limitations of each paper are summarized, while research areas of great promise that academics could address in future studies are highlighted.

The scale in *paper one* that captures regulatory focus in an entrepreneurial setting is of an exploratory nature, and given the many valid instruments to measure individual self-regulation (Haws et al., 2010), the results of the study may deviate when different measures are employed. Thus, further research applying the scale in the entrepreneurial context would be fruitful. Moreover, the classification of activities within the discovery and exploitation phase is still open to interpretation. Building upon the activities, business creation can be seen as an idiosyncratic process (Newbert, 2005), and depending on the type of business created, not all activities may be relevant to enable firm birth. Therefore, an extension of the proposed relationship within different industries could provide further insights into better understanding the successful emergence process. Additionally, given that not only activities but specific tasks are influenced by the regulatory focus (Brockner et al., 2004), a further breakdown of the exact type of tasks could prove valuable to delineate the evolvement of the gestation process.

For *paper two*, the measures for imitation and innovation are based on the judgments of individuals and thus could confound objective innovativeness with perceptual biases of the entrepreneurs (Hyytinen et al., 2015; Koellinger, 2008). Consequently, an objective ex-ante measurement for the innovativeness of a venture could prove extremely valuable for academia and industry. Moreover, the scale used for measuring various social identity types is still of an exploratory character. Thus, future research could apply recently developed scales (cf. Sieger et al., 2016) in a same setting to further scrutinize the results. Finally, the investigations in the first and second papers utilized data sourced from Switzerland. While offering a high level of generalizability through the longitudinal investigations, the insights and results yielded are mostly representative of the Swiss ecosystem and culture. While cognitive differences exist among countries (Meisenberg & Woodley, 2013), future avenues in the form of comparative studies in other countries could prove valuable.

While for *paper three* a diverse set of variables are included, the results indicate that the list is not conclusive. Therefore, future research is encouraged to examine additional emergence impact factors, thus possibly providing additional predictive power. Moreover, even though the

harmonized PSED dataset provides a significant pool of information, with its size and its longitudinal character, further data could not only improve the training process of the machine learning model but also increase the applicability of artificial or deep neural networks. Additionally, even though the applied methodologies are optimized via commonly accepted techniques, the parameters could be further tuned by using higher computational power, thus increasing the model performance even more. Consequently, the results of this paper build the fundament for applying innovative data mining methodologies for future research in the entrepreneurial field.

6.3 Conclusion

This cumulative dissertation reveals that a successful firm emergence process is influenced not only by the type of business idea (i.e., perceived innovative and imitative ideas) and activities (i.e., discovery and exploitation activities) but also by the entrepreneurial cognition (i.e., regulatory focus, social identity theory). The cumulative dissertation makes a significant contribution to the academic understanding of the firm emergence process by highlighting the role of specific cognitive factors and by evaluating the applicability of machine learning techniques and artificial neural networks to predict the early-stage venture outcome. The generated insights provide not only practical implications for the entrepreneur but also for the involved parties who provide guidance as well as financial support. The findings of the three papers and the applied machine learning models enable scholars and practitioners to thoroughly study and analyze the likelihood of the successful birth of a firm as well as the likelihood of firm abandonment. In this regard, practitioners are encouraged to apply predictive models to reduce possible costs related to the entrepreneurial process. For scholars, this dissertation encourages the expansion of research into the impact of cognitive factors on the venture outcome as well as the potential of up-to-date research methodologies to derive new insights in the study of the successful gestation process.

References

Abernathy, W. J., & Clark, K. B. (1985). Innovation: Mapping the winds of creative destruction. *Research Policy*, *14*(1), 3–22. https://doi.org/10.1016/0048-7333(85)90021-6

Abrams, D., & Hogg, M. A. (1988). Comments on the motivational status of self-esteem in social identity and intergroup discrimination. *European Journal of Social Psychology*, *18*(4), 317–334. https://doi.org/10.1002/ejsp.2420180403

Adomako, S. (2017). CEOs' regulatory foci and firm-level product innovativeness in competitive environments. *Journal of Business & Industrial Marketing*, *32*(5), 640–651. https://doi.org/10.1108/JBIM-06-2016-0126

Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing. https://doi.org/10.1007/978-3-319-14142-8

Aldrich, H. E., & Fiol, C. M. (1994). Entrepreneurship. *The Academy of Management Review*, *19*(4), 645–670. https://doi.org/10.1007/978-3-540-48543-8_5

Aldrich, H. E., & Martinez, M. A. (2001). Many are Called, but Few are Chosen: An Evolutionary Perspective for the Study of Entrepreneurship. *Entrepreneurship Theory and Practice*, *25*(4), 41–56. https://doi.org/10.1177/104225870102500404

Aldrich, H. E., & Ruef, M. (2006). *Organizations Evolving* (2nd ed.). SAGE Publications Ltd. https://doi.org/10.4135/9781446212509

Alharthi, H. (2019). Predicting physicians' satisfaction with electronic medical records using artificial neural network modeling. *Saudi Journal for Health Sciences*, *8*(2), 105. https://doi.org/10.4103/sjhs.sjhs_14_19

Amason, A. C., Shrader, R. C., & Tompson, G. H. (2006). Newness and novelty: Relating top management team composition to new venture performance. *Journal of Business Venturing*, *21*(1), 125–148. https://doi.org/10.1016/j.jbusvent.2005.04.008

Amir, M. T. (2015). Entrepreneurial Behavior and Innovative Behavior: A Conceptual Clarification. *The Asian Journal of Technology Management*, *8*(2), 160–171. https://doi.org/10.12695/ajtm.2015.8.2.7

Andrikopoulos, P., & Khorasgani, A. (2018). Predicting unlisted SMEs' default: Incorporating market information on accounting-based models for improved accuracy. *British Accounting Review*, *50*(5), 559–573. https://doi.org/10.1016/j.bar.2018.02.003

Arenius, P., Engel, Y., & Klyver, K. (2017). No particular action needed? A necessary condition analysis of gestation activities and firm emergence. *Journal of Business Venturing Insights*, *8*, 87–92. https://doi.org/10.1016/j.jbvi.2017.07.004

Audretsch, D. B., Bönte, W., & Keilbach, M. (2008). Entrepreneurship capital and its impact on knowledge diffusion and economic performance. *Journal of Business Venturing*, *23*(6), 687–698. https://doi.org/10.1016/j.jbusvent.2008.01.006

Autio, E., Dahlander, L., & Frederiksen, L. (2013). Information Exposure, Opportunity Evaluation, and Entrepreneurial Action: An Investigation of an Online User Community. *Academy of Management Journal*, *56*(5), 1348–1371. https://doi.org/10.5465/amj.2010.0328

Avnet, T., & Tory Higgins, E. (2006). Response to comments on "how regulatory fit affects value in consumer choices and opinions." *Journal of Marketing Research*, *43*(1), 24–27. https://doi.org/10.1509/jmkr.43.1.24

Bahl, A., Hellack, B., Balas, M., Dinischiotu, A., Wiemann, M., Brinkmann, J., Luch, A., Renard, B. Y., & Haase, A. (2019). Recursive feature elimination in random forest classification supports nanomaterial grouping. *NanoImpact*, *15*, 100179. https://doi.org/10.1016/j.impact.2019.100179

Baidoun, S. D., Lussier, R. N., Burbar, M., & Awashra, S. (2018). Prediction model of business success or failure for Palestinian small enterprises in the West Bank. *Journal of Entrepreneurship in Emerging Economies*, *10*(1), 60–80. https://doi.org/10.1108/JEEE-02-2017-0013

Baron, R. A. (2004). The cognitive perspective: A valuable tool for answering entrepreneurship's basic "why" questions. *Journal of Business Venturing*, *19*(2), 221–239. https://doi.org/10.1016/S0883-9026(03)00008-9

Baron, R. A. (2007). Behavioral and cognitive factors in entrepreneurship. *Strategic Entrepreneurship Journal*, *1*(1), 167–182. https://doi.org/10.1002/sej

Baron, R. A. (2012). *Entrepreneurship: An Evidence-based Guide*. Edward Elgar.

Baron, R. A., Mueller, B. A., & Wolfe, M. T. (2016). Self-efficacy and entrepreneurs' adoption of unattainable goals: The restraining effects of self-control. *Journal of Business Venturing*, *31*(1), 55–71. https://doi.org/10.1016/j.jbusvent.2015.08.002

Barsky, A. P., & Zyphur, M. J. (2016). Disentangling sunk-costs and completion proximity: The role of regulatory focus. *Journal of Experimental Social Psychology*, *65*, 105–108. https://doi.org/10.1016/j.jesp.2016.04.006

Baum, J. R., & Locke, E. A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology*, *89*(4), 587–598. https://doi.org/10.1037/0021-9010.89.4.587

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*, 281–305.

