What makes a responsible leader? - An empirical analysis of how personality characteristics affect leadership behaviour

Mischa Haberthür

Universität Zürich

Abstract

This master's thesis contributes to the ongoing discussion on responsible leadership and its antecedents by empirically analysing the relationship between various traits and responsible leadership effectiveness. In a quantitative experimental research study, the participants (62 men and 32 women) first filled out an online survey assessing their gender, age, and major field of study as well as their score in the personality dimensions honesty-humility, openness to experience, empathic concern, perspective taking, and holism. Afterwards, their second-to-fourth digit ratio was measured, and men had the option of providing a saliva sample which was used to determine their testosterone level. Finally, the participants' performance in the leadership roles professional, facilitator, citizen, idea-provider, and total was assessed with a think-aloud protocol. By performing a multiple linear regression analysis, I found that empathic concern positively influences the roles facilitator and total, holism the role citizen, and perspective taking the role idea-provider. I also found that men and psychology students are less likely to perform well in the role idea-provider and that openness to experience negatively affects facilitator.

Keywords: Responsible leadership, Ethical leadership, Corporate social responsibility, Stakeholder approach, Personality characteristics

1. Introduction

Over the course of the past few decades, political, economic, social, environmental, and technological activities have become increasingly interrelated on a worldwide scale (Figge and Martens (2014)). This ongoing process of globalisation has had a tremendous impact on the size and structural composition of companies as well. Nowadays, it is common for markets to be dominated by one or several multinational firms, which offer their products in a large variety of countries. Not only do these firms hold considerable market power, their influence frequently transcends a purely economic sphere. Many of the largest global corporations operate in countries with underdeveloped regulatory frameworks, granting them a lot of leeway in their regular business activities. Furthermore, their world-spanning reach impedes the nation states' capability of moderating their potentially harmful behaviour (Voegtlin et al. (2012)). This development in combination with the emergence of corporate scandals, such as the Enron case or the BP oil spill, has sparked a public debate on the role a market organisation should occupy in society and the responsibilities it has towards its environment.

The public discourse on corporate responsibility has been extended to incorporate the individual responsibility of leadership figures as well. Various researchers have argued that the traditional understanding of leadership effectiveness, which tends to put the focus on creating shareholder value and maximising profits, is inadequate for a globalised and interwoven business environment, among them Maak and Pless (2006), Patzer (2009), Waldman and Galvin (2008), and Waldman and Siegel (2008). Drawing on research from the fields of stakeholder theory and corporate social responsibility, these authors posit that previous leadership theories have neglected the importance of ethical behaviour and responsibility. They argue that the challenges presented by the globalised business world call for a new paradigm in leadership research, a concept that has been labelled responsible leadership.

While responsible leadership does not yet have a unifying definition that is accepted by all researchers in the field, it is generally considered to be related to fulfilling stakeholder needs and expectations. The underlying argument to this approach is that a leader's responsibility does not begin and end solely with the shareholders. Instead, he or she has an obli-

gation towards any person or group which is affected by the company's activities. To be truly responsible, a leader must therefore be willing to serve all stakeholders to the best of his or her abilities (Maak (2007)).

Based on these theoretical arguments, Voegtlin and Schwab (2015) have developed a multi-dimensional model of responsible leadership, in which leaders take on one of three different roles depending on which stakeholder group they engage. A responsible leader must be able to act as a professional towards shareholders or other owners of the company, as a facilitator towards colleagues and subordinates, and as a citizen towards external stakeholders. The better a leader is able to perform in all three roles simultaneously, the more he or she is considered to act responsibly.

While the link between responsibility and a focus on stakeholder demands is generally accepted in the literature, there is some dispute over what exactly motivates leaders to act in a responsible manner (Miska et al. (2014)). Some researchers suggest that responsible leadership is simply the result of calculated profit maximisation. Leaders engage stakeholders in order to gain a competitive advantage and increase the value of their business (Porter and Kramer (2006)). Other researchers posit that responsible leadership is at least partly informed by a leader's personal values or sense of duty (Miska et al. (2014); Waldman and Galvin (2008)).

The second approach would indicate that responsible leadership behaviour is not simply a rational choice that all leaders calculate in an equal manner, but is to some extent dependent on a leader's individual personality. This suggests that specific personality characteristics can increase or decrease the likelihood of a leader acting responsibly. I find some support for this assertion in the theoretical literature (Freeman and Auster (2011); Miska et al. (2014); Pless and Maak (2011)), but there is as of yet no clear consensus on which personality characteristics influence responsible leadership. The aim of my master's thesis is to help fill this gap in the current research by performing a quantitative empirical analysis to determine whether there are traits which have an impact on responsible leadership effectiveness.

To achieve this goal, I first established a set of ten traits which I argue could potentially influence responsible leadership behaviour. These are a person's gender, age, whether or not he or she is a psychology student, his or her testosterone level and second-to-fourth digit ratio (2D:4D) as well as his or her score in the personality dimensions honesty-humility, openness to experience, empathic concern, perspective taking, and holism. Additionally, I defined a set of outcome variables that can be used to measure responsible leadership effectiveness directly. For this, I relied on the multi-dimensional leadership model by Voegtlin and Schwab (2015) and its three roles professional, facilitator, and citizen as well as the two additional roles total and idea-provider. Using statistical regression analysis, I constructed a variety of multiple linear regression models that can be used to show whether any of the observed personality characteristics influence responsible leadership behaviour.

The data for these models was collected in collaboration with members of the Chair of Clinical Psychology and Psychotherapy and the Chair of Foundations of Business Administration and Theories of the Firm at the University of Zurich (UZH). We performed a scenario-based research study, in which 94 participants completed a survey on their individual personality traits and were then graded by us according to the responsible leadership effectiveness they displayed in a think-aloud protocol.

This master's thesis is divided into five chapters. In this chapter, the Introduction, I have given a brief overview of the thesis' subject, discussed my basic premise, and introduced the research goal I intend to achieve. Chapter 2 will focus on the theoretical foundations of my empirical regression study. In Subchapter 2.1, I will offer a general definition of leadership and discuss various research methods that have been used to measure it. I will then introduce the comparatively new field of responsible leadership, including the multi-dimensional model by Voegtlin and Schwab (2015), and highlight its differences from traditional approaches. Lastly, I will use these theoretical concepts to derive my research question. Subchapter 2.2 discusses the individual traits of my empirical regression analysis in greater detail. I will introduce each trait separately and form a hypothesis about its potential impact on responsible leadership effectiveness based on previous research.

Chapter 3 is concerned with the specific methods I used for collecting and analysing my data. I will talk about the research study conducted by members of the Chair of Clinical Psychology and Psychotherapy, the Chair of Foundations of Business Administration and Theories of the Firm, and myself, explaining our study design and the reasoning behind it. Subchapter 3.1 talks about the selection of participants and our recruitment process. Subchapter 3.2 is focused on the different variables we measured, while Subchapter 3.3 gives a brief overview of the actual procedures of the study. Finally, Subchapter 3.4 talks about the delimitations and limitations of our research study and their potential implications.

In Chapter 4, I will present the results of my empirical regression analysis. Each subchapter focuses on a different role of responsible leadership; Subchapter 4.1 talks about the models for the role of professional, Subchapter 4.2 about facilitator, Subchapter 4.3 about citizen, Subchapter 4.4 about idea-provider, and Subchapter 4.5 about total. Additionally, Subchapter 4.6 analyses the effects of testosterone levels on all five roles of responsible leadership.

Finally, Chapter 5 is going to feature the discussion and interpretation of my results from Chapter 4. I will once again look at each trait separately, compare the outcomes of my multiple linear regression models with the expected outcomes I have formulated in my original hypotheses, and determine whether there exist any congruencies or divergences. Subchapter 5.1 discusses the demographic and biological traits, while Subchapter 5.2 focuses on the personality characteristics. In Subchapter 5.3, I will analyse more general observations about my regression models that are not directly related to a single trait. Lastly, Subchapter 5.4

presents the conclusions of my thesis. I will give a brief summary of the most notable findings and highlight potential avenues that future research could focus on.

2. Theoretical foundations

This chapter discusses the theoretical frameworks which form the basis of my empirical research study. It is divided into two parts. In Subchapter 2.1, I will discuss the current state of the research on leadership and derive my generalised research question from it. In Subchapter 2.2, I will turn this research question into a specific set of hypotheses and introduce the traits I intend to measure.

2.1. Leadership

This subchapter discusses the current research in the field of leadership. In Part 2.1.1, I will give a brief overview of the traditional approaches to measuring leadership effectiveness. Part 2.1.2 will introduce the concept of responsible leadership, offer a comprehensive definition of the term, and discuss its implications. Part 2.1.3 concerns itself with the multi-dimensional leadership model and its applications to responsible leadership research. Finally, in Part 2.1.4, I will present my generalised research question and show how it was derived from combining the traditional approaches with the comparatively new field of responsible leadership.

2.1.1. Traditional approaches

The concept of leadership and its implications for management practices and organisational structures has always been one of the most extensively studied topics in the social sciences (Day and Antonakis (2012)). While the term leadership has different connotations depending on the context it is used in, researchers generally see it as a process of influencing other people in order to achieve a common goal (Yukl (2002)). A leader therefore is an individual capable of rallying followers and directing their efforts towards a specific task (Winston and Patterson (2006)).

The precise nature of such leaders and the question of what separates them from their followers has been the subject of extensive debates in the field. One of the earliest attempts to explain this phenomenon was the Great Man theory developed by the nineteenth century historian Thomas Carlyle, which posits that effective leadership is an inherent quality of specific individuals (Judge et al. (2002)). In this approach, leaders and followers are differentiated solely by certain heritable traits. Thus, great leaders are born rather than made (Hoffman et al. (2011)).

While the notion that only selected individuals are capable of becoming great leaders gradually fell out of favour over the course of the twentieth century, the idea that effective leadership results from specific traits has remained popular (Zaccaro (2007)). The trait perspective became the dominant approach in leadership research during the first half of the twentieth century, with researchers seeking to explore the connection between an individual's personality and his or her

effectiveness as a leader (Jago (1982)). Numerous studies attempted to quantify a set of traits that positively correlate with leadership effectiveness, operating under the assumption that identifying those traits could assist in selecting already qualified leaders or in fostering leadership behaviour within individuals who exhibit potential.

However, the empirical results from these studies often proved to be less significant than expected (Judge et al. (2002)). Many of the proposed traits did not show any correlations with leadership effectiveness, and for those that did, the effect often could not be replicated in subsequent studies. A seminal literature review by Stogdill (1948) found that while some traits indeed appear to impact leadership effectiveness, their importance is overstated when compared to more situational factors. The issue was further exacerbated by the fact that researchers sometimes used differing definitions for the same traits, which made comparing studies a difficult task. Additionally, there was no clear consensus on what exactly constitutes effective leadership and how it should be measured (Derue et al. (2011)).

The apparent problems of the purely trait-focused perspective caused a realignment in the field and led to the development of entirely new approaches to understanding leadership. Among them were behavioural theories, which claimed that effective leadership is the result of a leader's actions rather than his or her inherent qualities (Yukl (1989)). Proponents of this approach argued that there exist specific behaviours which generate effective leadership (House and Aditya (1997)). By analysing how leaders interact with their followers and replicating their successful strategies, leadership behaviour could therefore be learned by anyone.

As a reaction to both trait and behavioural perspectives on leadership, various contingency theories emerged in the second half of the twentieth century, for instance path-goal theory, situational leadership theory, and cognitive resource theory (House and Aditya (1997)). While these theories differed from each other in certain aspects, they all agreed that previous attempts at explaining effective leadership were too simplistic because they failed to take contextual factors into consideration (Judge et al. (2002)). Adherents of contingency theories argued that there are no universal behaviours or personality characteristics which always result in effective leadership. Instead, situational aspects, such as the social environment or a leader's degree of control, determine whether any given trait or action will contribute to the success of an organisation (Vroom and Jago (2007)).

Over the past few decades, trait theory has experienced a surge in popularity. By statistically analysing earlier studies, researchers found that the link between individual traits and leadership effectiveness was often more pronounced than had previously been assumed (Zaccaro et al. (2004)). In addition, the emergence of charismatic and transformative leadership once again highlighted the importance of individual characteristics on the leadership process (Zaccaro (2007)). More recently, several empirical studies have discovered positive correlations between specific personality attributes and leadership effectiveness, for instance Judge

et al. (2002), Judge et al. (2004), and McCormack and Mellor (2002).

What separates these new studies from traditional trait theory is that they no longer consider traits to be solely responsible for leadership effectiveness. Instead, they acknowledge the importance of situational effects and behavioural inputs, arguing that a comprehensive assessment of leadership needs to integrate all relevant factors (Derue et al. (2011)). The current research also puts a greater emphasis on harmonising the measurements used in empirical studies. Early research in the field was heavily fragmented, as studies often looked at completely different traits or analysed similar traits under different names (Judge et al. (2002)). More recent approaches try to mitigate this problem by focusing on measurements that are already well-established in psychological research, for instance the five-factor model (FFM) of personality (Judge et al. (2002)).

2.1.2. Responsible leadership

As has been shown in the previous part, leadership is generally considered to be a means of fulfilling a particular task or achieving a specific goal. However, the exact nature of this goal is not always clearly defined and strongly depends on the individual context. For instance, a military leader's objective might be to win a particular battle or ensure that casualties among his or her troops are minimised, whereas a civil rights leader would want to raise public awareness for his or her cause and implement social reforms. In both examples, effective leadership is important for reaching the desired goal, but a leadership approach which is successful in one situation may not necessarily produce the same results in the other.

Within the context of organisations competing in a market environment, a lot of empirical research has focused on leaders in the lower management, whose goals are predetermined by their superiors (House and Aditya (1997)). In those cases, leadership effectiveness has often been measured by looking at how well the managers accomplished the objectives set for them, while sometimes also evaluating the satisfaction of their employees. However, this approach is of limited use when analysing the leadership effectiveness of CEOs, who are not only able, but also expected to set the strategic goals for the entire organisation (Finkelstein and Boyd (1998)).

When evaluating the leadership effectiveness of a CEO, both researchers and the general public often tend to adopt a shareholder value approach (Waldman and Galvin (2008)). In this view, a CEO's sole responsibility is towards the shareholders of his or her company. As their direct employee, the CEO must ensure that the demands of the shareholders are met, a notion that is generally equated with maximising profits, stock prices, and future growth potential (Carson (1993)). Other considerations, such as employee or customer satisfaction, are only important in so far as that they contribute to maximising the shareholder value. Proponents of the shareholder value approach oftentimes argue that the focus on pure profit maximisation is not only beneficial to

the organisation itself, but also to society at large (Waldman and Galvin (2008)). Nevertheless, the theory has been heavily criticised for neglecting various groups who are of central importance to an organisation's continued operations (Russo and Perrini (2010)). Several researchers have pointed out that a pure shareholder-orientation does not adequately reflect actual business operations and that CEOs should instead try to incorporate other stakeholders in their decision-making process (Laplume et al. (2008)).

Criticism of the shareholder approach has not been solely confined to an academic setting. Corporate scandals, such as the Enron accounting fraud in 2001 or the BP oil spill in the Gulf of Mexico in 2010, have caused the public to take greater notice of corporate leaders, thereby sparking discussions about the specific role a business should occupy in society (Maak (2007)). The resulting public backlash led to a considerable loss of corporate legitimacy as well as a general distrust of leadership figures (Voegtlin et al. (2012)).

The notion of CEOs as pure profit-optimisers is further challenged by the ongoing globalisation process and the resulting lack of clearly-defined governance structures (Voegtlin et al. (2012)). Since businesses nowadays are often operating in areas with underdeveloped regulatory frameworks, the public has come to expect that they not only fulfil their basic obligations towards their shareholders, but also engage in behaviour which does not fall under the purview of standard market activities, such as combating corruption or poverty (Pless and Maak (2011)). Multinational firms thus transform from mere market participants into political actors (Voegtlin et al. (2012)).

With globalisation and scandals pushing the traditional understanding of corporate leadership to its limits, researchers have been trying to expand the term to incorporate these new challenges. Inspired by the academic discussions on stakeholder theory and corporate social responsibility, a new trend has emerged in the leadership literature in recent years; that of responsible leadership (Pless (2007)). As Waldman and Galvin (2008) note in their analysis of previous leadership theories:

> We also propose that the responsibility element is missing from these descriptors, and that it is actually this element that is at the heart of what effective leadership is all about. In a nutshell, to not be responsible is to not be effective as a leader. (p. 327)

Offering a comprehensive definition of responsible leadership is not an easy task, as different authors approach the term from their own unique perspectives (Miska et al. (2014)). In the following paragraphs, I will attempt to summarise the current state of the field by introducing two specific definitions, both of which either have been directly cited or whose meaning is reflected in multiple other research papers. I will then compare and contrast the two definitions to gain a greater understanding of what the responsible leadership concept actually entails. The first definition comes from an article by Maak and Pless (2006):

Responsible leadership is a relational and ethical phenomenon, which occurs in social processes of interaction with those who affect or are affected by leadership and have a stake in the purpose and vision of the leadership relationship. (p. 103)

The second definition is from an article by Waldman (2011):

[Responsible leadership] considers what managers should do in an effort to take into account the needs of stakeholders, other than shareholders, who may have legitimate interests in a firm's activities. (p. 77)

What these definitions have in common is that they embed leadership in the context of stakeholder interactions. The authors argue that a corporate leader has a responsibility not only to the company's shareholders, but to every person or group that is affected by the company's business activities. A responsible leader is therefore one who is able to fulfil stakeholder needs and expectations (Maak (2007)).

Both definitions also agree that responsible leadership is at least to some extent a question of morality. Maak and Pless (2006) explicitly refer to it as an ethical phenomenon, while Waldman (2011) draws comparisons to the related field of ethical leadership in his own article. This implies that responsible leadership transcends a purely self-serving perspective on business activities. A leader should not establish good relationships with stakeholders merely so they will grant legitimacy; rather, legitimacy is the natural result of a leader being a responsible member of the community. What separates responsible leadership from the concept of ethical leadership is that the former does not focus on a leader's own ethical norms and values. Instead, an ethical solution to a problem is created through the interaction between leaders and internal as well as external stakeholders (Waldman and Galvin (2008)).

One minor aspect in which the two presented definitions somewhat differ from each other is their conceptualisation of the actual leadership process. Maak and Pless (2006) explicitly stress the importance of direct interactions with stakeholders as a means of becoming a responsible leader. Their view is informed by the concept of deliberative democracy, which states that corporate legitimacy is gained through approaching stakeholders as equals and engaging them in a constructive dialogue (Palazzo and Scherer (2006)). In the ideal-type scenario, finding an appropriate solution to an issue would be done through a democratic process in which every affected person or entity has a voice. Waldman (2011), on the other hand, only states that a manager should consider the needs of stakeholders, but leaves open how strongly said stakeholders should be integrated into the final decision-making process. Under his definition, a leader could theoretically decide in an authoritative manner and still be considered responsible, as long as he gave due consideration to the stakeholders' demands in his deliberations.

2.1.3. Multi-dimensional leadership model

Despite their semantic differences, the definitions of Maak and Pless (2006) and Waldman (2011) seem to roughly agree on what responsible leadership is and how it can be considered distinct from other leadership concepts. Their basic assumption that the increasingly complex business environment necessitates a reconceptualisation of corporate leadership with a greater emphasis on stakeholder demands is reflected in many other writings on the subject, for instance Cameron (2011), Patzer (2009), Voegtlin (2011), and Waldman and Siegel (2008).

