NYENRODE BUSINESS UNIVERSITEIT

Public acceptance of AI-based detection of social security fraud

Executive Modular MBA

L. Vinke MSc

Nyenrode Business University

Straatweg 25, Breukelen

Core teacher: Prof. Dr. S. Nandram

Academic supervisor(s): Prof. Dr. R.J.M. Jeurissen and Dr. R. Zaal

Practice supervisor: J. Snijder (Head of Research Info Support)

Info Support

Kruisboog 42, Veenendaal

May 15, 2023

Preface

In recent years, I have worked on various IT projects within Info Support for our customers. One project was the development of a product to predict the unlawful use of social assistance benefits. We used an artificial intelligence system for this project – one with high impact in the society. The development of this system raised many questions, including the issue of whether the use of such a system is fair. In the media, journalists have published several articles about this system, often with a negative connotation. On the contrary, measurements have shown that the application of this system makes the detection of the unlawful use of social assistance benefits easier and faster. Weighing all the advantages and disadvantages, we ultimately decided not to further develop this system. The real result is that fraud is less easy to detect, with various consequences.

Sometimes I hear the argument that algorithms are less morally acceptable. However, does this case truly hold? Hence, my personal drive is to provide insight into how individual moral convictions influence the moral acceptability of algorithms. I intend to use this insight to improve the conversation/debate about artificial intelligence in a nuanced manner. I hope such learning contributes to the increased adoption of artificial intelligence in relation to predicting fraud not only in this specific domain but also in a broader sense.

I would like to personally thank several individuals and organizations that were part of my academic journey. First, I am thankful to Info Support as an employer, which offered me every opportunity to pursue this study. I am also grateful to Joop Snijder for his role as practice supervisor. The joint drive for artificial intelligence in a trustworthy manner spurred many excellent conversations. At Nyenrode, I met many people, including fellow students, teachers, and administrators of the program. In retrospect, I sincerely appreciate all the valuable conversations, learning moments, and helpful guidance. Michael Hermes and Jan Willem Torken were my buddies during the process; we discussed the progress on numerous Tuesday and Wednesday evenings, for which I am truly thankful. Ronald Jeurissen and Raymond Zaal supervised me during the thesis. For hours we ardently talked about the conceptual model and statistical models, and I received critical constructive feedback in the process. I learned so much from that experience. I would also like to thank my family who followed my studies with interest and my parents who encouraged me to study.

Finally, I would like to express my profound gratitude to my wife Louise and my son Laurens. I worked on the study for countless days, evenings, and Saturdays. Even during the holidays, I was regularly engrossed in books and papers that I had to read for the study. Without their support and time, I would not have succeeded in completing this study.

Elburg
May 2023
Lammert Vinke

Management summary

Data are increasingly growing, creating new possibilities with the use of data (Floridi, 2012). Artificial intelligence is one of these possibilities. Spector (2016) defines artificial intelligence as "the design and study of computer programs that behave intelligently" (p. 1251). Machine learning is one of the streams within artificial intelligence. Mittelstadt et al. (2015) describe machine learning as "any methodology and set of techniques that can employ data to come up with novel patterns and knowledge and generate models that can be used for effective predictions about data" (p. 3). One of the application areas is the detection of the unlawful use of social services (Ministerie van Sociale Zaken en Werkgelegenheid, 2017). Centraal Bureau voor de Statistiek (2023b) reports that there are currently about 95,000 claims outstanding against citizens who made an unlawful use of social assistance benefits with a total value of €346 million. Machine learning can be used for detecting this unlawful use.