Beyond Meat Inc. (n.d.). *Our Mission*. https://www.beyondmeat.com/

Bian, S., Deng, Z., Li, F., Monroe, W., Shi, P., Sun, Z., Wu, W., Wang, S., Wang, W. Y., Yuan, A., Zhang, T., & Li, J. (2018). IcoRating: A Deep-Learning System for Scam ICO Identification. *ArXiv Preprint*. http://arxiv.org/abs/1803.03670

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *14*(1), 106. https://doi.org/10.1186/1471-2105-14-106

Blanco-Oliver, A., Pino-Mejías, R., & Lara-Rubio, J. (2014). Modeling the Financial Distress of Microenterprise Start- Ups Using Support Vector Machines: A Case Study. *Innovar*, *24*(1), 153–168. https://doi.org/10.15446/innovar.v24n1spe.47615

Boyer, T., & Blazy, R. (2013). Born to be alive? The survival of innovative and non-innovative French micro-start-ups. *Small Business Economics*, *42*(4), 669–683. https://doi.org/10.1007/s11187-013-9522-8

Brändle, L., Berger, E. S. C., Golla, S., & Kuckertz, A. (2018). I am what I am - How nascent entrepreneurs' social identity affects their entrepreneurial self-efficacy. *Journal of Business Venturing Insights*, *9*, 17–23. https://doi.org/10.1016/j.jbvi.2017.12.001

Brinckmann, J., Grichnik, D., & Kapsa, D. (2010). Should entrepreneurs plan or just storm the castle? A meta-analysis on contextual factors impacting the business planning-performance relationship in small firms. *Journal of Business Venturing*, *25*(1), 24–40. https://doi.org/10.1016/j.jbusvent.2008.10.007

Brockner, J., Higgins, E. T., & Low, M. B. (2004). Regulatory focus theory and the entrepreneurial process. *Journal of Business Venturing*, *19*(2), 203–220. https://doi.org/10.1016/S0883-9026(03)00007-7

Brouder, P., & Eriksson, R. H. (2013). Staying Power: What Influences Micro-firm Survival in Tourism? *Tourism Geographies*, *15*(1), 125–144. https://doi.org/10.1080/14616688.2011.647326

Brush, C. G., Manolova, T. S., & Edelman, L. F. (2008). Properties of emerging organizations: An empirical test. *Journal of Business Venturing*, *23*(5), 547–566. https://doi.org/10.1016/j.jbusvent.2007.09.002

Bryant, P. (2007). Self-regulation and decision heuristics in entrepreneurial opportunity evaluation and exploitation. *Management Decision*, *45*(4), 732–748. https://doi.org/10.1108/00251740710746006

Buddelmeyer, H., Jensen, P. H., & Webster, E. (2010). Innovation and the determinants of company survival. *Oxford Economic Papers*, *62*(2), 261–285. https://doi.org/10.1093/oep/gpp012

Burmeister-Lamp, K., Lévesque, M., & Schade, C. (2012). Are entrepreneurs influenced by risk attitude, regulatory focus or both? An experiment on entrepreneurs' time allocation. *Journal of Business Venturing*, *27*(4), 456–476. https://doi.org/10.1016/j.jbusvent.2011.12.001

Burnaev, E., Erofeev, P., & Papanov, A. (2015, December 8). *Influence of resampling on accuracy of imbalanced classification.* Eighth International Conference on Machine Vision (ICMV 2015), Barcelona. https://doi.org/10.1117/12.2228523

Cardon, M. S., Wincent, J., Singh, J., & Drnovsek, M. (2009). The nature and experience of entrepreneurial passion. *Academy of Management Review*, *34*(3), 511–532. https://doi.org/10.5465/amr.2009.40633190

Carter, N. M., Gartner, W. B., & Reynolds, P. D. (1996). Exploring start-up event sequences. *Journal of Business Venturing*, *11*(3), 151–166. https://doi.org/10.1016/0883-9026(95)00129-8

Carter, R., & Auken, H. Van. (2006). Small Firm Bankruptcy. *Journal of Small Business Management*, *44*(4), 493–512. https://doi.org/10.1111/j.1540-627X.2006.00187.x

Casadesus-Masanell, R., & Zhu, F. (2013). Business model innovation and competitive imitation: The case of sponsor-based business models. *Strategic Management Journal*, *34*(4), 464–482. https://doi.org/10.1002/smj.2022

Cefis, E., & Marsili, O. (2005). A matter of life and death: innovation and firm survival. *Industrial and Corporate Change*, *14*(6), 1167–1192. https://doi.org/10.1093/icc/dth081

Cefis, E., & Marsili, O. (2006). Survivor: The role of innovation in firms' survival. *Research Policy*, *35*(5), 626–641. https://doi.org/10.1016/j.respol.2006.02.006

Cefis, E., & Marsili, O. (2012). Going, going, gone. Exit forms and the innovative capabilities of firms. *Research Policy*, *41*(5), 795–807. https://doi.org/10.1016/j.respol.2012.01.006

Chandler, G. N., Honig, B., & Wiklund, J. (2005). Antecedents, moderators, and performance consequences of membership change in new venture teams. *Journal of Business Venturing*, *20*(5), 705–725. https://doi.org/10.1016/j.jbusvent.2004.09.001

Chaudhuri, N., & Bose, I. (2020). Exploring the role of deep neural networks for post-disaster decision support. *Decision Support Systems*, *130*, 113234. https://doi.org/10.1016/j.dss.2019.113234

Chen, J., Jin, Q., & Chao, J. (2012). Design of Deep Belief Networks for Short-Term Prediction of Drought Index Using Data in the Huaihe River Basin. *Mathematical Problems in Engineering*, *6*, 1–16. https://doi.org/10.1155/2012/235929

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, 785–794. https://doi.org/10.1145/2939672.2939785

Ching, T., Zhu, X., & Garmire, L. X. (2018). Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Computational Biology*, *14*(4), 1–18. https://doi.org/10.1371/journal.pcbi.1006076

Chrisman, J. J. (1999). The influence of outsider-generated knowledge resources on venture creation. *Journal of Small Business Management*, *37*(4), 42–58.

Chuang, L.-M., & Chao, S.-T. (2013). Evolution of National Entrepreneurial Opportunity Recognition: A Neural Network Analysis. *Journal of Advanced Management Science*, *1*(1), 66–70. https://doi.org/10.12720/joams.1.1.66-70

Chwolka, A., & Raith, M. G. (2012). The value of business planning before start-up - A decision-theoretical perspective. *Journal of Business Venturing*, *27*(3), 385–399. https://doi.org/10.1016/j.jbusvent.2011.01.002

Cliff, J. E., Jennings, P. D., & Greenwood, R. (2006). New to the game and questioning the rules: The experiences and beliefs of founders who start imitative versus innovative firms. *Journal of Business Venturing*, *21*(5), 633–663. https://doi.org/10.1016/j.jbusvent.2005.02.010

Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. *Journal of Business Research*, *101*, 885–896. https://doi.org/10.1016/j.jbusres.2018.11.015

Cohen, W. M., & Klepper, S. (1996). Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D. *The Review of Economics and Statistics*, *78*(2), 232. https://doi.org/10.2307/2109925

Colombelli, A., Krafft, J., & Quatraro, F. (2013). Properties of knowledge base and firm survival: Evidence from a sample of French manufacturing firms. *Technological Forecasting and Social Change*, *80*(8), 1469–1483. https://doi.org/10.1016/j.techfore.2013.03.003

Cooper, A. C., Gimeno-Gascon, F. J., & Woo, C. Y. (1991). A resource-based prediction of new venture survival and growth. *Academy of Management Proceedings*, *1*, 68–72. https://doi.org/10.5465/ambpp.1991.4976561

Cortez, P., & Embrechts, M. J. (2013). Using sensitivity analysis and visualization techniques to open black box data mining models. *Information Sciences*, *225*, 1–17. https://doi.org/10.1016/j.ins.2012.10.039

Cowden, B. J., & Bendickson, J. S. (2018). Impacts of regulatory focus and institutions on innovation. *Management Decision*, *56*(5), 939–954. https://doi.org/10.1108/MD-11-2016-0826

Crowe, E., & Higgins, E. T. (1997). Regulatory Focus and Strategic Inclinations: Promotion and Prevention in Decision-Making. *Organizational Behavior and Human Decision Processes*, *69*(2), 117–132. https://doi.org/10.1006/obhd.1996.2675

D'souza, R. N., Huang, P.-Y., & Yeh, F.-C. (2020). Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size. *Scientific Reports*, *10*(1), 834. https://doi.org/10.1038/s41598-020-57866-2

Daft, R. L., & Weick, K. E. (1984). Toward a Model of Organizations as Interpretation Systems. *Academy of Management Review*, *9*(2), 284–295. https://doi.org/10.5465/amr.1984.4277657

Dahl, G. E., Dong Yu, Li Deng, & Acero, A. (2012). Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(1), 30–42. https://doi.org/10.1109/TASL.2011.2134090

Dahlqvist, J. (2007). *Assessing New Economic Activity: Process and performance in new ventures.* (Publication No. 32) [Doctoral dissertation, Jönköping International Business School]. JIBS Dissertation Series.