However, there is as of yet no clear consensus on how these theoretical concepts can be optimally implemented into practice, especially considering that stakeholders often have different and seemingly conflicting demands (Voegtlin and Schwab (2015)). A manager who attempts to incorporate all affected parties in the decision-making process may soon realise that he or she is faced with significant trade-offs. For instance, consumers usually want to pay as little as possible for a given product, but this desire may conflict with the employees demanding fair wages for their labour. Alternatively, the public's wish for stronger environmental protection can be difficult to reconcile with the shareholders' demands for continually increasing profits and growth. In addition to these inherent contradictions, managers are often faced with resource constraints as well. Temporal and financial limitations may make it difficult to give every stakeholder group the attention it requires or deserves.

Some researchers have proposed that these issues can be resolved by prioritising stakeholders based on their salience (Bundy et al. (2013)). A famous example of this approach is the framework developed by Mitchell et al. (1997), which measures stakeholder salience on the basis of three criteria: power, legitimacy, and urgency. They argue that stakeholders who possess all three attributes should be given high priority, while those with only one attribute are of lesser importance. This approach allows leaders to resolve conflicting stakeholder demands by focusing on whichever group matters more to the firm's operations. However, stakeholder salience is not always easy to determine and constantly fluctuating, which can make it difficult to objectively assess. Additionally, salience prioritisation does not seem to conform to the normative literature on responsible leadership, which tends to argue that all stakeholder demands should be recognized, regardless of their importance (Maak and Pless (2006); Waldman and Galvin (2008)).

In an attempt to approach the issue from a new perspective, Voegtlin and Schwab (2015) argue that the apparent gap between theoretical literature and practical considerations is not necessarily insurmountable. They propose that leaders might be able to deal with complex and paradoxical situations by becoming more complex themselves. Through the use of innovative approaches, a leader could reconcile contradictory stakeholder demands, thereby turning them from a seemingly unsolvable problem into a win-win situation for all involved parties. Thus, responsible leadership re-

quires a new paradigm of management which places a strong emphasis on behavioural complexity (Voegtlin and Schwab (2015)).

Building on these arguments and drawing inspiration from psychological contracts theory as well as stewardship theory, Voegtlin and Schwab (2015) developed a multi-dimensional model of leadership. They suggest that a leader has different responsibilities towards different groups. As such, managers in a globalised, highly complex business environment cannot afford to be inflexible; instead, they need to be able to take on different roles depending on which stakeholders they are addressing. With regards to shareholders or other owners of the company, a responsible leader must be able to act as a professional, fulfilling his or her fiduciary duty and ensuring that tasks are performed effectively and efficiently. Towards colleagues and subordinates, a responsible leader takes on the role of facilitator, caring for the other members of the workplace and creating a fair work environment. Finally, a responsible leader must act as a citizen towards external stakeholders, taking sustainability into consideration and analysing the firm's long-term effects on society (Voegtlin and Schwab (2015)). These three roles are summarised in Table 1 below.

Voegtlin and Schwab (2015) infer a causal relationship between their model and effective responsible leadership. They argue that the main challenge managers face when attempting to become responsible leaders is finding a balance between the roles of professional, facilitator, and citizen, thereby serving all stakeholder groups to the best of their abilities. The better a leader is able to perform multiple roles simultaneously, by observing high behavioural complexity and finding innovative solutions that satisfy all stakeholders, the more he or she is considered to act responsibly. Thus, their model does not only serve as a potential guideline for managers who wish to incorporate responsible leadership considerations into their business activities, it also offers a new perspective on how to empirically assess responsible leadership effectiveness.

2.1.4. Research question

In Part 2.1.1, I introduced the trait theory of leadership with its basic assumption that specific traits of individuals serve as antecedents to effective leadership. Empirical research in the field has focused on a large variety of personality characteristics that could potentially influence leadership behaviour, such as self-confidence, aggressiveness, dominance, and integrity (Judge et al. (2002)). So far, trait theory has predominantly worked with a traditional understanding of leadership, in which effectiveness is measured by how well the manager in question accomplished his or her given tasks or through surveying employee satisfaction (House and Aditya (1997)). However, proponents of responsible leadership posit that this understanding is too narrow for a globalised, highly interwoven business environment, arguing that it should be expanded to include aspects which are not reflected in a purely strategic approach. It seems reasonable to assume that this reconceptualisation would also result in a change of the antecedents. In other words, the traits which help one become an effective leader under a shareholder value paradigm are not necessarily the same traits that aid in becoming an effective responsible leader.

The question of what makes a responsible leader has been the subject of extensive debates in the theoretical literature. Various personality traits have been proposed as potential antecedents for effective responsible leadership, including relational and ethical intelligence (Maak and Pless (2006)), certain moral predispositions (Voegtlin et al. (2012)), authenticity (Freeman and Auster (2011)), virtuousness (Cameron (2011)), and cognitive complexity (Maak et al. (2016)). However, there is as of yet little empirical research into how an individual's personality affects responsible leadership effectiveness (Fernando (2016)). This may be partly due to the complexity in measurement and data collection, as traditional follower surveys are not necessarily comprehensive enough to assess the full spectrum of responsible leadership activities (Waldman (2011)).

The goal of my master's thesis is to help fill this gap in the current empirical research and contribute to the ongoing academic discussion through performing a quantitative empirical analysis. By drawing on the well-established trait theory and applying its methods to the comparatively new field of responsible leadership, I aim to answer the following research question:

> Research question: Which personality characteristics increase the likelihood that a leader will act responsibly?

As the field of responsible leadership is still in its early, formative years (Fernando (2016)), finding an answer to this question could provide valuable information on which directions are worth pursuing in future research; either by corroborating already established theories or by highlighting new avenues that hitherto have not been considered in the literature. Furthermore, exploring the link between personality characteristics and responsible leadership behaviour may offer practical advice for firms and their boards of directors. Finding the right person for a management position has always been a challenging task (Beck and Harter (2014)), and the increasing complexity of the globalised business environment compounds this issue even further. The consequences of a bad selection can be disastrous, as evidenced by corporate scandals, such as the Enron case, and the resulting loss of corporate legitimacy (Voegtlin et al. (2012)). Finding a relationship between certain traits and leadership effectiveness could help mitigate this problem, as it would give firms a better understanding of which qualities they should be looking for in an aspiring manager or promoting in their own organisational culture.

To answer my research question, I will proceed in two steps. Firstly, I define a group of specific traits that can be expected to have an influence on responsible leadership behaviour. My choice of traits and the justification for their selection will be discussed in-depth in the following subchapter. Secondly, I establish a set of criteria that allow

Table 1: Responsible leadership roles (Voegtlin and Schwab (2015))

Responsible leadership role	Moral obligation (that develops out of:)	Accountability toward stakeholders	Goal	Definition
Professional	Fiduciary duty task orientation	shareholders, investors, supervisor, customers	achieve performance goals	Leader perceives a moral obligation toward shareholders/owners, born out of fiduciary duty; displays instrumental behavior targeted at fulfilling organizational performance goals
Facilitator	Respect, care relational orientation	subordinates, colleagues	create fair working environment	Leader perceives moral obligations towards colleagues and subordinates; displays relational behavior in that she/he shows care and concern for others in the work environment
Citizen	Public welfare social welfare orientation	external stakeholders (NGOs, community, state, family, non-contractual employees, etc.)	create (long-term) value for society	Leader perceives moral obligations towards society and future generations, born out of social-welfare orientation; displays citizenship behaviors in that she/he considers social and environmental impact and emphasis long-term goals

me to directly measure responsible leadership effectiveness. I base these criteria on the multi-dimensional leadership model from Voegtlin and Schwab (2015) that was discussed in the previous part. An individual's effectiveness as a responsible leader is therefore measured by how well he or she performs in the three roles professional, facilitator, and citizen or in other words, how well he or she considers the needs of the company, the employees, and the external environment in the decision-making process. There are also two additional criteria which are not explicitly named in the multi-dimensional model, but are implicit in its fundamental assumptions. One is how well individuals fulfil all three roles simultaneously (total), the other is how well they perform as idea-provider; that is to say, how much they bring in their own ideas and creative solutions. The methods used for measuring these five criteria will be examined in Subchapter 3.2.

2.2. Characteristics

This subchapter focuses on the individual characteristics that I intend to measure within the confines of my empirical research study. I will introduce each characteristic separately, justify its inclusion into the study, and formulate a hypothesis about the way it will presumably affect responsible leadership effectiveness. Each part of this subchapter focuses on a different category of characteristics. Part 2.2.Demographic characteristics comprise all manner of socioeconomic factors in a given population, for instance income, religion, or birth rate. For the purpose of my empirical analysis, I measure three specific demographic characteristics of the study's par-

ticipants: their gender, age, and whether or not their university major is in psychology. For biological characteristics, I am interested in their right-hand second-to-fourth digit ratio and its correlation with testosterone levels.1 will talk about the demographic factors as well as the sole biological factor, while Part 2.2.2 will discuss the personality factors.

2.2.1. Demographic and biological characteristics

Demographic characteristics comprise all manner of socioeconomic factors in a given population, for instance income, religion, or birth rate. For the purpose of my empirical analysis, I measure three specific demographic characteristics of the study's participants: their gender, age, and whether or not their university major is in psychology. For biological characteristics, I am interested in their right-hand second-to-fourth digit ratio and its correlation with testosterone levels.

Gender

The relationship between gender and effectiveness as a leader has been extensively studied in the existent literature, although the results are somewhat inconclusive (Eagly and Johnson (1990)). As Barbuto et al. (2007) note: "For every study that has shown differences in leadership behaviors based on gender ..., another has shown no differences at all" (p. 71). Researchers who observe gender differences in their studies often argue that these can be explained by social stereotypes (Ridgeway (2001)). Men are expected to be task-oriented; as such, their leadership style tends to be autocratic with a strong emphasis on hierarchical authority, competitiveness, and achievements (Eagly and Johnson (1990)). Meanwhile, women are expected to display a more interper-

sonal approach. Their leadership style is generally more participatory, focusing on building up morale within a team and maintaining positive relationships (Barbuto et al. (2007)). In a similar vein, some studies have found that women are more likely to score well on emotional intelligence than their male peers (Mandell and Pherwani (2003)). However, these gender differences frequently disappear when looking at a formal organisational setting (Barbuto et al. (2007)), which indicates that women in management positions tend to adopt the leadership style of men. This may be because traditional notions of leadership have been dominated by a masculine view (Carli and Eagly (2001)), pressuring women into changing their approach if they want to be seen as equals.

Since responsible leadership theories tend to emphasise the limitations of a purely task-focused approach while simultaneously stressing the importance of building relationships and cooperating with stakeholders, it seems possible that stereotypically female traits would be more valuable than they are often considered to be in a traditional understanding of leadership. Women might be more likely to build up strong relationships with employees as well as external parties, thus performing better in the roles of facilitator and citizen. However, their performance as professional is harder to assess. Shareholder obligations often require a strong focus on performing tasks as efficiently as possible, which could favour a more stereotypically masculine leadership. The literature also gives no clear indication on whether gender has an influence on the role of idea-provider. While there is a wealth of research on the relationship between gender and creativity, findings in the field have proven inconclusive (Baer and Kaufman (2008)). To summarise, while the literature offers some indication that women could potentially perform better in the roles of facilitator and citizen and men in the role of professional, there is not enough evidence to make strong assumptions about how gender differences affect the ability to perform all three roles simultaneously or to act as idea-provider.

> Hypothesis 1: Women are more likely than men to perform well in the roles of facilitator and citizen, while men are more likely to perform well in the role of professional.

Age

The influence of age on leadership behaviour has not been as thoroughly explored as that of gender (Barbuto et al. (2007)). However, recent demographic changes in firms, such as the increasingly aging workforce widening the gap between the oldest and youngest employees and the greater prevalence of young managers in high positions, have spurred interest in the topic (Oshagbemi (2004)). In a study focusing on the Australian Public Service and National Health Service trust organisations in the United Kingdom, Korac-Kakabadse et al. (1998) found that time-related dimensions such as a manager's number of years in the organisation, number of years in the job, and age have a noticeable effect on his or her leadership behaviour. Young managers are more likely to be radicals who dislike control, but enjoy

challenges. They are primarily characterised by their flexibility and competitiveness as well as their independence. Meanwhile, older managers tend to be either bureaucrats who focus on rules and maintaining order or team players who adopt a more team-driven approach and encourage constructive dialogue (Korac-Kakabadse et al. (1998)). Similar findings were observed by other researchers in the field (Oshagbemi (2004)). Younger leaders generally appear to feel more comfortable in rapidly changing environments and operate with more energy, but they also tend to be more competitive and work towards promoting themselves first and foremost. Older leaders, on the other hand, try to minimise risks by relying on past experience and in-depth knowledge of their respective fields. They tend to cooperate more with their followers and primarily work towards promoting others (Oshagbemi (2004)).

Translating these findings into responsible leadership effectiveness is difficult, as both young and old leaders seem to possess beneficial qualities. The increased flexibility and drive of the younger leaders may allow them to more swiftly adapt to the increasingly complex business environment. However, their strong self-focus and competitiveness would seem to be at odds with an approach that strongly emphasises stakeholder concerns, whereas the team- and communications-oriented style of older leaders would be more suited for this task. Additionally, the increased knowledge and experience of the older leaders may potentially allow them to resolve complex and seemingly paradoxical situations better than their younger peers. As such, I would argue that age is likely to positively influence all aspects of responsible leadership effectiveness.

> Hypothesis 2: Age has a positive influence on all three roles of responsible leadership (professional, facilitator, and citizen) as well as the role of idea-provider.

Psychology student

While the question of whether or not leadership can be taught is the subject of extensive debates in the field (Elmuti et al. (2005)), research that looks at the specific impact a person's education has on his or her leadership behaviour is still relatively rare (Barbuto et al. (2007)). Only a small number of empirical studies have attempted to map the causal relationship between educational level and leadership effectiveness. Similarly, there is little research into the relationship between a student's university major and his or her leadership style, despite there being evidence that students exhibit personality differences depending on their chosen field of study (Baluch et al. (1996); Vedel (2014)).

The participants of our research study were almost exclusively students from the UZH or the Swiss Federal Institute of Technology in Zurich (ETHZ) and thus all on roughly the same educational level. However, there were noticeable differences in their chosen field of study. While we had students from a wide variety of different fields, more than half of the participants stated that they were majoring in psychology. I therefore decided to create a variable for my regression

models which measures whether the responsible leadership effectiveness of psychology students differs from that of their peers.

Several empirical studies have analysed the differences in personality characteristics between students enrolled in various college or university majors. A systematic literature review by Vedel (2014) found that psychology students tend to score high on the FFM dimensions neuroticism, openness to experience, and agreeableness. Liberal arts students on average also exhibit higher creativity than those majoring in business or natural sciences (Pringle et al. (2010)). In addition, Dimitrijević et al. (2011) found that psychology majors show more empathy and a greater motivation to help others than students who have chosen non-helping professions. These results indicate that psychology students might be better at performing the roles of facilitator and citizen, as they have a greater desire to serve stakeholder needs and ensure that everyone is satisfied. Higher creativity could also mean that they perform well in the role of idea-provider.

> Hypothesis 3: Students who major in psychology are more likely to perform well in the roles of facilitator, citizen, and idea-provider than students from other fields.

Originally, I had planned to include an additional variable for students who are business majors to determine whether they would differ from the other participants as well. However, only sixteen of the participants in our study actually majored in an economic field, including those who were enrolled in economics or economic chemistry. As such, I decided to discard the variable due to the low sample size.

Second-to-fourth digit ratio and testosterone

The effect of testosterone on human interactions has been extensively studied in a wide variety of medicinal and psychological research papers (Hines (2006)). Historically, the hormone has been closely linked to a variety of negative attributes, including aggressiveness, anti-social behaviour, and criminal activities. However, more recent evidence indicates that this relationship is not as clear-cut as it is commonly assumed to be (Dabbs Jr and Morris (1990)). While many studies did indeed find a correlation between aggressiveness and testosterone, others found no correlation at all (Van Bokhoven et al. (2006)). This has led some researchers to suggest that testosterone in humans is primarily linked to a desire for social dominance (Mazur and Booth (1998)). Higher levels of testosterone therefore indicate greater power motivation and increased vigilance against potential threats to one's social status (Eisenegger et al. (2011)). This inclination towards dominance can sometimes lead to aggressive or violent behaviour, but it can also manifest itself in alternate ways. For instance, Eisenegger et al. (2010) found that higher levels of testosterone increased the fairness of participants in the ultimatum bargaining game because those with high testosterone do not want their offer to be rejected by their partners. The greater degree of fairness is hereby not caused by altruism, but rather by the person's concern for his or her social status. In addition, while women on average have far lower levels of testosterone than men, an increase in testosterone seems to have a similar effect on both genders (Archer (2006); Stanton et al. (2011)).

Based on these findings, it seems reasonable to assume that testosterone levels would have an impact on responsible leadership effectiveness as well. However, the exact nature of this relationship is difficult to discern. The study conducted by Eisenegger et al. (2010) suggests that leaders with a high testosterone level might be more willing to fulfil stakeholder needs because finding a solution that is agreeable to everyone would increase their social status and assert their dominance. At the same time, such leaders would only engage in this behaviour if it were of personal benefit to them. This could indicate that they would focus solely on powerful stakeholders who have something to offer them, such as shareholders, while neglecting stakeholders whose claims may not be any less legitimate, but who cannot reward or punish the manager. Furthermore, high-testosterone leaders might also be more likely to fall back on more aggressive and competitive tactics if the cooperative strategy does not produce the desired results. I therefore argue that higher levels of testosterone are negatively correlated with effectiveness in the roles of facilitator and citizen. However, such leaders might show a better performance in the role of professional due to their increased competitiveness and desire to improve their social standing, which in the case of a CEO is often tied to the firm's earnings.

In our research study, we measured testosterone levels directly through the use of saliva samples. However, due to financial constraints, we could only analyse samples from male participants. Additionally, providing a saliva sample was not a necessary step for partaking in the study. As a result, we only received samples from roughly half of all participants. To measure the testosterone levels of the other participants, I rely on the 2D:4D, which is the relative length between the second and fourth finger of a hand. Studies have shown that this ratio is strongly correlated with a person's prenatal testosterone levels (Neave et al. (2003)). A low 2D:4D has been linked to higher levels of prenatal testosterone and correspondingly to lower risk aversion (Coates et al. (2009)) and increased aggressiveness (Perciavalle et al. (2013)), both of which are qualities that have also been linked to current testosterone levels (Sapienza et al. (2009); Van Bokhoven et al. (2006)). In my main regression models, I include the 2D:4D to see if it influences responsible leadership effectiveness. I will also analyse the influence of the testosterone values from the saliva samples, although that will be done in a separate regression to avoid the danger of overfitting. Additionally, I will look at the correlation between the two variables to see whether prenatal testosterone values correspond to current ones. While we measured the 2D:4D for both hands in our study, I will be using the ratio from the right hand in my analysis because it tends display more robust sex differences, indicating that it is more sensitive to differences in prenatal androgens (Coates et al. (2009)).

Hypothesis 4: A person with a low 2D:4D is more likely to perform well in the role of professional, but less likely to perform well in the roles of facilitator and citizen than one with a high 2D:4D.

Hypothesis 5: A person with high levels of testosterone is more likely to perform well in the role of professional, but less likely to perform well in the roles of facilitator and citizen than one with low levels of testosterone.