This thesis aims to answer the following research question: **To what extent do moral foundations influence the moral acceptability of algorithms in the social assistance benefits domain in the Netherlands?**

The first part of the research question concerns moral foundations. Individuals base moral decisions on moral foundations (Telkamp & Anderson, 2022). Graham et al. (2012) indicate that all individuals are different based on their moral foundations. Moral foundations can be conceptualized as five scales: care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and purity/degradation. Through a questionnaire, the moral foundations of an individual can be determined based on these five scales. Graham et al. (2012) suggest that composite concepts also exist, namely grouping and individualizing. The grouping concept consists of the care and fairness moral foundation. The individualizing concept comprises the notions of loyalty, authority, and purity. The individualizing and grouping concepts are used in the current research.

The second part of the research question is moral acceptability. Based on the description of Haidt and Kesebir (2010), the present research uses the definition that something is morally acceptable if, according to an individual, it meets his values, rights, and/or the interests of other people. A moral decision follows from a process that Haidt and Kesebir (2010) also refer to as moral reasoning.

In connection with the relationship between moral foundations and moral acceptability, Telkamp and Anderson (2022) assert that in the literature, moral foundations influence moral acceptability. Other components also influence this moral acceptability. Specifically for algorithms, Kodapannakal et al. (2020) show that the data (and thus data protection) used are a determinant of the moral acceptability of algorithms. Büchi et al. (2022) indicate, for example, that the use of social media data creates negative effects. Martin and Waldman (2022) underscore that the outcomes that an algorithm produces strongly determine the moral acceptability of the algorithm. One of the possible negative effects is discrimination (Mittelstadt et al., 2015; Barocas & Selbst, 2016). As Green (2020) and Chouldechova (2017) suggest, fairness metrics can be used for reducing the possible negative effects of discrimination, but this approach leads to a less effective algorithm.

To examine the relationship between moral foundations and moral acceptability, quantitative and deductive research is undertaken in this thesis. The research shows (n = 1,118) a negative relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms in the social assistance domain. In addition, a positive relationship is found between moral foundations in terms of grouping and the moral acceptability of algorithms in the welfare domain. More specifically, the use of public sources (including social media) apparently creates a stronger negative relationship between moral foundations in terms of individualization and the moral acceptability of algorithms. No significant moderating effect of using public data sources is found on the relationship between moral foundations in terms of grouping.

In relation to the fairness metrics, moderating effects seemingly occur when applying a fairness metric. The application of a fairness metric creates a weaker negative relationship between the moral foundations of individualization and the moral acceptability of algorithms. The application of a fairness metric produces a weaker positive relationship between moral foundations in terms of grouping and the moral acceptability of algorithms.

Contents

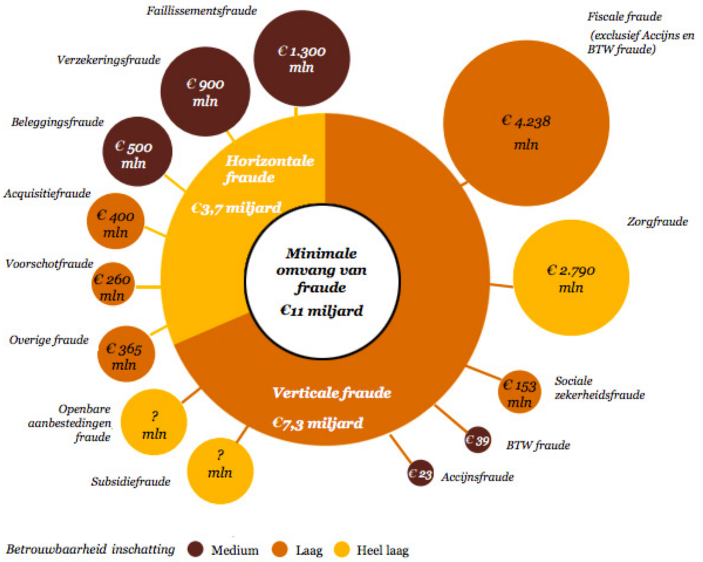
1 Introduction

1.1 Context

This thesis is about the public acceptance of the use of AI-based detection of social security fraud in the Netherlands. Fraud is a deliberate deception to gain an undue advantage. Fraud costs society a substantial amount of money. Research by PWC (2013) has shown that the total amount of fraud in the Netherlands was estimated at €11 billion annually in 2013 (see Figure 1). Furthermore, the size of social security fraud is reported to be €153 million annually.