Dahlqvist, J., & Wiklund, J. (2012). Measuring the market newness of new ventures. *Journal of Business Venturing*, *27*(2), 185–196. https://doi.org/10.1016/j.jbusvent.2010.12.001

Davidsson, P. (2003). The domain of entrepreneurship research: some suggestions. In J. A. Katz & D. A. Shepherd (Eds.), *Cognitive Approaches to Entrepreneurship Research (Advances in Entrepreneurship, Firm Emergence and Growth, Vol. 6)* (pp. 315–372). Emerald Group Publishing Limited. https://doi.org/10.1016/S1074-7540(03)06010-0

Davidsson, P. (2015). Data replication and extension: A commentary. *Journal of Business Venturing Insights*, *3*, 12–15. https://doi.org/10.1016/j.jbvi.2015.02.001

Davidsson, P. (2016). *Researching Entrepreneurship - Conceptualization and Design*. Springer International Publishing AG Switzerland.

Davidsson, P., & Gordon, S. R. (2012). Panel studies of new venture creation: A methods-focused review and suggestions for future research. *Small Business Economics*, *39*(4), 853–876. https://doi.org/10.1007/s11187-011-9325-8

Davidsson, P., & Honig, B. (2003). The role of social and human capital among nascent entrepreneurs. *Journal of Business Venturing*, *18*(3), 301–331. https://doi.org/10.1016/S0883-9026(02)00097-6

De Carolis, D. M., Litzky, B. E., & Eddleston, K. A. (2009). Why Networks Enhance the Progress of New Venture Creation: The Influence of Social Capital and Cognition. *Entrepreneurship Theory and Practice*, *33*(2), 527–545. https://doi.org/10.1111/j.1540-6520.2009.00302.x

Delmar, F. (2015). A response to Honig and Samuelsson (2014). *Journal of Business Venturing Insights*, *3*, 1–4. https://doi.org/10.1016/j.jbvi.2014.11.002

Delmar, F., & Shane, S. (2003). Does business planning facilitate the development of new ventures? *Strategic Management Journal*, *24*(12), 1165–1185. https://doi.org/10.1002/smj.349

Delmar, F., & Shane, S. (2004). Legitimating first: Organizing activities and the survival of new ventures. *Journal of Business Venturing*, *19*(3), 385–410. https://doi.org/10.1016/S0883-9026(03)00037-5

Delmar, F., & Shane, S. (2006). Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures. *Strategic Organization*, *4*(3), 215–247. https://doi.org/10.1177/1476127006066596

DeTienne, D. R., Shepherd, D. A., & De Castro, J. O. (2008). The fallacy of "only the strong survive": The effects of extrinsic motivation on the persistence decisions for under-performing firms. *Journal of Business Venturing*, *23*(5), 528–546. https://doi.org/10.1016/j.jbusvent.2007.09.004

Dimaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48*(2), 147–160. https://doi.org/10.2307/2095101

Dimov, D. (2010). Nascent Entrepreneurs and Venture Emergence: Opportunity Confidence, Human Capital, and Early Planning. *Journal of Management Studies*, *47*(6), 1123–1153. https://doi.org/10.1111/j.1467-6486.2009.00874.x

Diochon, M., Menzies, T. V., & Gasse, Y. (2005a). Canadian Nascent Entrepreneurs' Start-up Efforts: Outcomes and Individual Influences on Sustainability. *Journal of Small Business and Entrepreneurship*, *18*(1), 53–74. https://doi.org/10.1080/08276331.2005.10593332

Diochon, M., Menzies, T. V., & Gasse, Y. (2005b). Exploring the relationship between start-up activities and new venture emergence: a longitudinal study of Canadian nascent entrepreneurs. *International Journal of Management and Enterprise Development*, *2*(3–4), 408–426. https://doi.org/10.1504/ijmed.2005.006570

Diochon, M., Menzies, T. V., & Gasse, Y. (2008). Exploring the nature and impact of gestation-specific human capital among nascent entrepreneurs. *Journal of Developmental Entrepreneurship*, *13*(2), 151–165. https://doi.org/10.1142/S1084946708000909

du Jardin, P. (2018). Failure pattern-based ensembles applied to bankruptcy forecasting. *Decision Support Systems*, *107*, 64–77. https://doi.org/10.1016/j.dss.2018.01.003

Dyer, J. H., Gregersen, H. B., & Christensen, C. (2008). Entrepreneur behaviors, opportunity recognition, and the origins of innovative ventures. *Strategic Entrepreneurship Journal*, *2*(4), 317–338. https://doi.org/10.1002/sej.59

Epure, M., & Guasch, M. (2020). Debt signaling and outside investors in early stage firms. *Journal of Business Venturing*, *35*(2), 105929. https://doi.org/10.1016/j.jbusvent.2019.02.002

Ethiraj, S. K., Levinthal, D., & Roy, R. R. (2008). The Dual Role of Modularity: Innovation and Imitation. *Management Science*, *54*(5), 939–955. https://doi.org/10.1287/mnsc.1070.0775

Ethiraj, S. K., & Zhu, D. H. (2008). Performance effects of imitative entry. *Strategic Management Journal*, *29*(8), 797–817. https://doi.org/10.1002/smj.696

Farmer, S. M., Yao, X., & Kung-Mcintyre, K. (2011). The Behavioral Impact of Entrepreneur Identity Aspiration and Prior Entrepreneurial Experience. *Entrepreneurship Theory and Practice*, *35*(2), 245–273. https://doi.org/10.1111/j.1540-6520.2009.00358.x

Fauchart, E., & Gruber, M. (2011). Darwinians, Communitarians, and Missionaries: The Role of Founder Identity in Entrepreneurship. *Academy of Management Journal*, *54*(5), 935–957. https://doi.org/10.5465/amj.2009.0211

Fertsch, M. (2020). Artificial neural pseudo-network for production control purposes. *Logforum*, *16*(1), 7–13.

Fischer, D., Mauer, R., & Brettel, M. (2018). Regulatory focus theory and sustainable entrepreneurship. *International Journal of Entrepreneurial Behaviour and Research*, *24*(2), 408–428. https://doi.org/10.1108/IJEBR-12-2015-0269

Fisher, G. (2012). Effectuation, causation, and bricolage: A behavioral comparison of emerging theories in entrepreneurship research. *Entrepreneurship Theory and Practice*, *36*(5), 1019–1051. https://doi.org/10.1111/j.1540-6520.2012.00537.x

Forbes, D. P. (1999). Cognitive approaches to new venture creation. *International Journal of Management Reviews*, *1*(4), 415–439. https://doi.org/10.1111/1468-2370.00021

Förster, J., Higgins, E. T., & Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes*, *90*(1), 148–164. https://doi.org/10.1016/S0749-5978(02)00509-5

Freeland, R. E., & Keister, L. A. (2016). How Does Race and Ethnicity Affect Persistence in Immature Ventures? *Journal of Small Business Management*, *54*(1), 210–228. https://doi.org/10.1111/jsbm.12138

Freeman, J., Carroll, G. R., & Hannan, M. T. (1983). The Liability of Newness: Age Dependence in Organizational Death Rates. *American Sociological Review*, *48*(5), 692. https://doi.org/10.2307/2094928

Freitas, A. L., & Higgins, E. T. (2002). Enjoying goal-directed action: The role of regulatory fit. *Psychological Science*, *13*(1), 1–6. https://doi.org/10.1111/1467-9280.00401

Frid, C. J., Wyman, D. M., & Coffey, B. (2016). Effects of wealth inequality on entrepreneurship. *Small Business Economics*, *47*(4), 895–920. https://doi.org/10.1007/s11187-016-9742-9

Frid, C. J., Wyman, D. M., & Gartner, W. B. (2015). The influence of financial "skin in the game" on new venture creation. *Academy of Entrepreneurship Journal*, *21*(2), 1–14.

Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Fu, Y., & Tiez, M. A. (2015). When copycats can succeed? Business model imitation and the legitimacy of new ventures in emerging economies. *Frontiers of Entrepreneurship Research*, *35*(16), 412.

Fu, Y., & Tiez, M. A. (2017). When do investors prefer copycats? Conditions influencing the evaluation of innovative and imitative ventures (Summary). *Frontiers of Entrepreneurship Research*, *37*(1), 21.

Galbraith, C. S., DeNoble, A. F., Ehrlich, S. B., & Kline, D. M. (2007). Can experts really assess future technology success? A neural network and Bayesian analysis of early stage technology proposals. *The Journal of High Technology Management Research*, *17*(2), 125–137. https://doi.org/10.1016/j.hitech.2006.11.002

Gamache, D. L., McNamara, G., Mannor, M. J., & Johnson, R. E. (2015). Motivated to acquire? the impact of CEO regulatory focus on firm acquisitions. *Academy of Management Journal*, *58*(4), 1261–1282. https://doi.org/10.5465/amj.2013.0377

Gartner, W. B., & Liao, J. (2012). The effects of perceptions of risk, environmental uncertainty, and growth aspirations on new venture creation success. *Small Business Economics*, *39*(3), 703–712. https://doi.org/10.1007/s11187-011-9356-1

Gartner, W. B., & Shaver, K. G. (2012). Nascent entrepreneurship panel studies: Progress and challenges. *Small Business Economics*, *39*(3), 659–665. https://doi.org/10.1007/s11187-011-9353-4

Gatewood, E. J., Shaver, K. G., & Gartner, W. B. (1995). A longitudinal study of cognitive factors influencing start-up behaviors and success at venture creation. *Journal of Business Venturing*, *10*(5), 371–391. https://doi.org/10.1016/0883-9026(95)00035-7

Gerasimovic, M., & Bugaric, U. (2018). Enrollment Management Model: Artificial Neural Networks versus Logistic Regression. *Applied Artificial Intelligence*, *32*(2), 153–164. https://doi.org/10.1080/08839514.2018.1448146

Giotopoulos, I., Kontolaimou, A., & Tsakanikas, A. (2017). Drivers of high-quality entrepreneurship: what changes did the crisis bring about? *Small Business Economics*, *48*(4), 913–930. https://doi.org/10.1007/s11187-016-9814-x

Gordon, S. R. (2012). *Dimensions of the venture creation process: Amount, dynamics, and sequences of action in nascent entrepreneurship* (Publication No. 58078) [Doctoral dissertation, Queensland University of Technology]. https://eprints.qut.edu.au/58078/#?

Grégoire, D. A., Corbett, A. C., & Mcmullen, J. S. (2011). The Cognitive Perspective in Entrepreneurship: An Agenda for Future Research. *Journal of Management Studies*, *48*(6), 1443–1477. https://doi.org/10.1111/j.1467-6486.2010.00922.x

Grégoire, D. A., Cornelissen, J., Dimov, D., & van Burg, E. (2015). The mind in the middle: Taking stock of affect and cognition research in entrepreneurship. *International Journal of Management Reviews*, *17*(2), 125–142. https://doi.org/10.1111/ijmr.12060

Guo, R. (2019). Effectuation, opportunity shaping and innovation strategy in high-tech new ventures. *Management Decision*, *57*(1), 115–130. https://doi.org/10.1108/MD-08-2017-0799

Gupta, J., Barzotto, M., & Khorasgani, A. (2018). Does size matter in predicting SMEs failure? *International Journal of Finance and Economics*, *23*(4), 571–605. https://doi.org/10.1002/ijfe.1638

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. NJ: Prentice Hall.

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In D.-S. Huang, X.-P. Zhang, & G.-B. Huang (Eds.), *Advances in Intelligent Computing* (pp. 878–887). Springer-Verlag Berlin Heidelberg.

Haws, K. L., Dholakia, U. M., & Bearden, W. O. (2010). An assessment of chronic regulatory focus measures. *Journal of Marketing Research*, *47*(5), 967–982. https://doi.org/10.1509/jmkr.47.5.967

Hechavarría, D. M., Matthews, C. H., & Reynolds, P. D. (2016). Does start-up financing influence start-up speed? Evidence from the panel study of entrepreneurial dynamics. *Small Business Economics*, *46*(1), 137–167. https://doi.org/10.1007/s11187-015-9680-y

Hechavarria, D. M., Renko, M., & Matthews, C. H. (2012). The nascent entrepreneurship hub: Goals, entrepreneurial self-efficacy and start-up outcomes. *Small Business Economics*, *39*(3), 685–701. https://doi.org/10.1007/s11187-011-9355-2

Heeg, T., Jansen, J., & Knop, C. (2016, March 26). Die deutsche Internetrakete. *Frankfurter Allgemeine*. https://www.faz.net/aktuell/wirtschaft/unternehmen/oliver-samwer-erklaert-das-geschaeftsmodell-von-rocket-internet-14145109-p2.html

Helmers, C., & Rogers, M. (2010). Innovation and the survival of new firms in the UK. *Review of Industrial Organization*, *36*(3), 227–248. https://doi.org/10.1007/s11151-010-9247-7

Heydari, S. T., Ayatollahi, S. M. T., & Zare, N. (2012). Comparison of artificial neural networks with logistic regression for detection of obesity. *Journal of Medical Systems*, *36*(4), 2449–2454. https://doi.org/10.1007/s10916-011-9711-4

Higgins, E. T. (1998). Promotion and prevention: regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology.* (pp. 1–46). Academic Press.

Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist*, *55*(11), 1217–1230. https://doi.org/10.1037//0003-066X.55.11.1217

Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: Promotion pride versus prevention pride. *European Journal of Social Psychology*, *31*(1), 3–23. https://doi.org/10.1002/ejsp.27

Higgins, E. T., & Spiegel, S. (2004). Promotion and prevention strategies for self-regulation: A motivated cognition perspective. In E. Roy, F. Baumeister, & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (2nd ed.). Guilford.

Higgins, E. T., & Tykocinski, O. (1992). Self-discrepancies and biographical memory: Personality and cognition at the level of psychological situation. *Personality and Social Psychology Bulletin*, *18*(5), 527–535. https://doi.org/10.1177/0146167292185002

Hmielseki, K. M., & Baron, R. A. (2008). Regulatory Focus and new venture performance: A study of entrepreneurial opportunity exploitation under conditions of risk versus uncertainty. *Strategic Entrepreneurship Journal*, *2*(4), 285–299. https://doi.org/10.1002/sej.56

Hoang, H., & Gimeno, J. (2010). Becoming a founder: How founder role identity affects entrepreneurial transitions and persistence in founding. *Journal of Business Venturing*, *25*(1), 41–53. https://doi.org/10.1016/j.jbusvent.2008.07.002

Holland, D. V., & Shepherd, D. A. (2013). Deciding to Persist: Adversity, Values, and Entrepreneurs' Decision Policies. *Entrepreneurship Theory and Practice*, *37*(2), 331–358. https://doi.org/10.1111/j.1540-6520.2011.00468.x

Honig, B., & Karlsson, T. (2004). Institutional forces and the written business plan. *Journal of Management*, *30*(1), 29–48. https://doi.org/10.1016/j.jm.2002.11.002

Honig, B., & Samuelsson, M. (2012). Planning and the Entrepreneur: A Longitudinal Examination of Nascent Entrepreneurs in Sweden. *Journal of Small Business Management*, *50*(3), 365–388. https://doi.org/10.1111/j.1540-627X.2012.00357.x

Honig, B., & Samuelsson, M. (2014). Data replication and extension: A study of business planning and venture-level performance. *Journal of Business Venturing Insights*, *1–2*, 18–25. https://doi.org/10.1016/j.jbvi.2014.09.006

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–59. https://doi.org/10.21427/D7CF7R

Hopp, C. (2012). For Better or for Worse?—Nonprofit Experience and the Performance of Nascent Entrepreneurs. *Nonprofit and Voluntary Sector Quarterly*, *41*(6), 1251–1268. https://doi.org/10.1177/0899764012449517

Hopp, C., & Sonderegger, R. (2015). Understanding the Dynamics of Nascent Entrepreneurship-Prestart-Up Experience, Intentions, and Entrepreneurial Success. *Journal of Small Business Management*, *53*(4), 1076–1096. https://doi.org/10.1111/jsbm.12107