2.2.2. Personality characteristics

As has been discussed in Part 2.1.1, traditional trait theory has analysed a large variety of personality characteristics to determine whether they are correlated with effective leadership behaviour. However, researchers often worked with their own specific understanding of a given trait, making direct comparisons between studies difficult or even impossible (Judge et al. (2002)). To mitigate this issue, I based my selection of personality characteristics on validated scales which are already well-established in psychological research. My regression models include two dimensions from the HEXACO model of personality structure: honesty-humility and openness to experience. In addition, I also measure empathic concern and perspective taking from the Interpersonal Reactivity Index (IRI) as well as holism from the Analysis-Holism Scale (AHS).

Honesty-humility

One of the five dimensions of the FFM is agreeableness, which measures an individual's concern for cooperation and compliance with social norms (Jensen-Campbell and Graziano (2001)). As such, it puts a much stronger focus on interpersonal relationships than the other FFM characteristics (Graziano et al. (1996)). People who score high on agreeableness value social harmony and desire to get along well with everyone, often showing concern for the well-being of others. In social interactions, they are generally perceived to be warm, friendly, and helpful (Graziano and Eisenberg (1997)). Conversely, people who score low on agreeableness tend to be more suspicious of others and show less concern for their well-being. They are therefore also less likely to go out of their way in order to help someone and place a greater emphasis on competition rather than cooperation. Agreeableness has also been shown to correlate with empathy (Graziano et al. (2007)), indicating that those with high agreeableness display a greater degree of prosocial motivation and a higher desire to help rather than hurt.

For our research study, we used a modified version of the FFM called the HEXACO model of personality structure, which was developed by Michael Ashton and Kibeom Lee (Ashton et al. (2004)). Like the FFM, the HEXACO model incorporates the three dimensions extraversion, conscientiousness, and openness to experience, leaving them mostly unchanged. However, it replaces neuroticism and agreeableness with three new dimensions: agreeableness, emotionality, and honesty-humility. This restructuring was a reaction to a number of lexical studies which consistently found six factors of personality rather than five, indicating that the traditional FFM approach is somewhat limited (Ashton and Lee (2007)).

In the HEXACO model, both agreeableness and honesty-humility possess elements that are associated with the traditional understanding of agreeableness and can be seen as complementary aspects (Ashton and Lee (2007)). They are both used to measure reciprocal altruism, albeit from different vantage points. Honesty-humility is an expression of fairness. Individuals who score high on this dimension tend to cooperate with others even if they have something to gain by exploiting them. Agreeableness meanwhile measures tolerance and forgiveness. A person who exhibits high agreeableness is willing to cooperate with others even if he or she has been or is currently being exploited by them (Ashton and Lee (2007)).

It seems reasonable to suggest that these two dimensions have the potential of affecting responsible leadership effectiveness. A manager who scores high on honesty-humility may be more likely to cooperate with stakeholders and consider their demands even if there is no immediate financial benefit to be gained from such behaviour. Likewise, a manager who scores high on agreeableness might be willing to persevere and keep up negotiations despite potential setbacks. The general desire for social harmony and greater willingness to help others that is inherent to these two dimensions could also be seen as positive factors when dealing with employees and external stakeholders.

Originally, I had intended to include both agreeableness and honesty-humility as separate predictors in my multiple regression models. However, I did not want to incorporate too many independent variables due to my relatively low sample size and the associated danger of overfitting. As such, I decided to discard agreeableness and instead focus solely on honesty-humility. I argue that this choice is justified because CEOs of multinational companies are frequently in a position where they can exploit their stakeholders, while stakeholders themselves often have little to no means of exploiting the company. Therefore, honesty-humility seems to be more pertinent to a manager's business activities than agreeableness and accordingly a better predictor for responsible leadership effectiveness. In conclusion, I posit that displaying a high degree of honesty-humility will positively influence a manager's ability to perform well in the roles of facilitator and citizen.

Hypothesis 6: A person who scores high on honesty-humility is more likely to perform well in the roles of facilitator and citizen than one who scores low.

Openness to experience

Like agreeableness, openness to experience is one of the personality characteristics of the FFM. It can be broadly seen as a reflection of curiosity and open-mindedness, measuring the number of interests a person has and the degree to which he or she pursues them (Howard and Howard (1995)). Those who score high on openness to experience tend to welcome and actively seek out new aspects of life. They possess a high

level of emotional sensitivity and show an active imagination as well as strong aesthetic interests (Coan (1972)). Conversely, people who score low on openness to experience are often sceptical or hostile to new practices, preferring to rely on tradition and conventional approaches instead.

Empirical research studies have found a correlation between openness to experience and general creativity (Baer and Oldham (2006); McCrae (1987)) as well as adaptability (LePine et al. (2000)). High openness to experience seems to lead to greater flexibility, making it easier to adapt to rapidly changing conditions. Likewise, people who score high on openness to experience appear to be more likely to think outside of the proverbial box and come up with new and unique solutions to existing problems. George and Zhou (2001) theorise that this is due to said people having a broader range and depth of experiences to fall back on when attempting to solve an issue as well as being more open to the idea of trying something different. Thus, they have both the means and the inclination to discover radically new approaches. In a similar vein, openness to experience has been found to have a moderate influence on general intelligence (Harris (2004)).

The HEXACO model of personality structure we use in our research study defines openness to experience in the same way as the traditional FFM, although it slightly alters the respective subcategories (Ashton and Lee (2007)). In the HEXACO model, openness to experience is comprised of four dimensions: aesthetic appreciation, inquisitiveness, creativity, and unconventionality (Lee and Ashton (n.d.)). Aesthetic appreciation assesses the degree to which an individual derives enjoyment from the beauty inherent in art and nature, inquisitiveness measures the willingness to seek out new information and experiences, creativity showcases the preference for innovation and experimentation, and unconventionality indicates how accepting an individual is of people or ideas that defy societal expectations (Lee and Ashton (n.d.)).

Leaders in a modern, globalised business world are constantly faced with rapidly changing conditions and highly complex, seemingly paradoxical situations. Hence, openness to experience might be a valuable trait for those aspiring to be responsible leaders. Managers who score high on this personality dimension may be more capable of resolving conflicting stakeholder demands by drawing upon a greater wealth of knowledge and experience as well as by attempting radically new solutions that more traditionally minded managers would shy away from. Their increased flexibility may also allow them to more swiftly resolve issues or even recognise and prevent potential threats before they can become an acute problem. I therefore argue that a high degree of openness to experience positively influences a leader's ability to perform well in the role of idea-provider.

> Hypothesis 7: A person who scores high on openness to experience is more likely to perform well in the role of idea-provider than one who scores low.

Empathic concern
Emphatic concern is a widely used term in psychological research that refers to an emotional response of compassion and sympathy caused by perceiving another person to be in distress (Niezink et al. (2012)). It reflects an individual's intrinsic valuing of someone else's welfare (Batson (2009)) and often manifests itself in feelings of sadness, guilt, or emotional anguish. Empirical evidence suggests that this type of concern is a fundamental driver of prosocial or altruistic behaviour (Einolf (2008)), as experiencing negative emotions upon seeing another person suffer frequently causes the observer to actively try and help, if only to reduce his or her own personal level of mental distress (Schroeder et al. (1988)).

While there is some debate over the extent to which empathic concern is a fixed attribute rather than a spontaneously arising emotional reaction, some researchers do conceptualise it as a personality trait that can strongly vary between different individuals (Einolf (2008)). Empathic concern is one of the four categories measured in the IRI, where it is used to assess emotional empathy; that is to say, the degree to which a person experiences feelings of compassion and concern for an unfortunate other he or she is observing (Davis (1983)). Those who score high on empathic concern strongly feel for people in need and wish to ease their suffering, while those who score low are less affected by the well-being of others.

It seems possible that empathic concern would affect the manner in which a business leader interacts with stakeholders. Managers who display high levels of empathic concern may be more likely to feel sympathy for the plight of affected stakeholders and thus possess a greater incentive to find solutions that satisfy all parties. Conversely, managers with low empathic concern might not care as much about finding mutually beneficial solutions because the well-being of stakeholders is of lesser concern to them. Instead, they would focus their attention and efforts on satisfying their shareholders or any other stakeholders who are of direct importance to the company's survival. I therefore suggest that empathic concern is positively correlated with a leader's performance in the roles of facilitator and citizen.

> Hypothesis 8: A person who scores high on empathic concern is more likely to perform well in the roles of facilitator and citizen than one who scores low.

Perspective taking
The term perspective taking refers to the cognitive ability to conceptualise situations or events from the viewpoint of another individual (Galinsky et al. (2008)). It is often considered to be an important antecedent for empathy (Oswald (2002)) and has been linked to a variety of benefits. Research on the subject indicates that possessing the capability of leaving one's own personalised frame of reference and adapting that of someone else reduces stereotyping and facilitates the exchange between in- and out-groups (Galinsky and Moskowitz (2000)) while simultaneously providing an advantage in negotiations (Galinsky et al. (2008)).

Like empathic concern, perspective taking is one of the four personality aspects measured by the IRI. While these

two dimensions are often correlated (Davis (1983)), they describe distinct concepts. Empathic concern assesses emotional empathy, which can be described as the degree to which one sympathises with another's plight. Perspective taking, on the other hand, is a purely cognitive trait, showcasing the extent to which a person knows and understands what others are thinking or feeling. It is quite possible that someone would score high on perspective taking and therefore have a pronounced understanding of another person's inner turmoil, yet at the same time not be particularly concerned with said person's well-being, thus scoring low on empathic concern.

As such, varying levels of perspective taking may not necessarily influence a manager's motivation or goals when dealing with stakeholders. However, it still seems conceivable that perspective taking could be a valuable trait in the context of responsible leadership. Being able to fully understand the thoughts and feelings of all involved parties may facilitate the negotiation process, thereby leading to improved solutions. Leaders who possess this quality could thus resolve more complex situations than their peers who lack it. Based on this line of reasoning, I hypothesise that perspective taking positively influences the ability to perform well in the role of idea-provider.

> Hypothesis 9: A person who scores high on perspective taking is more likely to perform well in the role of idea-provider than one who scores low.

Holism

Holism is a concept that is used across a wide spectrum of scientific fields, although its connotations differ somewhat depending on the respective context to which it is applied (Pelletier (2012)). Generally speaking, it refers to the idea that the properties of a system can only be understood and analysed as a whole. Studying individual components in isolation will not produce meaningful insights because doing so ignores the dynamic interactions which occur among these components. Holism can thus be understood as an application of the Aristotelian principle that the whole is greater than the sum of its parts (Freeman (2005)).

The conception of holism as an inherent personality trait rather than an abstract ideal is a comparatively new phenomenon. Drawing from previous research on cultural differences, Choi et al. (2007) note that East Asians often seem to approach issues from another perspective than Westerners. While the former believe that all elements in the world are interconnected, the latter operate under the assumption that the universe is comprised of independent parts. Choi, Koo, and Choi expand on this observation by developing the AHS; a framework for measuring individual differences in analytic versus holistic thinking.

Those who score high on the AHS tend to exhibit a greater degree of holistic thinking. They focus their attention on the relationships between objects and their environment, seeing the world as a complex network of causal relationships that are dynamic and constantly changing (Choi et al. (2007)).

Subsequently, they are less likely to judge any information as irrelevant to a specific issue and report a greater connectedness to the environment (Leong et al. (2014)). When faced with contradictory demands, they are primarily concerned with preserving harmony by finding a common middle ground (Lechuga et al. (2011)). Conversely, those who score low on the AHS tend to employ analytic thinking, preferring to focus on individual objects removed from their contextual surroundings and being more willing to choose one side over the other when faced with potential disagreements.

Based on these findings, it seems possible that holistic thinking could be a beneficial attribute for responsible leaders. Managers who consider problems within their wider environmental context may be better at solving complex, interwoven issues that affect multiple stakeholder groups than those who only look at individual issues in isolation. Similarly, the fact that holistic thinkers display a greater connectedness to the environment and a strong concern for preserving harmony among different groups would presumably cause them to put more effort into analysing the needs of all stakeholders, even of those who are not vital to the company's continued operations or survival. I therefore argue that holism is positively correlated with the ability to succeed in the roles of facilitator, citizen, and idea-provider.

> Hypothesis 10: A person who scores high on holism is more likely to perform well in the roles of facilitator, citizen, and idea-provider than one who scores low.

3. Methods

In Chapter 2, I derived my research question from the theoretical literature and then turned it into a set of ten quantifiable hypotheses. Chapter 3 will discuss the specific process I used for collecting the necessary data as well as the statistical methods I employ to test my hypotheses.

The main goal of this master's thesis is to investigate the causal relationship between personality characteristics and responsible leadership effectiveness by performing a quantitative empirical analysis. For this purpose, I have constructed a number of multiple linear regression models, each with one of the roles of responsible leadership as dependent variable and the various demographic, biological, and personality traits I introduced in Subchapter 2.2 as independent variables. The data for these variables was collected through surveys and a scenario-based experimental study with a sample of students from the UZH and the ETHZ as participants.

This experimental research study was a collaboration between two chairs of the UZH: the Chair of Clinical Psychology and Psychotherapy from the Department of Psychology and the Chair of Foundations of Business Administration and Theories of the Firm from the Department of Business Administration. The study was designed by Dr Christian Vögtlin and Pascale Schwab in coordination with Andreas Walther and Prof. Dr Ulrike Ehlert and supervised by Dr Voegtlin.

The following subchapters will discuss our research study in greater detail. Subchapter 3.1 concerns itself with the participants; I will discuss their selection process and analyse their composition. Subchapter 3.2 focuses on the measures used in the study, detailing all observed variables and which instruments were employed to assess them. Subchapter 3.3 is dedicated to the procedures. It shows the data collection process as well as the statistical tools I use to construct my regression models. Lastly, Subchapter 3.4 discusses the limitations of our research study.

3.1. Participants

As discussed in Subchapter 2.1, responsible leadership is a complex phenomenon with a large variety of potential influence factors. In particular, it seems likely that people's personal leadership experiences or the organisational environment under which they are operating would have a profound impact on their view of responsible leadership and thus also on their effectiveness in this regard. Since our research study is primarily concerned with assessing the relationship between personality characteristics and responsible leadership, we wanted to minimise our participants' exposure to these two variables, which is why we decided to draw a sample from a pool of university students. Many of these students are likely to have never worked in an organisation, and those who did would presumably only have worked there for a few years at most. As such, their decision-making process may not be influenced by the organisational context as much as that of someone who has worked in the same company over a long period of time. Similarly, students who worked before starting their degree or who are working while finishing their bachelor's or master's degree are less likely to have ever been put in a leadership role, which suggests that their decisions would not be strongly affected by previous leadership experiences. We thus posit that using students for measuring the relationship between personality and responsible leadership is less likely to result in the aforementioned variables distorting the results than using participants who are already in the workforce. In addition, relying on students from the UZH and the ETHZ both facilitated our recruitment process and ensured that anyone who was interested would not have to travel far to partake in the experimental part of our study.

Instead of drawing a random sample, we relied on students who were interested in the subject matter and voluntarily wished to participate. Those who completed our study had the chance of winning an iPad mini or one of two Amazon vouchers with a value of 50 CHF each, but otherwise received no financial compensation for their efforts. We primarily recruited participants through two methods: By visiting various lectures and giving a short presentation to the attending students, and by sending out information over the mailing list of the Department of Psychology. In either case, we did not reveal specific details about the nature of our research study to the respective students. Instead, we simply told them that it would centre on decision making in a business context. This was done in order to prevent the participants from approaching our study with preconceived notions

and therefore a potential bias. Students who were interested in the research study upon hearing or reading our description could give us their email address, whereupon we provided them additional information as well as a link to the online survey we used to measure personality characteristics. Afterwards, they would visit the Chair of Foundations of Business Administration and Theories of the Firm to partake in the second part of the study. These procedures will be examined in greater detail in Subchapter 3.3.

Towards the end of our data collection period, we also started employing alternative recruitment methods. On several mornings and afternoons, we directly approached random students who were sitting in the main hall of the UZH and asked them if they were interested in participating. As before, we did not reveal the exact nature of our study, instead describing it as focusing on decision making in a business context. While the recruitment process discussed in the previous paragraph almost exclusively yielded students from economics or psychology, talking to those sitting in the main hall brought in quite a few participants from other fields as well. In addition to this approach, we also recruited colleagues or friends who were not affiliated with the study. While most of these people were students, four were already in the workforce. However, none of them were employed in a leadership position, and with one exception, they all had either just started working or were only working for a few years, which is why we decided to keep them in our sample.

In total, 97 people participated in the experimental part of our study. Three of them had to be discarded from our sample because they did not fill out the corresponding personality survey, leaving us with 94 participants who completed both the survey and the experimental part. The data we obtained from these 94 men and women form the basis of my multiple regression models.

Occasionally, a participant did not receive a score for one of the personality traits we measured in the survey. This could happen because the scores are computed by averaging the values a participant gives for several questions related to the personality dimension. If he or she left one of those questions out, the total score could no longer be assessed. As a result, there are three data entries missing for holism in my model, two entries for honesty-humility, openness to experience, and empathic concern, and one entry for perspective taking. Performing Little's test of missing completely at random for all independent variables in my model produced a nonsignificant result ($\chi^2 = 23.851$ with $p = 0.736$). This indicates that the entries are missing completely at random (Little (1988)), which allows me to fill them in using the expectation-maximisation algorithm (Do and Batzoglou (2008)).

When looking at the composition of the participants, there are two notable aspects. The first is the large number of psychology students. 51 people reported that they were majoring in psychology, more than half of all participants and far more than for any other field of study. This is partly because we spent a lot of time and effort on recruiting psychology students, visiting several of their lectures and

advertising on their mailing list, which is explicitly designed to introduce research studies. However, this in itself is not sufficient to explain the large discrepancy, as we also visited quite a few economic lectures, but only had 16 participants who reported that they were majoring in that field. The disproportional amount of psychology students can largely be attributed to the fact that the Bachelor of Science in Psychology is the only major at the UZH in which the students are required to collect hourly credits as trial subjects. Every psychological study conducted by members of the university rewards its participants with a certain number of hourly credits depending on the time investment said study requires. A psychology student needs to have collected at least ten of these hourly credits before he or she is allowed to graduate. Since our research study was undertaken in collaboration with the Chair of Clinical Psychology and Psychotherapy, we were authorised to reward hourly credits as well. Thus, while students from other fields were solely motivated by interest in the subject matter or the small chance of winning a price, psychology students also received a more tangible benefit for partaking in the study.

The second notable element is the ratio between male and female participants. Only 32 of the 94 people in our sample are women, while 62 are men. This gender discrepancy can be explained by our recruiting methods. Originally, we did not target specific individuals, instead introducing our study to large groups and then waiting for interested students to contact us. During that time period, the number of male and female participants was almost equal. However, later on, when we approached specific students sitting in the main hall, we focused almost exclusively on recruiting men. This was done because only men could deliver saliva samples, which we were then able to use for observing testosterone levels. Analysing these saliva samples is a costly process, and due to financial constraints, we knew from the start that we would not be able to do it for both genders. We therefore decided to solely analyse the samples of men because they tend to have higher testosterone levels on average (Sapienza et al. (2009)). Thus, the more men we recruited for our study, the more chances we had to obtain additional saliva samples and therefore collect more data on testosterone. We received a total of 52 samples, though I had to discard one from my analysis because we could not find the corresponding activity log. As such, I have 51 data points for the variable testosterone.