Figure 1

Impact of Fraud in the Netherlands in 2013 (PWC, 2013)



In the Netherlands, citizens can use social assistance benefits if they do not have sufficient money themselves. The surveys conducted by the Centraal Bureau voor de Statistiek (2023a) have shown that approximately 400,000 Dutch people use social assistance benefits each year. The figures from the Centraal Bureau voor de Statistiek (2023b) also reveal approximately 95,000 outstanding claims with a total balance of €346 million from citizens who have not lawfully used this social assistance. The municipality is responsible for ensuring that citizens lawfully use this assistance. The main objective of enforcement within the social security system is to maintain support for the social security system in society (Ministerie van Sociale Zaken en Werkgelegenheid, 2017). The government undertakes this step by focusing on three themes: prevention, guaranteeing that citizens lawfully use social assistance benefits, and punishment.

Data integration is one of the instruments that the government uses for checking whether citizens are lawfully using social assistance (Ministerie van Sociale Zaken en Werkgelegenheid, 2017). With data integration, the government combines different sources to detect the unlawful use of social assistance benefits more efficiently and effectively. Nederlands Forensisch Instituut (2017) suggests the possibility that the effectiveness of data integration is high. This research shows that when a data integration solution constructs a list with the 1,000 citizens who have the highest risk of fraud, 800 fraudsters are found to commit fraud.

Aside from advantages, the use of data integration also has disadvantages. One drawback is the impact on the privacy of citizens who use social assistance benefits. The Data Protection Authority (College Bescherming Persoonsgegevens, 2010) has made several recommendations about the preconditions under which these algorithms may be applied. In addition, in the Netherlands a court has ruled (Rechtbank Den Haag, 2020) that combining these sources is an excessive infringement for the individual if manifold sources are combined.

The debate that transpires specifically in the social security services domain in the Netherlands is also described in more abstract terms in scientific articles. A frequently cited scientific article by Barocas (2016) indicates that proponents of the application of algorithms contend that the application of algorithms removes human bias in decision-making. However, as algorithms are data-based, they maintain the bias that is already in the data.

To avoid the negative effects (externalities) of algorithms, public and commercial parties define different principles to which these algorithms must conform. Jobin et al. (2019) describe 84 documents with principles that are defined by multiple public and private parties. In their research, Jobin et al. (2019) underscore that the most important ethical principles are transparency, justice and fairness, non-maleficence, responsibility, and privacy. The research by Floridi (2019) is similar to the work of Jobin et al. (2019). Nonetheless, these ethical principles still raise certain questions, including the following: "What is a fair algorithm in this context?"

The principles as defined by public parties also lead to legal frameworks such as the General Data Protection Regulation (hereafter GDPR) and the upcoming Artificial Intelligence Act. Hoofnagle et al. (2019) describe, among other matters, that profiling algorithms fall into the highest risk category of the GDPR. Furthermore, the new European AI Act classifies profiling algorithms in the highest risk category (see European Commission, 2022). Both laws impose additional requirements on these profiling algorithms, including that AI algorithms are prohibited from making automated decisions about people.

In summary, AI can be used for predicting fraud in the social assistance benefits domain. The public debate includes both proponents and opponents of the application of profiling algorithms. Several parties have defined principles to prevent the negative consequences of algorithms.