Hopp, C., & Stephan, U. (2012). The influence of socio-cultural environments on the performance of nascent entrepreneurs: Community culture, motivation, self-efficacy and start-up success. *Entrepreneurship & Regional Development*, *24*(9–10), 917–945. https://doi.org/10.1080/08985626.2012.742326

Hu, G., Peng, X., Yang, Y., Hospedales, T. M., & Verbeek, J. (2018). Frankenstein: Learning Deep Face Representations Using Small Data. *IEEE Transactions on Image Processing*, *27*(1), 293–303. https://doi.org/10.1109/TIP.2017.2756450

Huang, L., & Pearce, J. L. (2015). Managing the Unknowable. *Administrative Science Quarterly*, *60*(4), 634–670. https://doi.org/10.1177/0001839215597270

Huang, W., Song, G., Hong, H., & Xie, K. (2014). Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning. *IEEE Transactions on Intelligent Transportation Systems*, *15*(5), 2191–2201. https://doi.org/10.1109/TITS.2014.2311123

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, *37*(4), 543–558. https://doi.org/10.1016/S0167-9236(03)00086-1

Hüsing, A. (2011, December 18). *Umbruch bei Rocket Internet: Mitten im Wandel zum globalen Internetkonzern geht fast das komplette Führungsteam*. https://www.deutsche-startups.de/2011/12/18/rocket-internet-umbruch/

Hyder, S., & Lussier, R. N. (2016). Why businesses succeed or fail: a study on small businesses in Pakistan. *Journal of Entrepreneurship in Emerging Economies*, *8*(1), 82–100. https://doi.org/10.1108/JEEE-03-2015-0020

Hyytinen, A., Pajarinen, M., & Rouvinen, P. (2015). Does innovativeness reduce startup survival rates? *Journal of Business Venturing*, *30*(4), 564–581. https://doi.org/10.1016/j.jbusvent.2014.10.001

Jarzebowski, A., & Palermo, J. (2012). When feedback is not enough : motivation after positive feedback. *International Coaching Psychology Review*, *7*(1), 14–32.

Joharestani, M. Z., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S. (2019). PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere*, *10*(7), 12–18. https://doi.org/10.3390/atmos10070373

Johnson, P. D., Smith, M. B., Wallace, J. C., Hill, A. D., & Baron, R. A. (2015). A Review of Multilevel Regulatory Focus in Organizations. *Journal of Management*, *41*(5), 1501–1529. https://doi.org/10.1177/0149206315575552

Karoly, P. (1993). Mechanisms of Self-Regulation: A Systems View. *Annual Review of Psychology*, *44*(1), 23–52. https://doi.org/10.1146/annurev.psych.44.1.23

Katz, J., & Gartner, W. B. (1988). Properties of Emerging Organizations. *Academy of Management Review*, *13*(3), 429–441. https://doi.org/10.5465/AMR.1988.4306967

Khan, S. A., Tang, J., & Joshi, K. (2014). Disengagement of Nascent Entrepreneurs from the Start-Up Process. *Journal of Small Business Management*, *52*(1), 39–58. https://doi.org/10.1111/jsbm.12032

Kiang, M. Y. (2003). A comparative assessment of classification methods. *Decision Support Systems*, *35*(4), 441–454. https://doi.org/10.1016/S0167-9236(02)00110-0

Kier, A., McMullen, J. S., & Kuratko, D. (2014). Trapped By the Entrepreneurial Mindset? an Examination of Regulatory Focus and Escalation of Commitment in Entrepreneurial Action. *Frontiers of Entrepreneurship Research*, *34*(5), 1–16.

Kim, H. J., Kim, T. S., & Sohn, S. Y. (2020). Recommendation of startups as technology cooperation candidates from the perspectives of similarity and potential: A deep learning approach. *Decision Support Systems*, *130*, 113229. https://doi.org/10.1016/j.dss.2019.113229

Koellinger, P. (2008). Why are some entrepreneurs more innovative than others? *Small Business Economics*, *31*(1), 21–37. https://doi.org/10.1007/s11187-008-9107-0

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, *104*, 38–48. https://doi.org/10.1016/j.dss.2017.10.001

Kröper, M., Fay, D., Lindberg, T., & Meinel, C. (2011). Interrelations between Motivation, Creativity and Emotions in Design Thinking Processes – An Empirical Study Based on Regulatory Focus Theory. In T. Taura & Y. Nagai (Eds.), *Design Creativity 2010* (pp. 97–104). Springer London. https://doi.org/10.1007/978-0-85729-224-7_14

Krueger, N. (2007). What Lies Beneath? The Experiential Essence of Entrepreneurial Thinking. *Entrepreneurship Theory and Practice, 30*(1), 123-138. https://doi.org/10.1111/j.1540-6520.2007.00166.x.

Ku, G. (2008). Learning to de-escalate: The effects of regret in escalation of commitment. *Organizational Behavior and Human Decision Processes*, *105*(2), 221–232. https://doi.org/10.1016/j.obhdp.2007.08.002

Kuckertz, A., Kollmann, T., Krell, P., & Stöckmann, C. (2017). Understanding, differentiating, and measuring opportunity recognition and opportunity exploitation. *International Journal of Entrepreneurial Behaviour and Research*, *23*(1), 78–97. https://doi.org/10.1108/IJEBR-12-2015-0290

Laird, J. J. (2012). Valuing the quality of strategic ferry services to remote communities. *Research in Transportation Business & Management*, *4*(5), 97–103. https://doi.org/10.1016/j.rtbm.2012.06.013

Laurikkala, J. (2002). Instance-based data reduction for improved identification of difficult small classes. *Intelligent Data Analysis*, *6*(4), 311–322. https://doi.org/10.3233/ida-2002-6402

Lawrence, T., Phillips, N., & Tracey, P. (2012). From the guest editors: Educating social entrepreneurs and social innovators. *Academy of Management Learning and Education*, *11*(3), 319–323. https://doi.org/10.5465/amle.2012.0224

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, J., Jang, D., & Park, S. (2017). Deep Learning-Based Corporate Performance Prediction Model Considering Technical Capability. *Sustainability*, *9*(6), 899. https://doi.org/10.3390/su9060899

Li, J., & Wang, Y. (2015). A new fast reduction technique based on binary nearest neighbor tree. *Neurocomputing*, *149*, 1647–1657. https://doi.org/10.1016/j.neucom.2014.08.028

Li, L., Zheng, Y., Sun, X. H., & Wang, F. S. (2014). The Application of Decision Tree Algorithm in the Employment Management System. *Applied Mechanics and Materials*, *543–547*, 1639–1642. https://doi.org/10.4028/www.scientific.net/AMM.543-547.1639

Liao, J., & Gartner, W. B. (2006). The effects of pre-venture plan timing and perceived environmental uncertainty on the persistence of emerging firms. *Small Business Economics*, *27*(1), 23–40. https://doi.org/10.1007/s11187-006-0020-0

Liao, J., & Gartner, W. B. (2007). The Influence of Pre-Venture Planning on New Venture Creation. *Journal of Small Business Strategy*, *18*(2), 1–21.

Liao, J., Welsch, H., & Moutray, C. (2009). Start-up resources and entrepreneurial discontinuance: The case of nascent entrepreneurs. *Journal of Small Business Strategy*, *19*(2), 89–103.

Lichtenstein, B. B., Carter, N. M., Dooley, K. J., & Gartner, W. B. (2007). Complexity dynamics of nascent entrepreneurship. *Journal of Business Venturing*, *22*(2), 236–261. https://doi.org/10.1016/j.jbusvent.2006.06.001

Lichtenstein, B. B., Dooley, K. J., & Lumpkin, G. T. (2006). Measuring emergence in the dynamics of new venture creation. *Journal of Business Venturing*, *21*(2), 153–175. https://doi.org/10.1016/j.jbusvent.2005.04.002

Loureiro, A. L. D., Miguéis, V. L., & da Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, *114*, 81–93. https://doi.org/10.1016/j.dss.2018.08.010

Lu, C. J., Lee, T. S., & Chiu, C. C. (2009). Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, *47*(2), 115–125. https://doi.org/10.1016/j.dss.2009.02.001

Lussier, R. N. (1995). A nonfinancial business success versus failure prediction model for young firms. *Journal of Small Business Management*, *33*(1), 8.