3.2. Measures

To properly assess the causal relationship between personality characteristics and responsible leadership effectiveness within the confines of a multiple linear regression model, I needed to turn my theoretical assumptions into a quantifiable set of dependent and independent variables. My independent variables are formed by the ten factors I have introduced in Subchapter 2.2: gender, age, psychology student, 2D:4D, testosterone level, honesty-humility, openness to experience, empathic concern, perspective taking, and holism.

The variables gender, age, and psychology student were measured by directly asking the participants about them. Gender and psychology student are dummy variables; for gender, 1 means male and 0 means female, while for psychology student, 1 is yes and 0 is no. Age is considered a discrete variable in my model because our survey only asked for full years. The 2D:4D was assessed when the participants personally came to the experimental part of the study. They were asked to sit down and lay their hand on their lap or a table. We then measured their second and fourth digit two times each and used the average of those values for computing the 2D:4D as a continuous variable. The testosterone level is also a continuous variable and was obtained by analysing the provided saliva samples.

The personality characteristics are all continuous variables which are computed by averaging multiple 5-point Likert questions, ranging from 1 = strongly disagree to 5 = strongly agree. Honesty-humility and openness to experience were assessed by using the 60-item version of the revised HEXACO personality inventory (Ashton and Lee (2009)), empathic concern and perspective taking by using the Saarbrücker Persönlichkeitsfragebogen, a validated German version of the IRI (Paulus (2009)), and holism through the AHS (Choi et al. (2007)). The mean and standard deviation of each variable can be found in Appendix A.

Cronbach's alpha is above 0.7 for the variables openness to experience (0.713), empathic concern (0.730), perspective taking (0.779), and holism (0.743), which is generally considered to be acceptable (Tavakol and Dennick (2011)). However, Cronbach's alpha for honesty-humility is only 0.651, indicating that this variable may be less reliable than the others. The difference is not enormous, so analysing honesty-humility may still yield useful insights, but its results should definitely be interpreted with caution.

In addition to these ten variables, our survey also assessed a variety of other personality characteristics, including the additional dimensions of the HEXACO model, internal and external loci of control, and the positive and negative affect schedule. However, I decided not to analyse those traits in this thesis partly because their relationship with responsible leadership behaviour is less apparent and partly because having too many parameters in my model would increase the possibility of overfitting. There is no clearly defined consensus in the theoretical literature as to how many predictors can be reliably measured with a given sample without the danger of overfitting (VanVoorhis and Morgan (2007)), but a commonly used rule of thumb is that for every independent variable, one should have a minimum of ten events (Harrell et al. (1996)). While I am analysing ten independent variables, I will be studying testosterone levels separately from the rest due to their low sample size and unique correlation with the 2D:4D. Therefore, my standard regression models include nine independent variables while my data set consists of 94 observations, which is in accordance with the one in ten rule.

Our selection of dependent variables was inspired by the multi-dimensional leadership model introduced in Part 2.1.3.

We decided that it would make sense to break down overall responsible leadership effectiveness into a set of continuous variables, ranging from one to five, which convey how well a participant is able to perform in each role of the leadership model. Thus, we have a dependent variable for how well someone considers the needs of the company (professional), the needs of the employees (facilitator), and the needs of the external environment (citizen). In addition, we measured how much the person in question was able to think outside the box and offer creative solutions to a problem (idea-provider) and how well he or she managed to consider all stakeholder needs simultaneously (total). Originally, we had also planned to include a sixth dependent variable which observes how well someone considers the needs of the customers, but critically evaluating this dimension revealed that it mostly measured the same elements as the variable professional because satisfying customer demands is usually seen as the best way to maximise profits. Therefore, we decided to discard it.

Observing five dependent variables does mean that the results of our study cannot be represented in a single regression model. Instead, I have created and analysed a number of distinct, yet related multiple linear regression models, each with one of the responsible leadership roles as dependent, and nine of the ten factors described in the previous paragraphs as independent variables.

The dependent variables were assessed by giving the participants a think-aloud protocol in which they had to take on the role of a CEO and decide on whether or not they would implement a new technology. We then graded their answer along the criteria outlined in the previous paragraphs. Every time they considered the needs of the company, employees, or external stakeholders, they received one point in the respective category. No negative points were given for stating that they did not care about a specific stakeholder group. At the end, we aggregated the number of points they received to form their score for each dimension, which could rank from 1 = lowest to 5 = highest. The advantage of this approach is that it allows us to analyse the participant's strengths and weaknesses separately, leading to a more nuanced portrayal than if we had simply focused on responsible leadership effectiveness as a whole. For instance, some people performed really well as professionals, but not well as citizens or vice versa. Additionally, we also gave the participants points for coming up with creative solutions that were not discussed in the protocol itself, which in the end formed their score for idea-provider. The score for how well they were able to perform all three roles simultaneously was determined by considering how they did in the three roles individually, but it was not a simple average. A participant who performed decently in all three roles would have a higher score in this category than one who performed really well in two roles, but not at all well in the remaining one.

It is fair to say that there is a certain degree of subjectivity inherent to this method, especially since participants often did not clearly structure their answer, which led to them repeating thought processes or mixing different considerations within the same sentence, making an objective assessment more difficult. To mitigate the issue of coder bias as much as possible, the answers were rated by five people: Dr Christian Vögtlin, Pascale Schwab, and myself as well as AL and CS, two naïve coders who were otherwise not involved in the study. I will be using my own values in the multiple regression models because I am the only person who rated all 94 participants. However, I did determine the inter-rater reliability with all other coders in order to measure the extent to which we were in agreement. I relied on Cohen's kappa for this task, as it is a commonly-used instrument for assessing inter-rater reliability (McHugh (2012)). Since the categories are ordered and the degree of disagreement plays a role, I focused on measuring the weighted kappa, using both linear and quadratic weighting. In both cases, I measured the inter-rater reliability with and without the dimension total. All calculations were performed by using the statistical tools available on the website VassarStats (Lowry (n.d.)). The full results of my analysis can be seen in Table 2 below.

While there is no universal consensus on what kappa values are considered good, a common interpretation is that values between 0.61 and 0.80 constitute a substantial agreement and values between 0.81 and 0.99 an almost perfect agreement (Viera et al. (2005)). Under this definition, my results are reassuring. The lowest observed value is 0.688, which still falls under the notion of substantial agreement. Several of the observed kappa values are above 0.8 or even 0.9, indicating that inter-rater reliability is very high when measuring the responsible leadership roles and that individual biases therefore did not distort the findings to a large extent. However, it should be pointed out that Pascale Schwab only rated the first 38 participants. As such, the reliability between her values and my own has to be analysed with caution.

There are two additional factors worth mentioning. Firstly, the kappa is always lower when the variable total is included, indicating that the inter-rater reliability for this variable is lower than it is for the others on average. This can be explained by the fact that total has less clear-cut criteria by which it could be measured, making it harder to objectively assess. Coders had to rate participants' performance over all three roles, which is more challenging than rating their performance for one individual role. Secondly, the inter-rater reliability between my values and those of the other researchers is higher than the one between my values and those of the two naïve coders. This is presumably because Christian Voegtlin, Pascale Schwab, and I were closely working together on the study and had read the same theoretical literature on responsible leadership, so we had a similar understanding of the concept and how to measure it. By contrast, the naïve coders were not familiar with the literature on responsible leadership and graded solely based on our instructions, so their conception of various terms may have differed from ours in certain respects.

Table 2: Inter-rater reliability between the values of other coders and my own

Coders	Observed Cohen's kappa			
	Linear without total	Linear with total	Quadratic without total	Quadratic with total
AL	0.715	0.688	0.831	0.816
CS	0.725	0.718	0.845	0.838
Voegtlin	0.825	0.806	0.903	0.892
Schwab	0.840	0.831	0.913	0.908

3.3. Procedures

The data gathering process occurred over a time period of several months, but with multiple breaks in-between. The first 43 participants were recruited and observed from the 15th of April 2015 to the 24th of June 2015, while the second set, consisting of 42 people, participated between the 22nd of October 2015 and the 18th of December 2015. Finally, the last nine participants joined our study after the New Year, in the time period between the 13th and 22nd of January 2016.

Everyone who wished to participate in our research study had to start by filling out the corresponding personality survey, which was hosted on the online platform Unipark (www.unipark.com). The full storybook, which includes all survey questions, can be found in Appendix B. In order to protect their confidentiality, the participants were asked to choose an alias at the start of the survey, which they would then keep for the remainder of the study. Once the survey was completed, they would make an appointment with us and come to the Chair of Foundations of Business Administration and Theories of the Firm for the experimental part of our study. While we tried to plan the appointments in such a way that there would always be two to four people at any given date, short-term cancellations meant that we occasionally only had one person show up on a particular day.

Regardless of the number of participants on a given day, the study always proceeded in the same manner. Once the people arrived, they were asked to sign a consent form, which can be found in Appendix C. Those who had volunteered to provide a saliva sample were then given the opportunity to do so. Afterwards, all participants filled out a mood scale assessing their current frame of mind as well as a short questionnaire focusing on the well-known trolley problem (Kamm (1989)). Once they were finished with these two items, we led them into a different room where we measured a number of biometrical factors: their height, weight, blood pressure, and the length of their second and fourth digit of both the left and right hand, which we later used to compute the 2D:4D. We also took a picture of each participant.

In the second stage of the experimental part, the participants were isolated from each other and given a think-aloud protocol in which they were asked to assume the role of a CEO of a multinational company. The full protocol is replicated below:

> Sie sind der CEO eines multinationalen Unternehmens in der Energiebranche mit Standort in der Schweiz. Sie müssen darüber entscheiden, ob Sie eine neue Technologie zur Förderung von Erdöl in ihrem Unternehmen einführen wollen oder nicht. Die neue Technologie ermöglicht eine höhere Fördermenge und eine Kostensenkung, da durch die Technologie weniger Personal benötigt wird. Doch die neue Technologie ist umstritten, da sie schädlicher für die Umwelt ist als die bisherige Technologie. Wie möchten sie, unter Berücksichtigung der verschiedenen Stakeholder (z.B. die Mitarbeiter, Kunden, Shareholder und die Gesellschaft), weiter vorgehen?

In English:

> You are the CEO of a multinational company in the energy sector headquartered in Switzerland. You have to decide whether you wish to implement a new technology for the extraction of crude oil in your company or not. The new technology would increase the extraction capacity and lead to a reduction in costs because fewer employees are needed for it. However, the new technology is controversial because it is worse for the environment than conventional technologies. How would you proceed when taking into account the various stakeholders (for instance, employees, customers, shareholders, and society)?

Once they had finished reading the protocol, the participants were given the opportunity to formulate and justify their own preferred strategy. We did not interfere with their thought process, nor did we set a time limit or specific formal requirements. We simply recorded their answers on tape so that we could later analyse how strongly they considered each stakeholder group and how much they brought in their own unique solutions to the problem. Both their final decision and the arguments they brought forward while reaching it were coded in accordance with the criteria outlined in the previous subchapter.

If more than one participant was present on that particular day, we also conducted a group discussion. The par-

ticipants were once again asked to assume the role of leaders in a company who have to make an important decision that affects numerous stakeholders. As with the individual think-aloud protocol, their discussion was graded along the dimensions of professional, facilitator, citizen, idea-provider, and total. In addition, we also assessed how strongly a participant dominated the discourse. While this group exercise yielded interesting insights in its own right, I decided not to include it in my thesis for two reasons. Firstly, not all participants were able to partake in one of these discussions, leaving us with fewer data points to analyse. Secondly, there are many factors which are not present in an individual assignment, but play an important role in group interactions. Properly analysing them would require setting up new hypotheses and variables, which would have been outside the scope of this master's thesis.

Originally, we had intended that the participants would grade each other's performance in the group discussion as well. However, we soon realised that they were often hesitant to give their partners a bad evaluation, leading to artificially inflated scores. As such, we decided to discard this variable halfway through the study. From that point onwards, the only thing participants had left to do after the group discussion was answering another short personality survey. This one wasn't directly related to responsible leadership, rather it was requested by a student of the Chair of Clinical Psychology and Psychotherapy for her own master's thesis. Finally, participants had the option of asking us to send them their results of the online survey. The feedback form they received is replicated in Appendix D.

Once all the data from both the initial personality survey and the experimental part had been collected, we combined it into a single file using the statistical software SPSS Statistics. I have also relied on SPSS for constructing and evaluating my multiple regression models, which will be discussed in-depth in Chapter 4.

3.4. Limitations

Our research study is one of the first attempts to empirically assess responsible leadership effectiveness. While we believe that it generally succeeds at this task, there are a number of limitations that need to be addressed. Some of them are deliberately imposed delimitations, while others are unintended side effects inherent to the study's design. In the following paragraphs, I will first discuss our specific delimitations. Afterwards, I will analyse the study's limitations and their potential influence on the observed variables while simultaneously offering suggestions as to how they can be avoided in future research.

In our study, we were solely interested in the relationship between responsible leadership and personality characteristics. As such, relying almost exclusively on students from the UZH and the ETHZ should not meaningfully distort our results. The fact that all participants joined us voluntarily, rather than being drawn from a random sample, should not have any adverse effects either because we were not trying to create generalised observations about the population of university students. However, this delimitation does mean that we cannot assess other factors that are generally considered to have an influence on leadership effectiveness, such as the organisational culture or a person's experience in the workforce. Additionally, focusing solely on traits neglects the important role different behavioural inputs and situational contingencies play in the leadership process. The results of our study can therefore never fully explain responsible leadership effectiveness. Instead, our intention was to provide valuable insights by analysing one single dimension of a complex, multi-faceted phenomenon.

Another delimitation lies in our method for assessing responsible leadership effectiveness. In Part 2.1.2, I presented two somewhat differing definitions of responsible leadership. The one from Waldman (2011) argued that being responsible simply requires a leader to carefully consider the needs of all stakeholders, while that of Maak and Pless (2006) posited that this in itself is not enough and that responsible leadership requires actively involving the affected stakeholders in the decision-making process. Our study is founded upon Waldman's definition, as we simply analysed the extent to which participants considered the various stakeholder groups in their decision. We chose to go with this understanding partly because it makes it easier to analyse the relationship between personality and leadership effectiveness in isolation from other factors and partly because it seems to be a more accurate reflection of the leadership process as it is currently being practiced in most multinational companies. That being said, it would certainly be interesting to see a study which focuses on the direct interactions between CEOs and their stakeholders and whether a close interaction between the two could facilitate responsible leadership effectiveness, and I hope to see future studies with a focus on this aspect.

In addition to these delimitations, there are a number of unintended factors which may inhibit the validity of our results. Our personality characteristics were all measured through self-assessments. The participants were asked a number of questions, and their answers in turn determined their score for a given personality dimension. Such self-assessments can lead to distorted results if the participants do not answer all questions truthfully, which can happen either intentionally if they want to present themselves in a manner that they deem to be socially desirable or unintentionally if the person in question did not read the survey carefully or has an unrealistic and inflated view of him- or herself (McDonald (2008)). In both cases, the participant's score on a personality scale would not match his or her actual personality. This issue is further exacerbated by the fact that we used two differing instruments to measure our independent and dependent variables and also assessed them on different days, making it more difficult to directly observe the relationship between personality traits and responsible leadership. While relying on validated scales should help mitigate the problem to an extent (McDonald (2008)), it is still something that has to be kept in mind when analysing the final results.

A similar problem arises when looking at the answers

given in the think-aloud protocol. While these answers were graded by us and therefore are not subject to the exact same biases as they would have been in a self-assessment, there is still some degree of uncertainty over how well they reflect a participant's actual views. Some participants may have deliberately chosen a solution to the think-aloud protocol that they deem to be socially acceptable, rather than one they would have selected if they had not been directly observed by us. In other words, there exists a potential risk that instead of giving their honest opinion on the matter, certain participants came up with answers which they thought we wanted to hear. Allowing a participant to answer the protocol isolated from any observers on a computer may be a possible remedy for this problem, although it could also be argued that this behaviour is not actually in itself a distortion of the leadership process. After all, CEOs of multinational companies are constantly facing social pressure. The willingness to change one's approach depending on what one considers socially desirable may be a useful trait to have in terms of attaining responsible leadership effectiveness.

There is also the question of how accurately our think-aloud protocol reflects the actual leadership process. Financial and temporal constraints restricted us from constructing an incredibly elaborate role-playing scenario, making it more difficult to realistically model decision making in a business context. Our participants had to choose a suitable strategy all by themselves in a very short period of time. They had no means of collecting additional information or consulting with employees or external advisors. By contrast, the implementation of a ground-breaking new technology in an actual company is frequently a long process during which various opinions and facts are collected and processed. Furthermore, our participants received no material compensation regardless of their final decision. In a real-life scenario, managers would presumably be faced with the prospect of different rewards depending on what strategy they choose to pursue, potentially skewing their priorities. A future study which responds to these limitations by expanding the scope of leadership activities participants can undertake and incorporating a reward structure that incentivises different behaviours would be a valuable contribution to the field.

Even if we operate under the assumption that the think-aloud protocol can be seen as a useful abstraction of the leadership process, there is still the risk that its specific wording influences the outcome. While we attempted to keep the protocol's description as neutral as possible so as not to prematurely influence the reader's opinion, I cannot entirely rule out the possibility that its formulations may have affected the participants in some form. Even seemingly minor and subtle differences in the phrasing could have a subconscious impact on them. For instance, the text of our protocol could be interpreted as suggesting that the central conflict lies in the additional profits generated from implementing the new technology versus the harm done to the environment. This could potentially lead to participants focusing strongly on these two dimensions while neglecting the employees. While this in itself is not necessarily a huge issue, it

does raise concerns about the replicability of the study's results. Another study using its own decision-making scenario may reach different conclusions by unintentionally steering the participants' thoughts into a certain direction through the way its central problem is framed. I thus have to be careful in my analysis. Not all observable results may have been caused by the causal relationship between personality and responsible leadership. Some could be predicated on the wording of the presented leadership problem.

One final aspect that has to be mentioned is our assessment of responsible leadership effectiveness. This is the first research study using the five dimensions professional, facilitator, citizen, idea-provider, and total to model responsible leadership. While we believe that our justification for selecting and grading these five dimensions is sound, they have not yet been validated in previous research. As such, I cannot entirely exclude the possibility that they are an inadequate instrument for accurately measuring responsible leadership effectiveness. The high inter-rater reliability between my own scores and those of the other coders is somewhat reassuring, as it indicates that the results are not strongly influenced by subjective biases. However, it is still possible to imagine that there exist conceptual issues with the scales. Responsible leadership is a complex phenomenon and trying to dismantle it into a set of distinguishable elements a challenging task, which is why I would welcome it if additional studies used our scales to aid in further calibrating them as well as increasing their accuracy and precision.

4. Results

In Chapter 3, I explained the methodology I used to collect my data. Chapter 4 will now present the actual results of my research. I will introduce each multiple linear regression model I have constructed separately and showcase its outcomes. The in-depth interpretation and analysis of my results will be the focus of Chapter 5.

Each subchapter of this chapter focuses on one of the five dependent variables I have previously introduced. Subchapter 4.1 will discuss the variable for professional, 4.2 the one for facilitator, 4.3 the one for citizen, 4.4 the one for idea-provider, and 4.5 the one for total. For each variable, I have constructed two regression models; one in which I include all nine independent variables, and one in which I only include the independent variables for which I hypothesised that they would have an influence on the respective dependent variable. I will then compare these two models to see which one produces more significant results. Whenever applicable, I have also constructed a third model which only includes variables from the first two models that were statistically significant.

For the reasons outlined in Subchapter 3.2, I did not want to include testosterone as an independent variable in my main regression models. However, I was still curious to see whether the hormone would have a noticeable effect on responsible leadership effectiveness and how it relates to the

2D:4D. Hence, I will be analysing testosterone separately in Subchapter 4.6.

4.1. Professional

This subchapter focuses on the dependent variable professional, which measures how strongly an individual considered the needs of the company in his or her decision-making process. I will first show the model in which all nine independent variables are included, henceforth referred to as model P1. It has a Durbin-Watson statistic of 2.068, which is very close to 2. Thus, we can reasonably assume that it suffers from neither positive nor negative autocorrelation and that the error terms are independent of each other (Durbin and Watson (1951)). Furthermore, looking at the skewness and kurtosis of the of the error terms reveals that their absolute value is less than twice the standard error, indicating that the error terms are approximately normally distributed (Adams and Lawrence (2014)). Performing the Breusch-Pagan test returns a p value of 0.582, which means I cannot reject the null hypothesis that the independent variables are homoscedastic for a significance level of alpha = 0.05 (Breusch and Pagan (1979)). In other words, I can operate under the assumption that no heteroscedasticity is present in my model. Table 3 below shows the findings for P1, including R, R^2, the adjusted (adj.) R^2, the standard error of the estimate (SEE), the F value, and the p value for the F-test.

P1 has an extremely low R^2 of 0.019, which suggests that only 1.9% of the variation in the dependent variable can be explained by the independent variables. The adjusted R^2 is even negative, which for the purpose of interpretation can be seen as equivalent to an adjusted R^2 of 0. Furthermore, the p value of the F-test is 0.996, meaning I cannot reject the null hypothesis that the intercept-only model is equal to my model for any reasonable significance level. Table 4 below shows the results for the individual variables, including the unstandardised regression coefficient b, the standard (std.) error, the standardised coefficient beta, the p value, and the variance inflation factor (VIF).

The VIF for all coefficients lies between 1.0 and 1.4, indicating that the model has no issues with multicollinearity (Alin (2010)). However, none of the included variables are statistically significant for an alpha of 0.05, as even the lowest observed p value is still above 0.6. As such, I cannot infer that any of the tested coefficients meaningfully influence the variable professional. These findings combined with the results from Table 3 indicate that the model P1 is a poor fit for the observed data.

In my second model, henceforth referred to as P2, I only included the independent variables for which I hypothesised that they would affect a leader's ability to perform well in the role of professional: gender, age, and the 2D:4D. The Durbin-Watson statistic for P2 is 2.034, indicating that there is no autocorrelation. The error terms are normally distributed, and the Breusch-Pagan test shows no signs of heteroscedasticity ($p = 0.486$). The results for the model as a whole and the individual coefficients are shown in Tables 5 and 6 below.

R^2 has decreased slightly, which is to be expected given that I have reduced the number of independent variables in the model. The adjusted R^2 and the p value of the F-test are slightly better than they were in P1, but still not satisfactory by any means. The same can be said about the individual coefficients. While the VIFs are all very low, and multicollinearity is thus presumably not a problem, neither of the three measured variables is significant for an alpha of 0.05. In conclusion, while the model P2 seems to be a small improvement over P1, it still is in no way a good fit for the data and lacks explanatory power. I thus find no support for my hypotheses that a leader's gender, age, and 2D:4D have any influence on his or her ability to perform well in the role of professional.

4.2. Facilitator

This subchapter discusses the models with the dependent variable facilitator, which assesses how strongly an individual considered the needs of the employees in his or her decision-making process. I will first introduce the model F1, which includes all nine independent variables. The Durbin-Watson statistic is 1.732, which is not as good as it was for the models P1 and P2, but still relatively close to 2. At most, F1 might display some very slight positive autocorrelation. Looking at the skewness and kurtosis indicates that the error terms are approximately normally distributed. The Breusch-Pagan test gives a p value of 0.222, which suggests that no heteroscedasticity is present. The findings of F1 are shown in Table 7 below.

R^2 is 0.213, which indicates that approximately 21% of the variation in the variable facilitator can be explained by the independent variables. Furthermore, the p value of the F-test is significant for an alpha of 0.05, which means that F1 is a better fit for the data than an intercept-only model would be. Table 8 below shows the results for the individual coefficients.

The VIFs are all relatively low, indicating that there is no multicollinearity in the model. From the p values, we can see that empathic concern and openness to experience are statistically significant for an alpha of 0.05. Empathic concern has a positive effect on facilitator; if the former increases by one standard deviation, the latter increases by 0.407 standard deviations. Conversely, facilitator's relationship with openness to experience is negative; if openness to experience increases by one standard deviation, facilitator decreases by 0.246 standard deviations.

The second model, F2, includes only the variables for which I hypothesised they would influence the role of facilitator. As such, I took out openness to experience and perspective taking. The Durbin-Watson statistic is 1.799. The error terms display a very slight positive skewness, indicating they might be asymmetrically distributed. Therefore, I used the White test for determining heteroscedasticity (White (1980)), rather than the Breusch-Pagan. Performing the White test returned a p value of 0.197, so the assumption of homoscedasticity seems to hold. The results for F2 are summarised in Tables 9 and 10 below.

Table 3: Summarised results for P1

R	R^2	Adj. R^2	SEE	F value	p value
0.136	0.019	-0.87	1.189	0.177	0.996

Table 4: Individual coefficients of P1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	4.658	4.350		0.287	
Gender	0.042	0.298	0.018	0.888	1.331
Age	-0.012	0.031	-0.044	0.700	1.090
Psychology	-0.149	0.285	-0.065	0.603	1.344
Honesty	0.110	0.229	0.055	0.631	1.115
Openness	-0.093	0.204	-0.050	0.652	1.055
Empathic	0.057	0.193	0.035	0.768	1.175
Perspective	0.082	0.218	0.046	0.707	1.246
Holism	-0.024	0.365	-0.008	0.949	1.186
2D:4D	-1.923	4.005	-0.054	0.632	1.098

Table 5: Summarised results for P2

R	R^2	Adj. R^2	SEE	F value	p value
0.079	0.006	-0.027	1.156	0.188	0.905

Table 6: Individual coefficients of P2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	5.421	3.817		0.159	
Gender	0.056	0.255	0.023	0.828	1.025
Age	-0.006	0.029	-0.023	0.827	1.005
2D:4D	-2.419	3.756	-0.068	0.521	1.022

Table 7: Summarised results for F1

R	R^2	Adj. R^2	SEE	F value	p value
0.462	0.213	0.129	0.849	2.530	0.013

Both the R^2 and the adjusted R^2 are smaller than in F1, which makes sense since F2 does not include the statistically significant variable openness to experience. The p value for the F-test is also slightly higher, though still significant for an alpha of 0.05. No multicollinearity seems to be present. The variable empathic concern is still highly significant, though its effect on facilitator has decreased slightly from 0.407 to 0.365. In general, it seems like F1 is a better fit for the data than F2. For the sake of completeness, I also constructed a model F3, which only includes the variables from F1 that were shown to be significant: empathic concern and openness to experience. The Durbin-Watson statistic for F3 is 1.788, and the error terms are normally distributed. However, the Breusch-Pagan test now returns a p value of 0.052, which is just barely above the threshold of 0.05. This suggests that F3 should be interpreted with caution, as slight heteroscedasticity may be present in the model. The results can be found in Tables 11 and 12 below.

F3 has the highest adjusted R^2 of all three facilitator models because it only includes significant variables. Similarly, it has the lowest p value for the F-test and the highest F value. When looking at the individual coefficients, the results are similar to F1. Both empathic concern and openness to experience are significant for an alpha of 0.05. The former has a

Table 8: Individual coefficients of F1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	0.731	3.107		0.814	
Gender	0.310	0.213	0.162	0.150	1.331
Age	-0.002	0.022	-0.009	0.930	1.090
Psychology	0.257	0.204	0.142	0.210	1.344
Honesty	0.084	0.163	0.052	0.609	1.115
Openness	-0.362	0.146	-0.246	0.015	1.055
Empathic	0.534	0.138	0.407	0.000	1.175
Perspective	-0.090	0.155	-0.062	0.565	1.246
Holism	-0.171	0.261	-0.069	0.513	1.186
2D:4D	1.032	2.861	0.037	0.719	1.098

Table 9: Summarised results for F2

R	R^2	Adj. R^2	SEE	F value	p value
0.387	0.150	0.080	0.872	2.163	0.045

Table 10: Individual coefficients of F2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	-0.760	3.131		0.809	
Gender	0.221	0.216	0.116	0.309	1.297
Age	-0.008	0.023	-0.034	0.739	1.067
Psychology	0.186	0.207	0.102	0.371	1.313
Honesty	0.069	0.167	0.043	0.681	1.105
Empathic	0.480	0.139	0.365	0.001	1.130
Holism	-0.212	0.253	-0.085	0.405	1.054
2D:4D	1.478	2.922	0.052	0.614	1.085

Table 11: Summarised results for F3

R	R^2	Adj. R^2	SEE	F value	p value
0.421	0.178	0.159	0.834	9.822	0.000

Table 12: Individual coefficients of F3

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	1.395	0.658		0.037	
Openness	-0.333	0.140	-0.227	0.020	1.011
Empathic	0.498	0.126	0.379	0.000	1.011

positive and the latter a negative influence on the dependent variable facilitator. I thus find support for the hypothesis that empathic concern positively affects a leader's ability to perform well in the role of facilitator, though I cannot say the same for gender, age, psychology student, honesty-humility, holism, or the 2D:4D.

4.3. Citizen

This subchapter is centred on the dependent variable citizen, which measures how strongly an individual considered

the needs of the external stakeholders and the environment in his or her decision-making process. The first model, C1, includes all nine independent variables. Its Durbin-Watson statistic is 2.027, which suggests that there is no autocorrelation. The error terms are normally distributed, and the Breusch-Pagan test gives a p value of 0.529, indicating that no heteroscedasticity is present. The general findings of the model are summarised in Table 13 below.

R^2 is 0.111 which indicates that 11.1% of the variation in the dependent variable citizen can be explained by the nine independent variables. The p value for the F-test is higher than the significance level of 0.05. Therefore, I once again cannot reject the null hypothesis that C1 is equal in explanatory power to an intercept-only model. These values suggest that C1 is not a good fit for the observed data. The results for the individual coefficients are shown in Table 14 below.

The low VIFs indicate that no multicollinearity is present. Of the nine observed variables, holism is significant with a p value of 0.007. The relationship between it and citizen appears to be positive; if holism increases by one standard deviation, citizen increases by 0.309 standard deviations. Thus, C1 hints at the possibility that holism might have a positive influence on the role of citizen.

The model C2 includes only the variables for which I hypothesised they would directly influence the role of citizen. As with the role of facilitator, this includes every variable except openness to experience and perspective taking. The Durbin-Watson statistic is 2.026, and the error terms are approximately normally distributed. The p value for the Breusch-Pagan test is 0.336, indicating that no heteroscedasticity is present. The summarised results for C2 can be seen in Tables 15 and 16 below.

The adjusted R^2 is slightly higher than it was for C1, although it is still relatively low. In the same vein, while the p value of the F-test is lower than in C1, it still does not meet the threshold of 0.05. Thus, I cannot confidently say that C2 is better than an intercept-only model. The low VIF of all coefficients suggest no problems with multicollinearity, but none of previously nonsignificant variables have become significant. Once again, the p value for holism is considerably lower than for any other predictor and statistically significant for an alpha of 0.05, although the p value is higher in this model than it was in C1. Looking at the bivariate correlations shows that perspective taking and holism are somewhat correlated ($r = 0.351$ with $p = 0.001$), which could explain why dropping the variable perspective taking from the model would increase the p value of holism (the full correlation table can be found in Appendix A). Since there was only a single significant variable, I decided not to construct an additional model C3. In conclusion, both C1 and C2 seem to support the hypothesis that holism has a positive influence on the role of citizen.

4.4. Idea-provider

In this subchapter, I will present the results for the dependent variable idea-provider, which shows how strongly an individual brought in his or her own unique ideas and creative solutions to the presented problem. Model I1 includes all nine independent variables. Its Durbin-Watson statistic is 1.931. The error terms display a slight positive skewness, which once again means that I cannot measure heteroscedasticity with the Breusch-Pagan test. Performing the White test instead returned a p value of 0.090, which means I cannot reject the null hypothesis that the model is homoscedastic for an alpha of 0.05. Table 17 below shows the results for I1.

R^2 is 0.185, indicating that 18.5% of the variation in the variable idea-provider can be explained through the independent variables. Furthermore, the F-test is significant for an alpha of 0.05, indicating that this model is not a bad fit for the data. The results for the individual coefficients are summarised in Table 18 below.

The VIFs show no sign of multicollinearity. Of the nine observed independent variables, three appear to be statistically significant for an alpha of 0.05: gender, psychology student, and perspective taking. Gender and psychology student are dummy variables and have a negative effect on the dependent variable. The score for men in the role of idea-provider was on average 0.519 points lower than for women. Similarly, the score for psychology students was 0.556 points lower on average than that of the other participants. Conversely, perspective taking has a positive influence on idea-provider; if perspective taking increases by one standard deviation, idea-provider increases by 0.235 standard deviations.

The second model, I2, only includes the independent variables for which I argued that they would have an impact on a leader's ability to perform well in the role of idea-provider: age, psychology student, openness to experience, perspective taking, and holism. The Durbin-Watson statistic is 1.827, and the error terms show a noticeable positive skewness. The p value of the White test is 0.688, indicating that no heteroscedasticity is present. The full results for I2 can be seen in the Tables 19 and 20 below.

The R^2 and adjusted R^2 are considerably lower than in I2, and the F-test no longer produces a significant outcome. This can be explained by the fact that the statistically significant variable gender is no longer included in the model. Looking at the individual coefficients shows that psychology student and perspective taking are still significant, while all other variables remain nonsignificant. The beta for perspective taking has barely changed from I1, but the influence of psychology student on idea-provider is now noticeably smaller. This may be because gender and psychology student have a relatively high negative correlation ($r = -0.434$ with $p = 0.000$), so the variable psychology student ends up partly compensating for the influence of gender. Given all these factors, I1 is definitely the more fitting model for this data set with a higher explanatory power.

Lastly, I once again constructed a model, I3, incorporating only the significant independent variables: gender, psychology student, and perspective taking. The Durbin-Watson statistic is 1.959, and the error terms display a positive skewness. The White test shows no signs of heteroscedasticity (p

Table 13: Summarised results for C1

R	R^2	Adj. R^2	SEE	F value	p value
0.333	0.111	0.016	0.934	1.165	0.328

Table 14: Individual coefficients of C1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	-1.162	3.419		0.735	
Gender	0.177	0.235	0.089	0.454	1.331
Age	0.028	0.025	0.120	0.268	1.090
Psychology	0.117	0.224	0.062	0.604	1.344
Honesty	-0.010	0.180	-0.006	0.957	1.115
Openness	-0.044	0.161	-0.029	0.783	1.055
Empathic	0.118	0.152	0.087	0.439	1.175
Perspective	-0.188	0.171	-0.127	0.274	1.246
Holism	0.793	0.287	0.309	0.007	1.186
2D:4D	0.824	3.148	0.028	0.794	1.098

Table 15: Summarised results for C2

R	R^2	Adj. R^2	SEE	F value	p value
0.311	0.097	0.023	0.931	1.313	0.254

Table 16: Individual coefficients of C2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	-1.775	3.340		0.596	
Gender	0.156	0.231	0.079	0.500	1.297
Age	0.023	0.024	0.102	0.339	1.067
Psychology	0.083	0.221	0.044	0.709	1.313
Honesty	-0.029	0.178	-0.018	0.869	1.105
Empathic	0.085	0.148	0.063	0.566	1.130
Holism	0.689	0.270	0.269	0.012	1.054
2D:4D	1.228	3.117	0.042	0.695	1.085

Table 17: Summarised results for I1

R	R^2	Adj. R^2	SEE	F value	p value
0.430	0.185	0.097	0.853	2.113	0.037

= 0.158). Tables 21 and 22 below show the full results for model I3.

The adjusted R^2 is slightly higher than it was in I1, as I have eliminated the nonsignificant variables. The F value is also higher, while the p value for the F-test is lower. The individual coefficients do not reveal any new insights though. All three independent variables are still statistically significant,

and while their exact value for beta has changed slightly, their general impact on idea-provider is still the same. The models I1 to I3 thus all seem to indicate that women, non-psychology students, and those who score high on the personality dimension perspective taking perform better in the role of idea-provider than their respective counterparts.

Table 18: Individual coefficients of I1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	1.621	3.120		0.605	
Gender	-0.519	0.214	-0.275	0.018	1.331
Age	-0.010	0.023	-0.047	0.648	1.090
Psychology	-0.556	0.205	-0.311	0.008	1.344
Honesty	0.249	0.164	0.158	0.132	1.115
Openness	0.091	0.147	0.063	0.534	1.055
Empathic	-0.171	0.138	-0.132	0.220	1.175
Perspective	0.333	0.156	0.235	0.036	1.246
Holism	-0.241	0.262	-0.099	0.361	1.186
2D:4D	0.234	2.873	0.008	0.935	1.098

Table 19: Summarised results for I2

R	R^2	Adj. R^2	SEE	F value	p value
0.299	0.090	0.038	0.880	1.730	0.136

Table 20: Individual coefficients of I2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	1.720	1.190		0.152	
Age	-0.002	0.023	-0.011	0.916	1.058
Psychology	-0.397	0.188	-0.221	0.037	1.061
Openness	0.025	0.149	0.017	0.869	1.021
Perspective	0.335	0.157	0.236	0.036	1.186
Holism	-0.243	0.266	-0.099	0.364	1.150

Table 21: Summarised results for I3

R	R^2	Adj. R^2	SEE	F value	p value
0.368	0.135	0.106	0.848	4.693	0.004

Table 22: Individual coefficients of I3

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	1.505	0.587		0.012	
Gender	-0.490	0.205	-0.260	0.019	1.233
Psychology	-0.602	0.196	-0.336	0.003	1.244
Perspective	0.282	0.140	0.199	0.047	1.014

4.5. Total

The last dependent variable I analyse within the confines of this thesis is total, which assesses how well an individual performed all three roles of responsible leadership simultaneously. The first model, T1, includes all nine independent variables. Its Durbin-Watson statistic shows no signs of au-

tocorrelation (2.078), and its error terms are approximately normally distributed. The Breusch-Pagan test has a p value of 0.150, which is not enough to reject the null hypothesis for an alpha of 0.05. The results for T1 are summarised in Table 23 below.

R^2 is 0.095, which means that roughly 9.5% of the vari-

Table 23: Summarised results for T1

R	R^2	Adj. R^2	SEE	F value	p value
0.308	0.095	-0.002	1.024	0.977	0.465

ation in the dependent variable total can be explained by the independent variables. However, the adjusted R^2 is less than zero, indicating that this result comes primarily from the large number of variables in the model, rather than said variables possessing any kind of explanatory power. Similarly, the p value of the F-test is larger than 0.05; therefore, I cannot reject the null hypothesis that the intercept-only model would be equal to T1. As for the individual coefficients, their results are shown in Table 24 below.

Empathic concern is the only significant variable for an alpha of 0.05. The relationship is positive; if empathic concern increases by one standard deviation, total increases by 0.247 standard deviations. This appears reasonable given that empathic concern had a noticeable effect on a leader's ability to perform well in the role of facilitator. This would also explain why openness to experience is close to being significant, though oddly enough, the p value for holism is very high despite it being statistically significant in the models C1 and C2.