Lussier, R. N., & Halabi, C. E. (2010). Success versus Failure Prediction Model. *Journal of Small Business Management*, *48*(3), 360–377. https://doi.org/10.1111/j.1540-627X.2010.00298.x

Malmström, M., Johansson, J., & Wincent, J. (2015). Cognitive Constructions of Low-Profit and High-Profit Business Models: A Repertory Grid Study of Serial Entrepreneurs. *Entrepreneurship Theory and Practice*, *39*(5), 1083–1109. https://doi.org/10.1111/etap.12096

Manolova, T. S., Edelman, L. F., Brush, C. G., & Rotefoss, B. (2012). Properties of emerging organizations: Empirical evidence from Norway. *Small Business Economics*, *39*(3), 763–781. https://doi.org/10.1007/s11187-011-9360-5

Markovits, Y. (2012). The Development of Regulatory Foci Characters and Moderation Effects on Satisfaction and Commitment. *International Journal of Psychology and Behavioral Sciences*, *2*(4), 71–80. https://doi.org/10.5923/j.ijpbs.20120204.01

Marom, S., & Lussier, R. N. (2014). A Business Success Versus Failure Prediction Model for Small Businesses in Israel. *Business and Economic Research*, *4*(2), 63. https://doi.org/10.5296/ber.v4i2.5997

McCann, & Chris. (n.d.). *Scaling Airbnb with Brian Chesky — Class 18 Notes of Stanford University's CS183C*. https://medium.com/cs183c-blitzscaling-class-collection/scaling-airbnb-with-brian-chesky-class-18-notes-of-stanford-university-s-cs183c-3fcf75778358

McMullen, J. S., & Kier, A. S. (2016). Trapped by the entrepreneurial mindset: Opportunity seeking and escalation of commitment in the Mount Everest disaster. *Journal of Business Venturing*, *31*(6), 663–686. https://doi.org/10.1016/j.jbusvent.2016.09.003

Mehmed, K. (2011). *Data Mining Concepts, Models, Methods, and Algorithms*. John Wiley & Sons.

Meisenberg, G., & Woodley, M. A. (2013). Are cognitive differences between countries diminishing? Evidence from TIMSS and PISA. *Intelligence*, *41*(6), 808–816. https://doi.org/10.1016/j.intell.2013.03.009

Melville, H. (1850). *Hawthorne and His Mosses*. The Literary World.

Mitchell, R. K., Busenitz, L. W., Bird, B., Gaglio, C. M., McMullen, J. S., Morse, E. A., & Smith, J. B. (2007). The Central Question in Entrepreneurial Cognition Research 2007. *Entrepreneurship Theory and Practice*, *31*(1), 1–27. https://doi.org/10.1111/j.1540-6520.2007.00161.x

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, *62*, 22–31. https://doi.org/10.1016/j.dss.2014.03.001

Muller, E., & Peres, R. (2018). The effect of social networks structure on innovation performance: A review and directions for research. *International Journal of Research in Marketing*, *36*(1), 3–19. https://doi.org/10.1016/j.ijresmar.2018.05.003

Muñoz-Bullon, F., Sanchez-Bueno, M. J., & Vos-Saz, A. (2015). Startup team contributions and new firm creation: the role of founding team experience. *Entrepreneurship & Regional Development*, *27*(1–2), 80–105. https://doi.org/10.1080/08985626.2014.999719

Murnieks, C. Y., Mosakowski, E., & Cardon, M. S. (2014). Pathways of Passion: Identity Centrality, Passion, and Behavior Among Entrepreneurs. *Journal of Management*, *40*(6), 1583–1606. https://doi.org/10.1177/0149206311433855

Newbert, S. L. (2005). New Firm Formation: A Dynamic Capability Perspective. *Journal of Small Business Management*, *43*(1), 55–77. https://doi.org/10.1111/j.1540-627X.2004.00125.x

Newbert, S. L., & Tornikoski, E. T. (2012). Supporter networks and network growth: A contingency model of organizational emergence. *Small Business Economics*, *39*(1), 141–159. https://doi.org/10.1007/s11187-010-9300-9

Newbert, S. L., Tornikoski, E. T., & Quigley, N. R. (2013). Exploring the evolution of supporter networks in the creation of new organizations. *Journal of Business Venturing*, *28*(2), 281–298. https://doi.org/10.1016/j.jbusvent.2012.09.003

Nuñez, E. (2015a). The differing impact of household income on firm emergence by heterogeneous start-up configuration. *New England Journal of Entrepreneurship*, *18*(2), 31–46. https://doi.org/10.1108/neje-18-02-2015-b003

Nuñez, E. (2015b). The Temporal Dynamics of Firm Emergence by Team Composition. *The Journal of Applied Management and Entrepreneurship*, *20*(2), 65–93. https://doi.org/10.9774/GLEAF.1158.2015.00006

Omri, A., & Boujelbene, Y. (2018). Success Factors of Entrepreneurial Teams: Modeling Through the Cognitive Mapping. *Journal of Entrepreneurship and Management*, *7*(1), 1–9.

Packard, M. D., Clark, B. B., & Klein, P. G. (2017). Uncertainty Types and Transitions in the Entrepreneurial Process. *Organization Science*, *28*(5), 840–856. https://doi.org/10.1287/orsc.2017.1143

Pan, N. D., Gruber, M., & Binder, J. (2019). Painting with All the Colors: The Value of Social Identity Theory for Understanding Social Entrepreneurship. *Academy of Management Review*, *44*(1), 213–215. https://doi.org/10.5465/amr.2017.0504

Park, A. (2019, September 10). *Why We Don't Need Animals to Keep Enjoying Meat*. https://time.com/5601980/beyond-meat-ceo-ethan-brown-interview/

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Pollack, J. M., Forster, W. R., Johnson, P. D., Coy, A., & Molden, D. C. (2015). Promotion- and Prevention-Focused Networking and Its Consequences for Entrepreneurial Success. *Social Psychological and Personality Science*, *6*(1), 3–12. https://doi.org/10.1177/1948550614543030

Pompe, P. P. M., & Bilderbeek, J. (2005). The prediction of bankruptcy of small- and medium-sized industrial firms. *Journal of Business Venturing*, *20*(6), 847–868. https://doi.org/10.1016/j.jbusvent.2004.07.003

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors.* The Free Press.

Powell, E. E., & Baker, T. E. D. (2017). In the Beginning : Identity Processes and Organizing in Multi-Founder Nascent Ventures. *Academy of Management Journal*, *60*(6), 2381–2414. https://doi.org/10.5465/amj.2015.0175

Pu, S., Yan, Y., & Zhang, L. (2020). Peers, Study Effort, and Academic Performance in College Education: Evidence from Randomly Assigned Roommates in a Flipped Classroom. *Research in Higher Education*, *61*(2), 248–269. https://doi.org/10.1007/s11162-019-09571-x

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for Activation Functions. *6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings*, 1–13. http://arxiv.org/abs/1710.05941

Randolph-Seng, B., Mitchell, R. K., Vahidnia, H., Mitchell, J. R., Chen, S., & Statzer, J. (2015). The Microfoundations of Entrepreneurial Cognition Research: Toward an Integrative Approach. *Foundations and Trends® in Entrepreneurship*, *11*(4), 207–335. https://doi.org/10.1561/0300000055

Raniszewski, M. (2010). The edited nearest neighbor rule based on the reduced reference set and the consistency criterion. *Biocybernetics and Biomedical Engineering*, *30*(1), 31–40.

Rauch, A., & Frese, M. (2007). Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success. *European Journal of Work and Organizational Psychology*, *16*(4), 353–385. https://doi.org/10.1080/13594320701595438

Ravisankar, P., Ravi, V., Raghava Rao, G., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, *50*(2), 491–500. https://doi.org/10.1016/j.dss.2010.11.006

Reid, G. C., & Smith, J. A. (2000). What Makes a New Business Successful ?*. *Small Business Economics*, *14*(3), 165–182.

Renko, M. (2013). Early Challenges of Nascent Social Entrepreneurs. *Entrepreneurship Theory and Practice*, *37*(5), 1045–1069. https://doi.org/10.1111/j.1540-6520.2012.00522.x

Rest, J. (2013, May 21). *Samwers, die Klon-Krieger*. https://www.fr.de/wirtschaft/samwers-klon-krieger-11263791.html

Reynolds, P. D. (1987). New firms: Societal contribution versus survival potential. *Journal of Business Venturing*, *2*(3), 231–246. https://doi.org/10.1016/0883-9026(87)90011-5

Reynolds, P. D. (2016). Start-up Actions and Outcomes: What Entrepreneurs Do to Reach Profitability. *Foundations and Trends® in Entrepreneurship*, *12*(6), 443–559. https://doi.org/10.1561/0300000071

Reynolds, P. D. (2017). When is a Firm Born? Alternative Criteria and Consequences. *Business Economics*, *52*(1), 41–56. https://doi.org/10.1057/s11369-017-0022-8

Reynolds, P. D., Bosma, N., Autio, E., Hunt, S., Bono, N., Servais, I., Lopez-Garcia, P., & Chin, N. (2005). Global entrepreneurship monitor: Data collection design and implementation 1998-2003. *Small Business Economics*, *24*(3), 205–231. https://doi.org/10.1007/s11187-005-1980-1

Reynolds, P. D., & Curtin, R. (2007). *Panel Study of Entrepreneurial Dynamics Program Rationale and Description.* www.psed.isr.umich.edu/psed/background

Reynolds, P. D., & Curtin, R. T. (2009). Business creation in the United States: Entry, startup activities, and the launch of new ventures. In *The Small Business Economy*. United States Government Printing Office.

Reynolds, P. D., Hechavarria, D., Tian, L. R., Samuelsson, M., & Davidsson, P. (2016a). Panel study of entrepreneurial dynamics: A five cohort outcomes harmonized data set. *Res. Gate*, 1–48. https://doi.org/1.13140/RG.2.1.2561.7682

Reynolds, P. D., Hechavarria, D., Tian, L. R., Samuelsson, M., & Davidsson, P. (2016b). *Panel study of entrepreneurial dynamics: A five cohort outcomes harmonized dataset* [Data set]. http://www.psed.isr.umich.edu/psed/data

Reynolds, P. D., & Miller, B. (1992). New firm gestation: Conception, birth, and implications for research. *Journal of Business Venturing*, *7*(5), 405–417. https://doi.org/10.1016/0883-9026(92)90016-K

Reynolds, P. D., & White, S. B. (1997). *The entrepreneurial process: Economic growth, men, women and minorities.* CT: Quorum Books. https://doi.org/10.1080/01580370701403696

Ribeiro, B., & Lopes, N. (2011). Deep belief networks for financial prediction. In B.-L. Lu, L. Zhang, & J. Kwok (Eds.), *Neural Information Processing* (pp. 766–773). Springer.

Rosenbusch, N., Brinckmann, J., & Bausch, A. (2011). Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of Business Venturing*, *26*(4), 441–457. https://doi.org/10.1016/j.jbusvent.2009.12.002

Rotefoss, B., & Kolvereid, L. (2005). Aspiring, nascent and fledging entrepreneurs: An investigation of the business start-up process. *Entrepreneurship and Regional Development*, *17*(2), 109–127. https://doi.org/10.1080/08985620500074049

Sadat-Hashemi, S. M., Kazemnejad, A., Lucas, C., & Badie, K. (2005). Predicting the type of pregnancy using artificial neural networks and multinomial logistic regression: a comparison study. *Neural Computing and Applications*, *14*(3), 198–202. https://doi.org/10.1007/s00521-004-0454-8

Saemundsson, R., & Dahlstrand, Å. L. (2005). How Business Opportunities Constrain Young Technology-Based Firms from Growing into Medium-Sized Firms. *Small Business Economics*, *24*(2), 113–129. https://doi.org/10.1007/s11187-003-3803-6

Samuelsson, M., & Davidsson, P. (2009). Does venture opportunity variation matter? Investigating systematic process differences between innovative and imitative new ventures. *Small Business Economics*, *33*(2), 229–255. https://doi.org/10.1007/s11187-007-9093-7

Sassetti, S., Marzi, G., Cavaliere, V., & Ciappei, C. (2018). Entrepreneurial cognition and socially situated approach: a systematic and bibliometric analysis. *Scientometrics*, *116*(3), 1675–1718. https://doi.org/10.1007/s11192-018-2809-4

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent Variable Analysis_ Applications to Developmental Research.* (pp. 399–419). SAGE Publications Ltd.

SBA. (2012, September 2). *Frequently Asked Questions about Small Business*. https://www.sba.gov/sites/default/files/FAQ_Sept_2

Scherer, F. M., & Harhoff, D. (2000). Technology policy for a world of skew-distributed outcomes. *Research Policy*, *29*(4–5), 559–566. https://doi.org/10.1016/S0048-7333(99)00089-X

Schott, T., & Sedaghat, M. (2014). Innovation embedded in entrepreneurs' networks and national educational systems. *Small Business Economics*, *43*(2), 463–476. https://doi.org/10.1007/s11187-014-9546-8

Schumpeter, J. A. (1934). *The Theory of Economic Development.* Oxford University Press.

Sevim, C., Oztekin, A., Bali, O., Gumus, S., & Guresen, E. (2014). Developing an early warning system to predict currency crises. *European Journal of Operational Research*, *237*(3), 1095–1104. https://doi.org/10.1016/j.ejor.2014.02.047

Shah, J., Higgins, E. T., & Friedman, R. S. (1998). Performance Incentives and Means: How Regulatory Focus Influences Goal Attainment. *Journal of Personality and Social Psychology*, *74*(2), 285–293. https://doi.org/10.1037/0022-3514.74.2.285

Shane, S. (2012). Reflections on the 2010 "AMR" Decade Award : Delivering on the promise of entrepreneurship as a field of research. *The Academy of Management Review*, *37*(1), 10–20. https://doi.org/10.5465/amr.2011.0078

Shane, S., & Venkataraman, S. (2000). The Promise of Entrepreneurship as a Field of Research. *Academy of Management Review*, *25*(1), 217–226. https://doi.org/10.5465/amr.2000.2791611

Shepherd, D. A. (2015). Party On! A call for entrepreneurship research that is more interactive, activity based, cognitively hot, compassionate, and prosocial. *Journal of Business Venturing*, *30*(4), 489–507. https://doi.org/10.1016/j.jbusvent.2015.02.001

Shepherd, D. A., Douglas, E. J., & Shanley, M. (2000). New venture survival: Ignorance, external shocks, and risk reduction strategies. *Journal of Business Venturing*, *15*(5–6), 393–410. https://doi.org/10.1016/S0883-9026(98)00032-9

Shepherd, D. A., Williams, T. A., & Patzelt, H. (2015). Thinking About Entrepreneurial Decision Making: Review and Research Agenda. *Journal of Management*, *41*(1), 11–46. https://doi.org/10.1177/0149206314541153

Shim, J., & Davidsson, P. (2018). Shorter than we thought: The duration of venture creation processes. *Journal of Business Venturing Insights*, *9*, 10–16. https://doi.org/10.1016/j.jbvi.2017.12.003

Sieger, P., Gruber, M., Fauchart, E., & Zellweger, T. (2016). Measuring the social identity of entrepreneurs: Scale development and international validation. *Journal of Business Venturing*, *31*(5), 542–572. https://doi.org/10.1016/j.jbusvent.2016.07.001

Spanjol, J., Tam, L., Qualls, W. J., & Bohlmann, J. D. (2011). New product team decision making: Regulatory focus effects on number, type, and timing decisions. *Journal of Product Innovation Management*, *28*(5), 623–640. https://doi.org/10.1111/j.1540-5885.2011.00833.x

Spiegel, S., Grant-Pillow, H., & Higgins, E. T. (2004). How regulatory fit enhances motivational strength during goal pursuit. *European Journal of Social Psychology*, *34*(1), 39–54. https://doi.org/10.1002/ejsp.180

Steffens, P., Terjesen, S., & Davidsson, P. (2012). Birds of a feather get lost together: New venture team composition and performance. *Small Business Economics*, *39*(3), 727–743. https://doi.org/10.1007/s11187-011-9358-z

Stinchcombe, A. L. (1965). Social Structure and Organizations. In J. G. March (Ed.), *Handbook of Organizations.* (pp. 142–193). Rand McNally & Company.

Stryker, S., & Burke, P. J. (2000). The Past, Present, and Future of an Identity Theory. *Social Psychology Quarterly*, *63*(4), 284–297. https://doi.org/10.2307/2695840

Tajfel, H., & Turner, J. C. (1979). An Integrative Theory of Intergroup Conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33–47). Brooks/Cole.