In contrast to the other four dimensions of responsible leadership, I did not create any specific hypotheses as to which traits would affect a leader's ability to perform all three roles simultaneously. However, I was curious to see whether openness to experience would become significant if I were to remove some of the nonsignificant variables from T1. Hence, I created a second model, T2, which only includes the two independent variables empathic concern and openness to experience. T2's Durbin-Watson statistic is 2.153, showing no signs of autocorrelation. The error terms are approximately normally distributed, and the Breusch-Pagan test gives a p value of 0.207. Tables 25 and 26 below show the full results.

While the R^2 is slightly lower than it was in T1, the adjusted R^2 is higher due to the removal of several nonsignificant variables. Additionally, the F-test now becomes significant for an alpha of 0.05. The low VIFs indicate that no multicollinearity is present. However, while the p value of openness to experience is slightly lower than it was in T1, it still does not pass the threshold. Empathic concern shows no strong differences compared to the previous model; its p value, standard error, and beta are almost the same as they were before. Given the results of both T1 and T2, it seems reasonable to suggest that empathic concern does have a positive influence on a leader's ability to perform well in all three roles of responsible leadership simultaneously.

4.6. Testosterone

Male participants of our research study had the option of giving a saliva sample, which we then used to analyse their respective levels of the hormone testosterone. In total, 52 participants provided us with such a sample, one of which I

had to discard due to missing the corresponding activity log. While I hypothesised in Part 2.1.2 that testosterone would have an effect on responsible leadership effectiveness, I opted not to include it in my main regression model for two reasons. Firstly, research indicates that prenatal and current testosterone values are correlated (Sapienza et al. (2009)). Since I already measure the former with the right-hand 2D:4D, it seemed reasonable to leave out the latter due to potential multicollinearity issues. Secondly, the fact that I have far fewer data points for testosterone than for any other independent variable means there is a potential danger of overfitting. As mentioned in Subchapter 3.2, a common rule of thumb in regression analysis is that one ought to have a minimum of ten observed events for every independent variable. If I were to run a multiple regression with testosterone and eight or nine other independent variables, I would not be able to fulfil this condition.

Nevertheless, I was interested to see whether testosterone levels would affect any of the five responsible leadership dimensions. Therefore, I created a separate regression model which only includes the independent variable testosterone as well as the controlled variable age. There is no reason to control for gender in this model because all saliva samples came from men. I then ran this regression separately for each of the five dependent variables professional, facilitator, citizen, idea-provider, and total. The Durbin-Watson statistic for every model indicates that there is no issue with autocorrelation (the values are 1.990, 1.884, 2.177, 2.480, and 2.213 respectively). The error terms appear to be normally distributed, except in the model for idea-provider, which exhibits a slight positive skewness. I therefore performed a White test for this model ($p = 0.986$) and a Breusch-Pagan test for the other four ($p = 0.565, 0.493, 0.831,$ and 0.417). None of the five tests show any sign of heteroscedasticity. The VIFs of the coefficients further suggest that no multicollinearity is present. Table 27 on the next page shows the values for the independent variable testosterone in each of the five tested regression models. The full tables can be found in Appendix A.

As shown in the table, none of the models yielded a statistically significant result. Even the lowest p value, which is 0.217 for the model professional, is still far higher than the alpha of 0.05. I therefore cannot reject the null hypothesis that testosterone has no effect on the five dimensions of responsible leadership.

In addition to testosterone, we also analysed the levels of cortisol present in the saliva samples of our participants because recent research suggests that this hormone may have a moderating effect on the relationship between testosterone and aggressive behaviour. Popma et al. (2007) found that subjects with higher levels of testosterone were more likely

Table 24: Individual coefficients of T1

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	2.944	3.747		0.434	
Gender	0.162	0.257	0.075	0.531	1.331
Age	-0.019	0.027	-0.078	0.475	1.090
Psychology	-0.059	0.246	-0.029	0.810	1.344
Honesty	0.062	0.197	0.034	0.755	1.115
Openness	-0.304	0.176	-0.184	0.087	1.055
Empathic	0.365	0.166	0.247	0.031	1.175
Perspective	-0.101	0.187	-0.063	0.590	1.246
Holism	0.113	0.315	0.040	0.722	1.186
2D:4D	-0.109	3.451	-0.003	0.975	1.098

Table 25: Summarised results for T2

R	R^2	Adj. R^2	SEE	F value	p value
0.278	0.078	0.057	0.993	3.824	0.025

Table 26: Individual coefficients of T2

	Coefficient b	Std. Error	Beta	p value	VIF
(Constant)	2.812	0.784		0.001	
Openness	-0.309	0.167	-0.187	0.068	1.011
Empathic	0.335	0.149	0.227	0.028	1.011

Table 27: Values for the variable testosterone in each of the five models

	Coefficient b	Std. Error	Beta	p value
Professional	-0.005	0.004	-0.177	0.217
Facilitator	0.004	0.003	0.171	0.233
Citizen	-0.001	0.003	-0.034	0.814
Idea-provider	0.003	0.003	0.140	0.331
Total	-0.003	0.004	-0.125	0.373

to display overt aggressive behaviour, but only if their cortisol levels were low. If both testosterone and cortisol levels were high, this effect would disappear. Since my hypothesis argues that testosterone influences responsible leadership effectiveness at least partly because it is related to aggressive behaviour, it seems reasonable to suggest that I would have to look at the ratio between testosterone and cortisol in order to obtain meaningful results. I therefore constructed five additional regression models, one for each of the dependent variables, in which I keep the independent variable age, but replace testosterone with C:T, the ratio between a participant's cortisol and testosterone levels.

Once again, the number of observations for every model is 51. The Durbin-Watson statistics are all relatively close

to two (1.985, 1.864, 2.181, 2.565, and 2.214), and the error terms for professional, facilitator, citizen, and total are approximately normally distributed, while idea-provider exhibits a slight positive skewness. The White test ($p = 0.692$) and the four Breusch-Pagan tests ($p = 0.744$, 0.133, 0.678, and 0.524) give no indication of heteroscedasticity. The VIFs show no sign of multicollinearity either. Table 28 on the next page shows the results for the variable C:T. The full tables can be found in Appendix A.

As before, none of the models produce any statistically significant results for an alpha of 0.05. Thus, even when controlling for the potential influence of cortisol, I cannot find any support for the hypothesis that testosterone has an influence on responsible leadership effectiveness. Lastly, I

Table 28: Values for the variable C:T in each of the five models

	Coefficient b	Std. Error	Beta	p value
Professional	-0.101	0.235	-0.062	0.669
Facilitator	-0.208	0.165	-0.178	0.213
Citizen	-0.012	0.195	-0.009	0.950
Idea-provider	-0.229	0.155	-0.208	0.147
Total	-0.133	0.211	-0.088	0.531

looked at the correlation between the variables testosterone and 2D:4D. I obtained an *r* of 0.268 with a *p* value of 0.057. As such, the correlation is just barely nonsignificant for an alpha of 0.05.

5. Discussion

In Chapter 4, I presented the results of my empirical regression analysis without any value judgements. This chapter will now focus on the interpretation of said results. I will see how well they compare to my proposed hypotheses and attempt to find reasonable explanations for any discrepancies.

My analysis will follow the same basic structure as Subchapter 2.2 did. Thus, Subchapter 5.1 is concerned with the demographic and biological factors I measured, while Subchapter 5.2 addresses the personality traits. Additionally, Subchapter 5.3 will note more general observations about the models I have constructed, which are not related solely to one single variable. Lastly, Subchapter 5.4 will summarise the conclusions of my thesis and discuss avenues that future research could focus on.

5.1. Demographic and biological characteristics

This subchapter discusses the five demographic and biological traits that were measured in my regression models: gender, age, psychology student, and the 2D:4D as well as testosterone levels.

Gender

In Hypothesis 1, I stated that men are more likely to perform well in the role of professional because stereotypically male traits would aid in maximising the shareholder value, while women are more likely to perform well in the roles of facilitator and citizen because stereotypically female traits would be more conductive to interacting with employees and the external environment. Looking at the results from Chapter 4 reveals that this hypothesis is not supported by the data. Gender has no significant influence on any of the three responsible leadership roles.

The lack of a direct causal relationship may be due to a variety of reasons. As has been mentioned before, research studies focusing on gender differences in leadership have produced mixed results, and observed variations frequently disappear when looking at a formal organisational setting (Barbuto et al. (2007)). While our research study did not take place in such a setting, it is still conceivable that

by assuming the role of a leader in the think-aloud protocol, the participants adapted a different leadership approach than they would have used in a less standardised environment. Furthermore, Barbuto et al. (2007) also found that gender differences seem to be more apparent at lower levels of education, while disappearing at higher levels. Since our study focused almost exclusively on students from the UZH and the ETHZ, the gender differences would therefore be less pronounced.

Finally, it has to be pointed out that our research study did not assess the direct interactions between leaders and their relevant stakeholder groups; instead, we measured how strongly a participant considered the needs of his or her stakeholders. While it is possible that stereotypically female traits may indeed be useful for cooperating with employees and external associates, this does not necessarily mean that said traits would also make one more willing to actually consider the needs of these groups. Likewise, even if stereotypically male traits provide an advantage when it comes to generating shareholder value, there is no guarantee that these same traits would affect a person's desire to actually prioritise shareholder demands. Thus, the fact that I found no gender effect on the three responsible leadership roles does not necessarily mean that no such effect exists, but could rather indicate that our study design is inadequate for measuring it.

While I did not detect any significant influences for professional, facilitator, and citizen, gender differences do seem to have an impact on the role of idea-provider. Women scored noticeably higher on this dimension than their male peers. One potential explanation for this phenomenon is that it is caused by gender differences in creativity. Women may be more creative on average than men, which would allow them to come up with more inventive solutions and subsequently perform better as idea-providers. However, while some research on the subject does indicate that women perform better in specific domains of creative thinking (Averill (1999); Bowers (1971); Dudek et al. (1993)), the overall results are less conclusive. As Baer and Kaufman (2008) put it in their review of the literature: "In some cases … there are significant numbers of studies in which one group or the other scores higher, but these are generally counter-balanced by studies showing just the opposite" (p. 28). These findings suggest that creativity on its own may not be sufficient to fully explain the observed gender differences.

Additional factors that might merit consideration are the potential differences in motivation or emotional involvement between genders. While a high level of creativity can certainly be advantageous for coming up with unique solutions, it in itself is meaningless if the person in question has no desire to actually think about alternative strategies. Thus, a person who performs well in the role of idea-provider would need both the means of thinking outside of the proverbial box and the inclination to actually do so. Within the confines of our research study, this could mean that participants who scored high on idea-provider were more willing to bring in their own thoughts and ideas because they felt more emotionally involved in the think-aloud protocol. Various studies have shown that women display more empathy than men on average (Gault and Sabini (2000); Toussaint and Webb (2005)). This could indicate that women experience a stronger connection to the plight of the various stakeholders involved in the think-aloud protocol, which in turn could cause them to come up with new solutions in an attempt to make everyone content. Meanwhile, men might have felt more distant from the presented problem on average, causing them to stick to more conventional approaches because they did not care enough about satisfying everyone to explore alternative ideas.

Age

Hypothesis 2 stated that age would have a positive effect on all three roles of responsible leadership (professional, facilitator, and citizen) as well as on the role of idea-provider. I justified this statement by arguing that older leaders tend to be more team-oriented and can use their accumulated experience to find better solutions to existing problems. However, my regression models do not support the hypothesis, as age was found to be nonsignificant for every dependent variable. This may be partly because, as discussed in Part 2.2.1, the relationship between leadership behaviour and age is more complex than Hypothesis 2 would suggest. Younger leaders tend to be highly motivated and flexible (Korac-Kakabadse et al. (1998)), traits that could potentially counterbalance some of the advantages older leaders possess.

Perhaps more importantly, the age span we surveyed in our empirical research study was extremely narrow. Almost all of our participants were between 18 and 30 years old. Only three participants were older than 30, and only one of those was older than 40. This limits our study's ability to gauge age-related effects, as minor differences in years would presumably not affect leadership behaviour to an enormous extent. Furthermore, age-related effects are often caused by differences in experience. Older leaders do not differ from their younger peers purely because of the additional years they have accumulated, but rather because during those years, they were able to collect additional experience in their respective field or organisation. Since almost all of our participants were students, it seems reasonable to assume that their experiences would be relatively similar. Hence, it is not too surprising that I was unable to find any support for the assumption that age affects responsible lead-

ership effectiveness; not necessarily because such effects do not exist, but because our study design does not allow us to optimally measure them. Controlling for age in my multiple regression models was still a worthwhile endeavour, but if one wanted to specifically assess the impact of age on responsible leadership behaviour, one would have to conduct a study whose subjects exhibit greater differences in age and correspondingly a large variety of unique experiences. Perhaps it would even make sense to replace age with a new variable that measures years of experience instead.

Psychology student

Hypothesis 3 argued that students who major in psychology would be more likely to perform well in the roles of facilitator, citizen, and idea-provider than students from other fields because they tend to show greater empathy and a higher degree of creativity. However, the results of my regression models do no support this statement. The variable psychology student was found to have no significant impact on facilitator and citizen. Additionally, while it did influence idea-provider, the relationship was negative, which means psychology students were less likely to score high on this dimension than students who major in other fields.

The lack of a causal relationship between psychology students and the roles facilitator and citizen may indicate that these students are not necessarily more empathic than the other participants we observed. My assumption in this regard was based on a single research study conducted with students from universities in Serbia (Dimitrijević et al. (2011)). It is possible that these results cannot be generalised without any caveats. Alternatively, the results of my regression analysis could also be caused by the fact that my models already control for empathy to a certain extent by assessing the personality characteristic empathic concern, which does have a significant effect on the role of facilitator. Looking at the bivariate correlations reveals that psychology student and empathic concern are positively correlated ($r = 0.207$ with $p = 0.045$). Thus, the variable psychology student might provide no additional explanatory power because its potential influence on the dependent variables is already covered by empathic concern.

The negative correlation between psychology student and idea-provider is similarly notable, as it would seemingly contradict the notion that those students are more creative than others. However, there are other potential explanations as well. One is that the psychology students simply lacked the necessary framework to conceptualise unique approaches to the problem. Our think-aloud protocol asks them to imagine themselves as CEOs of a multinational company, a role which they presumably did not encounter during their normal studies. Whereas students with a background in economics would be more accustomed to this type of scenario and consequently also more able to come up with highly inventive concepts. The difference in score could therefore be explained as a matter of experience rather than creativity. One limitation to this interpretation is that only a small number of non-psychology students actually majored in an eco-

nomic field and thus have a study-related advantage, while the other participants would be just as unfamiliar with these kinds of scenarios as the psychology students are.

Another, perhaps more convincing, explanation for this phenomenon lies in the study design. As mentioned in Sub-chapter 3.1, psychology students were the only participants in our research study who received a tangible compensation for their efforts, as they were granted hourly credits. Hence, while the other participants joined out of interest in the subject matter, some of the psychology students may have participated solely because of the promised reward. In turn, this might have affected their motivation and the level of emotional engagement they had with our think-aloud protocol. It is conceivable that they did not care as deeply about the presented scenario as the other participants on average and therefore did not wish to expand any strong effort on thinking about alternative solutions either.

Second-to-fourth digit ratio and testosterone
Hypotheses 5 and 6 were concerned with the 2D:4D and the testosterone level respectively. I argued that a low 2D:4D and high levels of testosterone would be positively correlated with the role of professional and negatively correlated with the roles of facilitator and citizen. However, this did not turn out to be the case in my regression models. Neither of the two variables had a significant effect on any of the five responsible leadership dimensions.

As discussed in Part 2.2.1, several researchers argue that testosterone increases the desire for social dominance (Mazur and Booth (1998)). Following this line of argumentation, those with a low 2D:4D or high levels of testosterone would be more likely to choose strategies which increase their social standing. In a corporate setting, this could mean that they are more willing to fulfil the role of professional because the shareholders have the best means of rewarding them for their behaviour. However, in our research study, the participants were not rewarded for maximising profits. Focusing on the role of professional to the exclusion of facilitator and citizen therefore did not provide them with any social advantages. It is possible that the lack of a specific reward structure in our think-aloud protocol would cause them to behave differently than they would have in a real-life example, neutralising the effects which form the foundation of my proposed hypotheses. Additionally, there is also the issue of accurately measuring the 2D:4D. In our study, we asked participants to keep their hand still and then directly assessed the length of the second and fourth digit using a Vernier caliper. While we measured both fingers twice and took the average of the two values to ensure greater consistency, our method is still susceptible to slight variations. Many of the observed ratios were very close to each other, so even a small error in the reported value could have a significant impact on the relative comparisons. Thus, the lack of significance could at least partly be the result of imprecise measuring. This would also explain why I could find no significant correlation between the 2D:4D and current testosterone levels. If further research studies wish to incorporate the 2D:4D in their analysis, I would recommend relying on more precise measurement methods, for instance analysing photocopies of the respective hands.

5.2. Personality characteristics

This subchapter discusses the five personality characteristics I have observed: honesty-humility, openness to experience, empathic concern, perspective taking, and holism. As before, I will briefly restate my original hypothesis, compare it to the actual results from Chapter 4, and attempt to explain potential divergences.

Honesty-humility
In Hypothesis 6, I argued that those who score high on honesty-humility would perform better in the roles of facilitator and citizen because they are more willing to engage stakeholder demands even if there is no immediate financial incentive involved in doing so. However, the findings in Chapter 4 do not support my statement, as honesty-humility was not found to have a significant influence on any of the five dimensions of responsible leadership effectiveness. Though, as I have pointed out in Subchapter 3.2, Cronbach's alpha for honesty-humility was below the commonly accepted threshold of 0.7, indicating that the results for this variable may be less reliable than for the other dimensions of personality.

The lack of significant relationships could potentially indicate that honesty-humility is not as important to the responsible leadership process as the underlying theory would suggest, or it could mean that its impactful elements are already covered by different variables in the models. I have mentioned in Part 2.2.2 that honesty-humility is seen as an expression of fairness, but fairness is also commonly attributed to the personality dimension empathy (Page and Nowak (2002)). It is thus imaginable that adding honesty-humility to the model provides no additional explanatory power because its positive effects are already incorporated into empathic concern.

Additionally, the non-significance could also partly be the result of our study's design, specifically its lack of direct financial compensation for any of the participants. An important facet of honesty-humility is the dimension greed avoidance, which measures how strongly an individual is motivated by material gains. Those who score high on greed avoidance and correspondingly also high on honesty-humility overall tend to be less inclined to profit from the exploitation of others (Ashton and Lee (2007)). This behavioural trait may potentially encourage CEOs to act in a responsible manner in a real-life setting, as it would deter them from focusing on profit maximisation at the expense of everything else, but it does not play a large role in our think-aloud protocol because the individuals received no specific compensation regardless of whatever strategy they chose to pursue.

Openness to experience
Hypothesis 7 was concerned with the personality dimension openness to experience. Specifically, I argued that openness to experience would have a positive effect on a leader's ability

to perform well in the role of idea-provider due to its correlation with creativity. However, my observed data does not support this assumption, as I could find no significant relationship between the variables idea-provider and openness to experience. Somewhat surprisingly, openness to experience does have a significant impact on the role facilitator, but this relationship is negative. In other words, the higher an individual scored on openness to experience, the less likely he or she was to consider the needs of the employees.