Thiess, D., Sirén, C., & Grichnik, D. (2016). How does heterogeneity in experience influence the performance of nascent venture teams?: Insights from the US PSED II study. *Journal of Business Venturing Insights*, *5*, 55–62. https://doi.org/10.1016/j.jbvi.2016.04.001

Thorgren, S., & Wincent, J. (2015). Passion and habitual entrepreneurship. *International Small Business Journal: Researching Entrepreneurship*, *33*(2), 216–227. https://doi.org/10.1177/0266242613487085

Topuz, K., Zengul, F. D., Dag, A., Almehmi, A., & Yildirim, M. B. (2018). Predicting graft survival among kidney transplant recipients: A Bayesian decision support model. *Decision Support Systems*, *106*, 97–109. https://doi.org/10.1016/j.dss.2017.12.004

Tornikoski, E. T. (2008). Legitimating characteristics and firm emergence. *Journal of Enterprising Culture*, *16*(03), 233–256. https://doi.org/10.1142/S0218495808000144

Tornikoski, E. T. (2009). Legitimating behaviors and firm emergence: a resource dependence perspective. *International Entrepreneurship and Management Journal*, *5*(2), 121–138. https://doi.org/10.1007/s11365-007-0049-9

Tornikoski, E. T., & Newbert, S. L. (2007). Exploring the determinants of organizational emergence: A legitimacy perspective. *Journal of Business Venturing*, *22*(2), 311–335. https://doi.org/10.1016/j.jbusvent.2005.12.003

Tornikoski, E. T., & Puhakka, V. (2009). Exploring firm emergence: Initially conditioned or actively created? *International Journal of Entrepreneurship and Small Business*, *7*(1), 123–138. https://doi.org/10.1504/IJESB.2009.021613

Trevelyan, R. (2011). Self-regulation and effort in entrepreneurial tasks. *International Journal of Entrepreneurial Behavior & Research*, *17*(1), 39–63. https://doi.org/10.1108/13552551111107507

Tu, J., Lin, A., Chen, H., Li, Y., & Li, C. (2019). Predict the Entrepreneurial Intention of Fresh Graduate Students Based on an Adaptive Support Vector Machine Framework. *Mathematical Problems in Engineering*, *2019*, 1–16. https://doi.org/10.1155/2019/2039872

Tumasjan, A., & Braun, R. (2012). In the eye of the beholder: How regulatory focus and self-efficacy interact in influencing opportunity recognition. *Journal of Business Venturing*, *27*(6), 622–636. https://doi.org/10.1016/j.jbusvent.2011.08.001

Utterback, J. M. (1996). *Mastering the Dynamics of Innovation: How Companies Can Seize Opportunities in the Face of Technological Change.* Harvard Business School Press.

Van-Dijk, D., & Kluger, A. N. (2004). Feedback Sign Effect on Motivation: Is it Moderated by Regulatory Focus? *Applied Psychology*, *53*(1), 113–135. https://doi.org/10.1111/j.1464-0597.2004.00163.x

Van Dijk, D., & Kluger, A. N. (2011). Task type as a moderator of positive/negative feedback effects on motivation and performance: A regulatory focus perspective. *Journal of Organizational Behavior*, *32*(8), 1084–1105. https://doi.org/10.1002/job.725

van Gelderen, M., Thurik, R., & Bosma, N. (2005). Success and Risk Factors in the Pre-Startup Phase. *Small Business Economics*, *26*(4), 319–335. https://doi.org/10.1007/s11187-004-6837-5

Veganzones, D., & Séverin, E. (2018). An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, *112*, 111–124. https://doi.org/10.1016/j.dss.2018.06.011

Vegetti, F., & Adăscăliţei, D. (2017). The impact of the economic crisis on latent and early entrepreneurship in Europe. *International Entrepreneurship and Management Journal*, *13*(4), 1289–1314. https://doi.org/10.1007/s11365-017-0456-5

Vo, N. N. Y., He, X., Liu, S., & Xu, G. (2019). Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems*, *124*, 113097. https://doi.org/10.1016/j.dss.2019.113097

von Grone, J., Mueller, S., Grégoire, D. A., & Volery, T. (2016). Developing and validating a new instrument for measuring regulatory foci in entrepreneurship (summary). *Frontiers of Entrepreneurship Research, 36*(4), 3.

von Hippel, E. (1994). "Sticky Information" and the Locus of Problem Solving: Implications for Innovation. *Management Science*, *40*(4), 429–439. https://doi.org/10.1287/mnsc.40.4.429

Wagner, S., & Cockburn, I. (2010). Patents and the survival of Internet-related IPOs. *Research Policy*, *39*(2), 214–228. https://doi.org/10.1016/j.respol.2009.12.003

Wang, J. S., Lin, C. W., & Yang, Y. T. C. (2013). A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. *Neurocomputing*, *116*, 136–143. https://doi.org/10.1016/j.neucom.2011.10.047

WBAF. (2020). *Global Fundraising Stage - GFRS 2020 An international co-investment platform*. https://www.wbaforum.org/upload/07GFRS_2020_745.pdf

Westhead, P., Wright, M., & Ucbasaran, D. (2001). The internationalization of new and small firms: A resource-based view. *Journal of Business Venturing*, *16*(4), 333–358. https://doi.org/10.1016/S0883-9026(99)00063-4

Wiklund, J., Yu, W., & Patzelt, H. (2018). Impulsivity and Entrepreneurial Action. *The Academy of Management Perspectives*, *32*(3), 379–403. https://doi.org/10.5465/amp.2016.0177

Williams, D. (2017). Modeling business failure among SMEs: An artificial neural networks and logistic regression analysis. *Journal of Business and Entrepreneurship*, *28*(2), 1–27.

Wood, M. S., Williams, D. W., & Grégoire, D. A. (2012). The Road to Riches? A Model of the Cognitive Processes and Inflection Points Underpinning Entrepreneurial Action. In A. C. Corbett & J. A. Katz (Eds.), *Entrepreneurial Action* (pp. 207–252). Emerald Group Publishing Ltd. https://doi.org/10.1108/S1074-7540(2012)0000014010

Wry, T., & York, J. G. (2017). An Identity-Based Approach to Social Enterprise. *Academy of Management Review*, *42*(3), 437–460. https://doi.org/10.5465/amr.2013.0506

Wu, Y., Gaunt, C., & Gray, S. (2010). A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting and Economics*, *6*(1), 34–45. https://doi.org/10.1016/j.jcae.2010.04.002

Wu, Yue, Xu, Y., & Li, J. (2019). Feature construction for fraudulent credit card cash-out detection. *Decision Support Systems*, *127*, 113155. https://doi.org/10.1016/j.dss.2019.113155

Xu, L. (2014). A Tale of Trinity in Founder's Identity: The Case of New Venture Creation. *Journal of Management Policy & Practice*, *15*(5), 11–31.

York, J. G., O'Neil, I., & Sarasvathy, S. D. (2016). Exploring Environmental Entrepreneurship: Identity Coupling, Venture Goals, and Stakeholder Incentives. *Journal of Management Studies*, *53*(5), 695–737. https://doi.org/10.1111/joms.12198

Yusuf, J. E. (2012). A tale of two exits: Nascent entrepreneur learning activities and disengagement from start-up. *Small Business Economics*, *39*(3), 783–799. https://doi.org/10.1007/s11187-011-9361-4

Zekić-Sušac, M., Pfeifer, S., & Šarlija, N. (2015). A Comparison of Machine Learning Methods in a High-Dimensional Classification Problem. *Business Systems Research Journal*, *5*(3), 82–96. https://doi.org/10.2478/bsrj-2014-0021

Curriculum vitae

Name	Paris Koumbarakis, M.A. HSG
Date of Birth	21.10.1989
Place of Birth	Egg, Switzerland
Citizenship	Swiss
Marital Status	Single
Address	Zelgmatt 5, 8132 Egg, Switzerland
E-Mail	paris.koumbarakis@gmail.ch

Education

Since 2017	University of St.Gallen, Switzerland, Ph.D. in Management
2017 - 2020	University of St.Gallen, Switzerland, Business Education
2013 - 2016	University of St.Gallen, Switzerland, Master Studies in Accounting and Finance
2014 - 2016	ESADE Business School, Universitat Ramon Llull, Master Studies in Innovation and Entrepreneurship
2009 - 2013	University of St.Gallen, Switzerland, Bachelor Studies in Business Management
2011 - 2012	HEC Lausanne, Exchange

Work Experience

Since 2020	BMS Zurich, Teacher
Since 2018	FTS Consulting AG, Co-Founder
Since 2018	CV Pics (cvpics.ch/de), General Manager
2017 - 2018	Business Broker AG, Broker
2016	Biotronik AG, Inhouse Consultant
2013 - 2014	University of St. Gallen, Institute of Management & Strategy, Student Assistent
2012 - 2013	Mazars Coresa AG, Internship
2011	Groupon AG, Internship

Extracurricular Activities

2016 - 2017	Co-Founder HiRating (hirating.ch)
2014 - 2015	Co-Founder Kidventions, Regional Finalist of the Hult Prize 2015
2011 - 2013	Member of the Student Union, ESN/Ressort International

Languages

German (native); English (fluent), Greek (very good command); French (very good command); Spanish (basic); Russian (basic)

Hobbies

Guitar, Piano, Freeletics, Cooking, Traveling