The fact that openness to experience is nonsignificant for the role of idea-provider seems surprising, as this personality dimension is frequently considered to be strongly correlated with creative behaviour (Baer and Oldham (2006); McCrae (1987); Silvia et al. (2014)). Thus, one would assume that it simultaneously affects an individual's ability to come up with new and improved solutions to existing problems. However, as I have discussed before, our assessment of the role of idea-provider does not solely measure creativity, but also the motivation to actually employ it. It is possible that participants who score high on openness to experience are indeed more creative on average than their peers, but that does not necessarily mean that they were more interested in the presented scenario. If they did not care enough about the think-aloud protocol, they may simply have been unwilling to expend any effort on coming up with innovative approaches.

Perhaps the most surprising finding of my regression analysis is that openness to experience appears to have a negative influence on the role of facilitator. This relationship was found in both models that included openness to experience as a factor, F1 and F3, and has a *p* value that is considerably lower than the significance level of 0.05. The beta was -0.246 in F1 and -0.227 in F3, indicating a moderate influence. What makes these results so surprising is that research on the interactions between openness to experience and empathy has found a positive relationship; that is to say, higher levels of openness to experience correspond to higher levels of empathy (Magalhães et al. (2012)). Based on these results, one would expect that openness to experience also has a positive effect on facilitator, if any, but the opposite can be observed in my models.

It is difficult to pinpoint the exact cause of this phenomenon. A potential explanation may be the correlation between openness to experience and narcissism. Several studies have found that these two personality dimensions are related to each other (Furnham et al. (2013); Lee and Ashton (2005)); subjects who score high on openness to experience are also more likely to display narcissistic traits. Since narcissism is characterised by feelings of superiority, a desire for dominance, and a general willingness to exploit others for personal gain (Raskin and Terry (1988)), it could conceivably diminish one's ability to perform well in the role of facilitator.

On the other hand, the negative relationship could be a result of the specific scenario presented in our think-aloud protocol. It seems reasonable to assume that participants with a high openness to experience would also tend to be more open towards newer technologies. They may therefore be more likely to implement the improved crude oil extraction method than others even if this decisions harms the employees. If this interpretation is true, the negative impact of openness to experience on facilitator would be specific to our scenario, rather than being a general principle, and therefore not necessarily appear in further studies. However, both of the mentioned explanations are highly speculative in nature, and I would recommend that future research on the subject of responsible leadership continues to analyse the impact of openness to experience to see whether my findings can be confirmed or refuted.

Empathic concern

I argued in Hypothesis 8 that empathic concern would positively influence a leader's performance in the roles of facilitator and citizen because he or she would be more strongly affected by the plight of stakeholders and therefore also more motivated to find mutually beneficial solutions. The results of my empirical regression analysis partly support this hypothesis. Empathic concern does indeed seem to positively affect an individual's performance in the role of facilitator. However, it appears to have no significant effect on the role of citizen.

The positive relationship between empathic concern and facilitator is present in all three models F1 to F3 and significant even for a very low alpha of 0.001. The beta values are 0.407 for F1, 0.365 for F2, and 0.379 for F3, indicating that the impact of empathic concern is quite considerable. Furthermore, empathic concern is the only personality trait that affects a person's ability to perform all three roles simultaneously, showing a significant influence on the variable total in models T1 and T2. These results seem to confirm my initial assumption that empathy plays an important role in fostering responsible leadership effectiveness. It appears that highly empathic individuals are more likely to consider the needs of their employees and thereby act more responsibly in general. This may be either because they are better able to figure out what concerns said employees, because they are more willing to go out of their way to accommodate them, or a combination of both of these factors.

It is a bit surprising that empathic concern does not affect the performance as citizen, but this may be a result of the abstract nature of said role. In our think-aloud protocol, the negative consequences for the employees are immediately apparent. There is a direct causal link between implementing the new technology and having to fire parts of the staff. Conversely, it is not entirely clear how the environmental pollution will affect other human beings. While it is likely to have some negative impact, there is a degree of abstraction to the decision that makes it difficult to directly relate to the plight of the affected stakeholders. Participants who score high on empathic concern may thus be more likely to feel the suffering of the employees and try to ease their burden, while they are less concerned with the problems of external stakeholders, which are not as immediately obvious to them.

Perspective taking

In Hypothesis 9, I argued that scoring high on the person-

ality dimension perspective taking increases one's likelihood of performing well in the role of idea-provider. This hypothesis seems to be supported by the results of my empirical regression analysis. Perspective taking does indeed have a significant and positive effect on the dependent variable idea-provider, while having no influence on any of the other four dimensions of responsible leadership effectiveness.

Perspective taking's relationship with the role of idea-provider is significant in all three models I1 to I3 for an alpha of 0.05, though its p values are slightly higher than the ones for the variables gender and psychology student. The beta coefficient is 0.235 for I1, 0.236 for I2, and 0.199 for I3, indicating that the effect is moderately strong. These results do seem to support the theoretical notion that, unlike empathic concern, perspective taking does not directly affect an individual's performance in the three roles of responsible leadership, but it does aid in coming up with innovative and mutually beneficial solutions to existing problems. Hence, I posit that it is a valuable trait for establishing overall responsible leadership effectiveness.

Holism

My final hypothesis, Hypothesis 10, stated that holism would positively influence a leader's performance in the roles facilitator, citizen, and idea-provider because holistic thinking may help in solving complex, interwoven issues and is correlated with a greater concern for the environment. My models for the dependent variable citizen offer some evidence for part of this hypothesis, as holism has a significant and positive effect there. However, it does not even come close to having a significant effect on facilitator and idea-provider.

The relationship between holism and the role of citizen is significant in both models C1 and C2. The beta coefficient for holism is 0.309 in C1 and 0.269 in C2, which would indicate a moderately strong positive influence on the role of citizen. As mentioned, those who score high on holism report a greater connectedness with nature (Leong et al. (2014)), which might cause them to be more concerned about the potential environmental pollution that the new technology would bring than those who subscribe to a more analytic view of the world. In a similar vein, holistic thinkers would presumably be more likely to regard the company they work for as part of an interconnected system, rather than a completely separate institution. As such, they might be more receptive to the idea that harming the environment would negatively affect the entire network and therefore also the corporation in the long-term, causing them to give greater consideration to their responsibilities as citizen. However, it should be pointed out that holism had more missing observations than any of the other variables. While I used the expectation maximisation algorithm to fill in these empty data points, I cannot rule out the possibility that the missing entries reduce the reliability of the results.

Holism does not significantly impact the role of facilitator. Perhaps this is an indication that holistic thinkers are less concerned with what they perceive to be relatively small-scale issues. Their attention is centred on the larger, network-

spanning implications of the strategic decisions they have to undertake, and they may therefore not be as concerned with the plight of the individual employees. Holism's lack of influence on the role of idea-provider might be due to its positive correlation with perspective taking, a variable which is significant in all three models I1 to I3. It is possible that holism does not impact the role of idea-provider because its potential effects on creativity are already covered by perspective taking. In conclusion, while holism does not affect facilitator and idea-provider, my analysis does seem to suggest that it has a positive influence on citizen. As such, I would welcome it if further studies could take a more in-depth look at this potential relationship.

5.3. General observations

In addition to the individual personality characteristics that have been discussed in the previous two subchapters, there are a few general observations on my multiple linear regression models that are worth mentioning as well. For one, all constructed models have very low R^2 values. Even F1, which seems to be the most accurate model of the entire group, only has an R^2 of 0.213, indicating that almost 80% of the variation in the dependent variable facilitator cannot be explained by my chosen predictors. The adjusted R^2 is similarly low in all models and even negative in some select cases. However, these results were not entirely unexpected. Leadership is a multi-faceted phenomenon with a nearly unlimited number of potential influence factors. As such, even complex models will have trouble representing all relevant aspects. Furthermore, our study focused by design on the relationship between personality traits and responsible leadership effectiveness. We did not analyse the impact of behavioural attributes or contingency factors, which are considered to be of great importance in the theoretical literature (Derue et al. (2011); House and Aditya (1997)). It would therefore have been rather surprising if my models were able to explain all or even the vast majority of the variation in my dependent variables. My findings suggest that although traits play an important role in generating leadership effectiveness, they are not the only factors capable of doing so.

When looking at the individual models, it appears that the role of professional is the hardest one to properly assess and measure. In both P1 and P2, none of the observed independent variables are even remotely significant for an alpha of 0.05. Correspondingly, the p value for the F-test in both cases is extremely high. Neither P1 nor P2 thus seems to hold any explanatory power. This may simply indicate that my selection of traits was misguided. Perhaps the nature of the professional role is different from how I conceptualised it theoretically. I primarily argued that assertiveness and a desire for increased social status and dominance would cause leaders to focus more strongly on fulfilling shareholder needs and expectations, but there may be alternative aspects which I have failed to consider. It is conceivable that the role of professional is more heavily influenced by traits such as loyalty or devotion to the company. Furthermore, some participants in our study argued that the long-term survival of the firm

also benefits the employees who otherwise would lose their jobs. This line of thought suggests that the interactions between professional and facilitator are not as conflicting as some of my hypotheses indicate and that traits which positively influence one role would potentially also help in the other.

Another possible explanation for the lack of any significant variables in regard to the role of professional may be that the participants were already predisposed to give it special attention. While public discussions on leadership have put a stronger focus on ethical decision making in recent years, society still tends to view CEOs as being obligated towards their shareholders first and foremost (Waldman and Galvin (2008)). One would expect that the participants of our study reflect these values to some extent as well. In other words, it is possible that most participants, regardless of their respective personality traits, felt responsible for ensuring the survival of the corporation and the fulfilment of shareholder demands because they consider these tasks to be the primary duty of their position. This would also explain why the average score in the category professional was higher than for the other two roles (2.894 compared to 2.011 for facilitator and 2.872 for citizen). Furthermore, listening to the individual recordings of the think-aloud protocol reveals that while some participants did not analyse the needs of the employees or the external stakeholders in their answers, almost everyone gave at least some consideration to the needs of the company. This observation is congruent with the notion that professional is somehow seen as being more important or more pertinent to a CEO's activities than facilitator and citizen.

In contrast to professional, the models for facilitator were by far the most significant, with very low p values for the F-tests and the highest R^2 and adjusted R^2. Looking at the individual results of the participants also reveals that the average score in the role of facilitator was lower than for the roles professional and citizen. This could indicate that the needs of the employees are treated as being less important, perhaps once again reflecting societal values. As discussed before, professional is indisputably seen as a central aspect of leadership. Meanwhile, recent corporate scandals, such as the BP oil spill or Nestlé's usage of unsustainable palm oil, have put a strong emphasis on the environmental components. Participants may thus have been primed to critically analyse any potential harm to the environment, whereas issues facing the workers are not as deeply embedded in their frame of reference. However, the lower average score in the role of facilitator may also have resulted from the wording of our think-aloud protocol. Our described setup could be understood primarily as a conflict between higher profits and greater harm to the environment, with the potential dismissal of employees being an ancillary problem. If some participants conceptualised it this way, it would explain why their score for professional and citizen is higher than the one for facilitator.

The models for the variable total are difficult to comprehensively analyse, as I did not create any specific hypotheses for them. Furthermore, the inter-rater reliability between my

own scores and those of my colleagues are lower for this dimension than the others, which is not too surprising given that it is harder to objectively assess how well an individual performs in all three roles simultaneously than how well he or she performs in each role separately. What is interesting to note is that the only significant variable in models T1 and T2 is empathic concern, with openness to experience having the second-lowest p value. Both variables also influence facilitator, which would suggest that those who performed better than average in this role also performed well in all three roles simultaneously. Once again, this may be due to the average score for facilitator being lower than for the other two roles. It seems that most participants were relatively good at performing as professional and citizen, which is why their score for all three roles together strongly depends on whether or not they also performed well as facilitator.

Lastly, it should be mentioned that the role of idea-provider had the lowest average score of all five dimensions (1.968). This may be partly due to the way it was measured, but it could also indicate that most participants were hesitant to bring in their own ideas and solutions. Lack of motivation may be a reason for this behaviour, although the fact that participation was voluntary and that many students received no material compensation for their efforts would suggest that they had at least some interest in the subject matter. Another explanation is that the participants felt they lacked the necessary expertise to offer unique solutions, as most of them were not enrolled in a major which is concerned with decision making in a business context. Finally, it could indicate that it did not even occur to them to try unconventional ideas, instead believing that relying on standard approaches would be the best idea. If this was the case, then perhaps fostering an individual's willingness to think outside of the box and move beyond conventional narratives would be an effective way of increasing his or her effectiveness as a responsible leader.

5.4. Conclusions

The goal of this master's thesis was to empirically assess and subsequently analyse the relationship between a variety of personality characteristics and responsible leadership effectiveness. For this purpose, I first established a set of ten demographic, biological, and personality factors and then hypothesised about their potential influence on responsible leadership behaviour. Using data gathered in an experimental research study in collaboration with members of the Chair of Clinical Psychology and Psychotherapy and the Chair of Foundations of Business Administration and Theories of the Firm at the UZH, I constructed a number of multiple linear regression models to test whether my assumptions would match reality. In this subchapter, I will summarise the most important results, briefly discuss the limitations of our study, and highlight aspects that future research could focus on. Table 29 on the next page lists all variables which were shown to be significant in one or more models for a significance level of 0.05.

Table 29: Values for the variable C:T in each of the five models

	Model	Coefficient b	Beta	p value
Gender	I1	-0.519	-0.275	0.018
	I3	-0.490	-0.260	0.019
Psychology student	I1	-0.556	-0.311	0.008
	I2	-0.397	-0.221	0.037
	I3	-0.602	-0.336	0.003
Openness to experience	F1	-0.362	-0.246	0.015
	F3	-0.333	-0.227	0.020
Empathic concern	F1	0.534	0.407	0.000
	F2	0.480	0.365	0.001
	F3	0.498	0.379	0.000
	T1	0.365	0.247	0.031
	T2	0.335	0.227	0.028
Perspective taking	I1	0.333	0.235	0.036
	I2	0.335	0.236	0.036
	I3	0.282	0.199	0.047
Holism	C1	0.793	0.309	0.007
	C2	0.689	0.269	0.012

Of my ten hypotheses, only one was fully supported by the empirical results: Hypothesis 9. Perspective taking is positively correlated with a leader's ability to perform well in the role of idea-provider in my regression models, suggesting that the ability to put oneself in someone else's shoes and see a situation from his or her point of view is a valuable trait to have as a responsible leader.

Additionally, I found partial support for Hypothesis 8, as empathic concern has a positive influence on the roles of facilitator and total. Those who score high on this personality dimension are thus more likely to consider the needs of the employees and the needs of all stakeholder groups simultaneously, which would indicate that responsible leadership effectiveness increases when someone is strongly concerned with the well-being of others. This finding is congruent with the theoretical literature in the field, which frequently ascribes an ethical dimension to responsible leadership behaviour (Maak and Pless (2006); Waldman (2011)). However, contrary to my expectations, empathic concern does not affect the role of citizen.

My results partially support Hypothesis 10, as holism has a significant impact on the role of citizen for an alpha of 0.05. This positive relationship is congruent with the theoretical literature and earlier research on the subject, which is why I believe that following up on this result would be of great value for the field of responsible leadership as a whole. However, it appears that holism has no influence on the roles of facilitator and idea-provider, despite my initial assumptions to the contrary.

I also found some unexpected results that were not covered by my ten initial hypotheses. Gender and psychology student both had an influence on the role of idea-provider, as men and psychology students were less likely to perform well in this dimension than the other participants. With regard to the former, this result may be related to gender differences in empathy and therefore in emotional involvement in the think-aloud protocol. As for psychology students, the result may very well have been caused by the peculiarities of our study design and its reward structure.

Furthermore, openness to experience was shown to have a negative influence on the role of facilitator, which I did not foresee in Hypothesis 7. This may be because I have neglected its correlation with narcissism, or it may be a result of the specific subject matter of our think-aloud protocol. Either way, the result is surprising, and it would be interesting to see whether it can be replicated in subsequent studies.

None of the other hypotheses I have created were supported by the results of my regression analysis. Of the ten observed traits, four had no influence on any of the five responsible leadership dimensions: age, the 2D:4D, testosterone levels, and honesty-humility. To a certain extent, this is not unexpected. The lack of empirical studies on the subject meant that I did not have a wealth of data to go on when choosing the traits for my models. As such, I based my selection on theoretical notions or studies that have been undertaken in related fields. The non-significance of some

of these selected traits may indicate that my respective hypotheses were founded on faulty premises or that there exist additional aspects to these dimensions which I have failed to appropriately consider. At the same time, I cannot entirely exclude the possibility that these results were at least partly caused by limitations in our research study and that the non-significant variables may exhibit a significant influence in a study with a different design.

In general, I found that participants scored higher on average in the roles of professional and citizen than in the role of facilitator, perhaps reflecting societal values and expectations. Fostering their awareness for the needs of the employees would therefore seem to be a promising way to increase their overall responsible leadership effectiveness. Similarly, most participants scored low in the role of idea-provider, indicating that they were unwilling or unable to come up with innovative approaches. As with facilitator, training or educating them in this regard could be a valuable step in turning them into responsible leaders.

There are a number of limitations that have to be mentioned when critically evaluating our empirical research study. For one, we did not include any behavioural or contingency factors. Nor did we analyse the importance the organisational environment or previous work and leadership experiences play in creating responsible leadership effectiveness. These omissions were intentional, as we were solely focused on analysing the relationship between personality traits and responsible leadership, but they can be seen as interesting starting points for future research.

Since our input factors were collected through a self-assessment, there is the potential danger that the participants intentionally or unintentionally distorted the results by having an unrealistic view of themselves. Similarly, some of them may have given replies in the think-aloud protocol that they thought we wanted to hear, rather than answering completely truthfully. Furthermore, our think-aloud protocol is not necessarily a realistic representation of the actual leadership process. We did not incorporate a reward structure in our study, nor did we give participants the option to collect additional information or discuss the issue with stakeholders or other leaders. Lastly, our assessment of the responsible leadership roles was based on scales which have not been validated in previous empirical studies. As such, I cannot say with absolute certainty that they are a suitable measuring instrument.

In conclusion, I do find some support for the theoretical notion that personality traits affect responsible leadership effectiveness, although the results of my empirical regression analysis indicate that this interaction is perhaps less strong and far-reaching than I had initially assumed. My results suggest that empathic concern and openness to experience influence the role of facilitator and that holism influences the role of citizen. I believe that it is worth following up on these results and would therefore welcome it if further research included these three personality dimensions to see whether my results can be corroborated. Likewise, it would be valuable to see if the influence of gender, psychology stu-

dent, and perspective taking on the role of idea-provider can be replicated in subsequent studies, or if it is confined to my empirical regression analysis. Having additional research use our scales for measuring the five responsible leadership roles would also give us more opportunities to assess their validity and improve upon them if necessary.

In addition to building on top of my previously discussed results, it would also be great to see future research circumvent some of the limitations that inhibit the explanatory power of our empirical study. In particular, constructing a more realistic and in-depth representation of the leadership process with an actual reward structure could help to improve the validity of the obtained results and potentially give greater insights into aspects that have hitherto been neglected. In a similar vein, including situational and behavioural factors in addition to personality traits would grant a more comprehensive picture of responsible leadership, although the complexity of the subject and the huge number of potentially influential variables would make constructing such a study a challenging task.

In any case, I believe that undertaking further research on the relationship between personality traits and responsible leadership effectiveness is an important and worthwhile endeavour. There is little reason to assume that the degree of globalisation will decline in the near future. On the contrary, if past trends are any indication, the business environment of multinational companies will become even more complex and interwoven in the coming years. As such, I expect that responsible leadership will continue to play an important role and that understanding its antecedents will be of great value to both companies, who are looking for this type of leader to fill their top positions, as well as society as a whole, which benefits from having CEOs whose motivation does not solely lie in maximising profits and who understand the importance of considering stakeholders in their decision-making process.

References

Adams, K. A. and Lawrence, E. K. *Research methods, statistics, and applications*. Sage Publications, 2014.

Alin, A. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):370–374, 2010.

Archer, J. Testosterone and human aggression: an evaluation of the challenge hypothesis. *Neuroscience & Biobehavioral Reviews*, 30(3):319–345, 2006.

Ashton, M. C. and Lee, K. Empirical, theoretical, and practical advantages of the hexaco model of personality structure. *Personality and social psychology review*, 11(2):150–166, 2007.

Ashton, M. C. and Lee, K. The hexaco–60: A short measure of the major dimensions of personality. *Journal of personality assessment*, 91(4):340–345, 2009.

Ashton, M. C., Lee, K., Perugini, M., Szarota, P, De Vries, R. E., Di Blas, L., Boies, K., and De Raad, B. A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of personality and social psychology*, 86(2):356–366, 2004.

Averill, J. R. Individual differences in emotional creativity: Structure and correlates. *Journal of personality*, 67(2):331–371, 1999.

Baer, J. and Kaufman, J. C. Gender differences in creativity. *The Journal of Creative Behavior*, 42(2):75–105, 2008.

Baer, M. and Oldham, G. R. The curvilinear relation between experienced creative time pressure and creativity: moderating effects of openness to experience and support for creativity. *Journal of Applied Psychology*, 91 (4):963–970, 2006.

Baluch, B., Martin, G. N., Christian, L., and Corulla, W. J. Psychology and non-psychology students' estimation of their desirable and undesirable personality traits. *Personality and individual differences*, 21(4):617–620, 1996.

Barbuto, J. E., Fritz, S. M., Matkin, G. S., and Marx, D. B. Effects of gender, education, and age upon leaders' use of influence tactics and full range leadership behaviors. *Sex Roles*, 56(1-2):71–83, 2007.

Batson, D. Empathic concern and altruism in humans, 2009. URL http://nationalhumanitiescenter.org/on-the-human/2009/10/empathic-concern-and-altruism-in-humans/. Retrieved from.

Beck, R. and Harter, J. Why good managers are so rare. *Harvard Business Review*, 2014.

Bowers, K. S. Sex and susceptibility as moderator variables in the relationship of creativity and hypnotic susceptibility. *Journal of Abnormal Psychology*, 78(1):93–100, 1971.

Breusch, T. S. and Pagan, A. R. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, pages 1287–1294, 1979.

Bundy, J., Shropshire, C., and Buchholtz, A. K. Strategic cognition and issue salience: Toward an explanation of firm responsiveness to stakeholder concerns. *Academy of Management Review*, 38(3):352–376, 2013.

Cameron, K. Responsible leadership as virtuous leadership. In *Responsible Leadership*, pages 25–35. Springer, 2011.

Carli, L. L. and Eagly, A. H. Gender, hierarchy, and leadership: An introduction. *Journal of Social issues*, 57(4):629–636, 2001.

Carson, T. Friedman's theory of corporate social responsibility. *Business and Professional Ethics Journal*, 12(1):3–32, 1993.

Choi, I., Koo, M., and Choi, J. A. Individual differences in analytic versus holistic thinking. *Personality and Social Psychology Bulletin*, 33(5):691–705, 2007.

Coan, R. W. Measurable components of openness to experience. *Journal of Consulting and Clinical Psychology*, 39(2):346, 1972.

Coates, J. M., Gurnell, M., and Rustichini, A. Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proceedings of the National Academy of Sciences*, 106:623–628, 2009.

Dabbs Jr, J. M. and Morris, R. Testosterone, social class, and antisocial behavior in a sample of 4,462 men. *Psychological Science*, 1(3):209–211, 1990.

Davis, M. H. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113–126, 1983.

Day, D. V. and Antonakis, J. Leadership: Past, present, and future. *The nature of leadership*, 2:3–25, 2012.

Derue, D. S., Nahrgang, J. D., Wellman, N., and Humphrey, S. E. Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. *Personnel psychology*, 64(1):7–52, 2011.

Dimitrijević, A., Hanak, N., and Milojević, S. Psychological characteristics of future helping professionals: Empathy and attachment of psychology students. *Psihologija*, 44(2):97–115, 2011.

Do, C. B. and Batzoglou, S. What is the expectation maximization algorithm? *Nature biotechnology*, 26(8):897–899, 2008.

Dudek, S. Z., Strobel, M., and Runco, M. A. Cumulative and proximal influences on the social environment and children's creative potential. *The Journal of Genetic Psychology*, 154(4):487–499, 1993.

Durbin, J. and Watson, G. S. Testing for serial correlation in least squares regression. ii. *Biometrika*, 38:159–178, 1951.

Eagly, A. H. and Johnson, B. T. Gender and leadership style: A meta-analysis. *Psychological bulletin*, 108(2):233–256, 1990.

Einolf, C. J. Empathic concern and prosocial behaviors: A test of experimental results using survey data. *Social Science Research*, 37(4):1267–1279, 2008.

Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M., and Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature*, 463:356–359, 2010.

Eisenegger, C., Haushofer, J., and Fehr, E. The role of testosterone in social interaction. *Trends in cognitive sciences*, 15(6):263–271, 2011.

Elmuti, D., Minnis, W., and Abebe, M. Does education have a role in developing leadership skills? *Management Decision*, 43(7/8):1018–1031, 2005.

Fernando, M. *Leading Responsibly in the Asian Century*. Springer, 2016.

Figge, L. and Martens, P. Globalisation continues: The maastricht globalisation index revisited and updated. *Globalizations*, 11(6):875–893, 2014.

Finkelstein, S. and Boyd, B. K. How much does the ceo matter? the role of managerial discretion in the setting of ceo compensation. *Academy of Management journal*, 41(2):179–199, 1998.

Freeman, J. Towards a definition of holism. *Br J Gen Pract*, 55(511):154–155, 2005.

Freeman, R. E. and Auster, E. R. Values, authenticity, and responsible leadership. In *Responsible Leadership*, pages 15–23. Springer, 2011.

Furnham, A., Hughes, D. J., and Marshall, E. Creativity, ocd, narcissism and the big five. *Thinking Skills and Creativity*, 10:91–98, 2013.

Galinsky, A. D. and Moskowitz, G. B. Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, 78(4):708–724, 2000.

Galinsky, A. D., Maddux, W. W., Gilin, D., and White, J. B. Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological science*, 19(4):378–384, 2008.

Gault, B. A. and Sabini, J. The roles of empathy, anger, and gender in predicting attitudes toward punitive, reparative, and preventative public policies. *Cognition & Emotion*, 14(4):495–520, 2000.

George, J. M. and Zhou, J. When openness to experience and conscientiousness are related to creative behavior: an interactional approach. *Journal of applied psychology*, 86(3):513–524, 2001.

Graziano, W. G. and Eisenberg, N. Agreeableness: A dimension of personality. In *Handbook of personality psychology*, pages 795–824. Elsevier, 1997.

Graziano, W. G., Jensen-Campbell, L. A., and Hair, E. C. Perceiving interpersonal conflict and reacting to it: the case for agreeableness. *Journal of personality and social psychology*, 70(4):820–835, 1996.

Graziano, W. G., Habashi, M. M., Sheese, B. E., and Tobin, R. M. Agreeableness, empathy, and helping: A person× situation perspective. *Journal of personality and social psychology*, 93(4):583–599, 2007.

Harrell, F. E., Lee, K. L., and Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387, 1996.

Harris, J. A. Measured intelligence, achievement, openness to experience, and creativity. *Personality and individual differences*, 36(4):913–929, 2004.

Hines, M. Prenatal testosterone and gender-related behaviour. *European Journal of Endocrinology*, 155(suppl 1):115–121, 2006.

Hoffman, B. J., Woehr, D. J., Maldagen-Youngjohn, R., and Lyons, B. D. Great man or great myth? a quantitative review of the relationship between individual differences and leader effectiveness. *Journal of Occupational and Organizational Psychology*, 84(2):347–381, 2011.

House, R. J. and Aditya, R. N. The social scientific study of leadership: Quo vadis? *Journal of management*, 23(3):409–473, 1997.

Howard, P. J. and Howard, J. M. The big five quickstart: An introduction to the five-factor model of personality for human resource professionals. *Charlotte, NC: Center for Applied Cognitive Studies*, 1995.

Jago, A. G. Leadership: Perspectives in theory and research. *Management science*, 28(3):315–336, 1982.

Jensen-Campbell, L. A. and Graziano, W. G. Agreeableness as a moderator of interpersonal conflict. *Journal of personality*, 69(2):323–362, 2001.

Judge, T. A., Bono, J. E., Ilies, R., and Gerhardt, M. W. Personality and leadership: a qualitative and quantitative review. *Journal of applied psychology*, 87(4):765–780, 2002.

Judge, T. A., Colbert, A. E., and Ilies, R. Intelligence and leadership: a quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, 89(3):542–552, 2004.

Kamm, F. M. Harming some to save others. *Philosophical Studies*, 57(3): 227–260, 1989.

Korac-Kakabadse, A., Korac-Kakabadse, N., and Myers, A. Demographics and leadership philosophy: exploring gender differences. *Journal of Management Development*, 17(5):351–388, 1998.

Laplume, A. O., Sonpar, K., and Litz, R. A. Stakeholder theory: Reviewing a theory that moves us. *Journal of management*, 34(6):1152–1189, 2008.

Lechuga, J., Santos, B. M., Garza-Caballero, A. A., and Villarreal, R. Holistic reasoning on the other side of the world: Validation of the analysis-holism scale in mexicans. *Cultural Diversity and Ethnic Minority Psychology*, 17 (3):325–330, 2011.

Lee, K. and Ashton, M. C. Scale descriptions, n.d. URL `http://hexaco.o rg/scaledescriptions`. Retrieved from.

Lee, K. and Ashton, M. C. Psychopathy, machiavellianism, and narcissism in the five-factor model and the hexaco model of personality structure. *Personality and Individual differences*, 38(7):1571–1582, 2005.

Leong, L. Y. C., Fischer, R., and McClure, J. Are nature lovers more innovative? the relationship between connectedness with nature and cognitive styles. *Journal of Environmental Psychology*, 40:57–63, 2014.

LePine, J. A., Colquitt, J. A., and Erez, A. Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel psychology*, 53(3):563–593, 2000.

Little, R. J. A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, 83(404): 1198–1202, 1988.

Lowry, R. Kappa as a measure of concordance in categorical sorting, n.d. URL `http://vassarstats.net/kappa.html`. Retrieved from.

Maak, T. Responsible leadership, stakeholder engagement, and the emergence of social capital. *Journal of Business Ethics*, 74(4):329–343, 2007.

Maak, T. and Pless, N. M. Responsible leadership in a stakeholder society–a relational perspective. *Journal of business ethics*, 66(1):99–115, 2006.

Maak, T., Pless, N. M., and Voegtlin, C. Business statesman or shareholder advocate? ceo responsible leadership styles and the micro-foundations of political csr. *Journal of Management Studies*, 53(3):463–493, 2016.

Magalhães, E., Costa, P., and Costa, M. J. Empathy of medical students and personality: evidence from the five-factor model. *Medical teacher*, 34 (10):807–812, 2012.

Mandell, B. and Pherwani, S. Relationship between emotional intelligence and transformational leadership style: A gender comparison. *Journal of business and psychology*, 17(3):387–404, 2003.

Mazur, A. and Booth, A. Testosterone and dominance in men. *Behavioral and brain sciences*, 21(3):353–363, 1998.

McCormack, L. and Mellor, D. The role of personality in leadership: An application of the five-factor model in the australian military. *Military Psychology*, 14(3):179–197, 2002.

McCrae, R. R. Creativity, divergent thinking, and openness to experience. *Journal of personality and social psychology*, 52(6):1258–1265, 1987.

McDonald, J. D. Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, 1(1):1–19, 2008.

McHugh, M. L. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, 22(3):276–282, 2012.

Miska, C., Hilbe, C., and Mayer, S. Reconciling different views on responsible leadership: A rationality-based approach. *Journal of Business Ethics*, 125 (2):349–360, 2014.

Mitchell, R. K., Agle, B. R., and Wood, D. J. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of management review*, 22(4):853–886, 1997.

Neave, N., Laing, S., Fink, B., and Manning, J. T. Second to fourth digit ratio, testosterone and perceived male dominance. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1529):2167–2172, 2003.

Niezink, L. W., Siero, F. W., Dijkstra, P., Buunk, A. P., and Barelds, D. P. Empathic concern: Distinguishing between tenderness and sympathy. *Motivation and emotion*, 36(4):544–549, 2012.

Oshagbemi, T. Age influences on the leadership styles and behaviour of managers. *Employee Relations*, 26(1):14–29, 2004.

Oswald, P. A. The interactive effects of affective demeanor, cognitive processes, and perspective-taking focus on helping behavior. *The Journal of social psychology*, 142(1):120–132, 2002.

Page, K. M. and Nowak, M. A. Empathy leads to fairness. *Bulletin of mathematical biology*, 64(6):1101–1116, 2002.

Palazzo, G. and Scherer, A. G. Corporate legitimacy as deliberation: A communicative framework. *Journal of business ethics*, 66(1):71–88, 2006.

Patzer, M. *Führung und ihre Verantwortung unter den Bedingungen der Globalisierung*. Berlin, Germany: Patzer Verlag, 2009.

Paulus, C. Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index, 2009. URL `http://psydok.sulb. uni-saarland.de/volltexte/2009/2363/`. Retrieved from.

Pelletier, F. J. Holism and compositionality. *The Oxford handbook of compositionality*, pages 149–175, 2012.

Perciavalle, V., Di Corrado, D., Petralia, M. C., Gurrisi, L., Massimino, S., and Coco, M. The second-to-fourth digit ratio correlates with aggressive behavior in professional soccer players. *Molecular medicine reports*, 7(6): 1733–1738, 2013.

Pless, N. M. Understanding responsible leadership: Role identity and motivational drivers. *Journal of Business Ethics*, 74(4):437–456, 2007.

Pless, N. M. and Maak, T. Responsible leadership: Pathways to the future. *Journal of Business Ethics*, 98:3–13, 2011.

Popma, A., Vermeiren, R., Geluk, C. A., Rinne, T., van den Brink, W., Knol, D. L., Jansen, L. M., Van Engeland, H., and Doreleijers, T. A. Cortisol moderates the relationship between testosterone and aggression in delinquent male adolescents. *Biological psychiatry*, 61(3):405–411, 2007.

Porter, M. E. and Kramer, M. Strategy and society: The link between competitive advantage and corporate social responsibility. *Harvard business review*, 84(12):78–92, 2006.

Pringle, C. D., DuBose, P. B., and Yankey, M. D. Personality characteristics and choice of academic major: Are traditional stereotypes obsolete? *College Student Journal*, 44:131–142, 2010.

Raskin, R. and Terry, H. A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of personality and social psychology*, 54(5):890–902, 1988.

Ridgeway, C. L. Gender, status, and leadership. *Journal of Social issues*, 57 (4):637–655, 2001.

Russo, A. and Perrini, F. Investigating stakeholder theory and social capital: Csr in large firms and smes. *Journal of Business ethics*, 91(2):207–221, 2010.

Sapienza, P., Zingales, L., and Maestripieri, D. Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences*, 106:15268–15273, 2009.

Schroeder, D. A., Dovidio, J. F., Sibicky, M. E., Matthews, L. L., and Allen, J. L. Empathic concern and helping behavior: Egoism or altruism? *Journal of Experimental Social Psychology*, 24(4):333–353, 1988.

Silvia, P. J., Beaty, R. E., Nusbaum, E. C., Eddington, K. M., Levin-Aspenson, H., and Kwapil, T. R. Everyday creativity in daily life: An experience-sampling study of "little c" creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2):183–188, 2014.

Stanton, S. J., Liening, S. H., and Schultheiss, O. C. Testosterone is positively associated with risk taking in the iowa gambling task. *Hormones and behavior*, 59(2):252–256, 2011.

Stogdill, R. M. Personal factors associated with leadership: A survey of the literature. *The Journal of psychology*, 25(1):35–71, 1948.

Tavakol, M. and Dennick, R. Making sense of cronbach's alpha. *International journal of medical education*, 2:53–55, 2011.

Toussaint, L. and Webb, J. R. Gender differences in the relationship between empathy and forgiveness. *The Journal of social psychology*, 145(6):673–685, 2005.

Van Bokhoven, I., Van Goozen, S. H., Van Engeland, H., Schaal, B., Arse-
    neault, L., Séguin, J. R., Assaad, J.-M., Nagin, D. S., Vitaro, F., and Trem-
    blay, R. E. Salivary testosterone and aggression, delinquency, and social
    dominance in a population-based longitudinal study of adolescent males.
    Hormones and Behavior, 50(1):118–125, 2006.

VanVoorhis, C. W. and Morgan, B. L. Understanding power and rules of
    thumb for determining sample sizes. *Tutorials in Quantitative Methods
    for Psychology*, 3(2):43–50, 2007.

Vedel, A. The big five and tertiary academic performance: A systematic
    review and meta-analysis. *Personality and Individual Differences*, 71:66–
    76, 2014.

Viera, A. J., Garrett, J. M., et al. Understanding interobserver agreement:
    the kappa statistic. *Fam Med*, 37(5):360–363, 2005.

Voegtlin, C. and Schwab, P. Good professional, colleague and citizen at the
    same time? an empirical examination of antecedents and outcomes of
    a role model of responsible leadership. *Unpublished manuscript, Depart-
    ment of Business Administration, University of Zurich, Switzerland*, 2015.

Voegtlin, C. Development of a scale measuring discursive responsible lead-
    ership. In *Responsible Leadership*, pages 57–73. Springer, 2011.

Voegtlin, C., Patzer, M., and Scherer, A. G. Responsible leadership in global
    business: A new approach to leadership and its multi-level outcomes.
    Journal of Business Ethics, 105(1):1–16, 2012.

Vroom, V. H. and Jago, A. G. The role of the situation in leadership. *American
    psychologist*, 62(1):17–24, 2007.

Waldman, D. A. Moving forward with the concept of responsible leadership:
    Three caveats to guide theory and research. *Journal of Business Ethics*,
    pages 75–83, 2011.

Waldman, D. A. and Galvin, B. M. Alternative perspectives of responsible
    leadership. *Organizational Dynamics*, 37(4):327–341, 2008.

Waldman, D. A. and Siegel, D. Defining the socially responsible leader. *The
    Leadership Quarterly*, 19(1):117–131, 2008.

White, H. A heteroskedasticity-consistent covariance matrix estimator and a
    direct test for heteroskedasticity. *Econometrica: Journal of the Econometric
    Society*, pages 817–838, 1980.

Winston, B. E. and Patterson, K. An integrative definition of leadership.
    International journal of leadership studies, 1(2):6–66, 2006.

Yukl, G. A. Leadership in organizations (5th ed.). *Upper Saddle River, NJ:
    Prentice Hall*, 2002.

Yukl, G. Managerial leadership: A review of theory and research. *Journal of
    management*, 15(2):251–289, 1989.

Zaccaro, S. J. Trait-based perspectives of leadership. *American psychologist*,
    62(1):6–16, 2007.

Zaccaro, S. J., Kemp, C., and Bader, P. Leader traits and attributes. *The
    nature of leadership*, pages 101–124, 2004.