1.2 Problem statement

Floridi (2019) and Telkamp and Anderson (2022) argue that the principles and legal frameworks for the application of algorithms have several deficiencies. For example, the principles inventoried by Jobin et al. (2019) and Floridi (2019) mention justice and fairness as frequently used principles. However, the principle of fairness does not solve the impossibility of fairness, as outlined by Chouldechova (2017), Green (2020), and Ruf (2021). The impossibility of fairness signifies that "any effort to improve decision-making using algorithms will violate at least one normatively desirable fairness principle"

(Green, 2020, p. 3). Additionally, individuals think differently about these ethical principles for AI. Telkamp and Anderson (2022) contend that the moral views of individuals are often disregarded when making decisions about AI. This factor regularly leads to AI solutions that are morally not, or differentially, acceptable to people while complying with all the principles acknowledged in the literature and legal frameworks. In the context of liberal ethical theories, in which the validity of moral judgments is considered wholly dependent on their congruence with an individual's moral conscience (Kant's principle of ethical autonomy), a neglect of the moral views of individuals is highly problematic. Any decision-making algorithm that suffers from this neglect will likely produce outcomes without ethical legitimacy.

Telkamp and Anderson (2022) underscore the need for additional empirical research into the relationship between moral foundations and the moral acceptability of algorithms. Moral foundations can be conceptualized as five axes on which individuals base their moral judgments (Graham et al., 2012). These five axes are care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and purity/degradation. Graham (2006) suggests that individuals differ on these moral foundations. A different moral foundation may therefore result in a different moral judgment (and thus the moral acceptability) of algorithms. At present, the relationship between these moral foundations and moral acceptability has yet to be investigated.

1.3 Research goal and scope

The aim of the research is to obtain knowledge of and insight into how moral foundations influence the moral acceptability of algorithms in the social assistance benefits domain.

According to Martin and Waldman (2022) and Araujo et al. (2020), individuals judge algorithms differently in different contexts. Therefore, the current research is limited to a specific context. The scope is limited to the application of algorithms in the social assistance benefits domain in the Netherlands. The reason for selecting this domain is threefold: this study concerns the application of algorithms in a domain that the media has criticized, it relates to a vulnerable target group in the data, and the researcher's practical experience with this domain.

1.4 Research question

Based on the problem statement and the goal, the following research question is addressed: To what extent do moral foundations influence the moral acceptability of algorithms in the social assistance benefits domain in the Netherlands?

To answer this question, the researcher conducts a quantitative and deductive research. A quantitative study is undertaken because the researcher intends to gain insight into the influence of moral foundations on the moral acceptability of algorithms within the social assistance benefits domain in the Netherlands. A quantitative study is typically chosen to validate a theory through data (Saunders et al., 2019). The present research examines the theory as described by Telkamp and Anderson (2022).

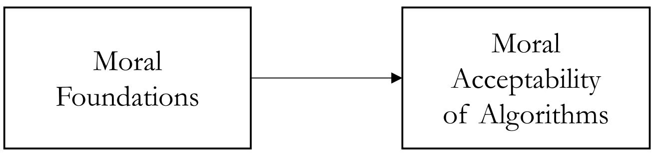
1.5 Conceptual model

Several factors influence the **moral acceptability** of algorithms. For example, Kodapanakkal et al. (2020) indicate that various factors (outcome favorability, data sharing, and data protection) impact the moral acceptability of algorithms. Ditto et al. (2009) describe the process by which individuals judge morally. Kodapanakkal (2020) uses a scale from 0 to 100 (0 is morally unacceptable, and 100 is morally acceptable) for measuring the extent to which individuals regard an algorithm as morally acceptable.

Citizens make moral judgments about the application of algorithms. These moral judgments are based on moral foundations. Graham et al. (2012) defined moral foundations theory. This theory holds that individuals primarily base moral judgments on **moral foundations**. The moral foundations are care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and purity/degradation. Based on an existing questionnaire (see Moral Foundations Theory, 2017), each moral foundation is given a score from 0 to 5. This questionnaire asks respondents to indicate the extent to which they agree with statements that belong to moral foundations based on a Likert scale. Based on the answers in the questionnaire, a score for each moral foundation is calculated. When more moral judgments of an individual comply with moral foundations, an algorithm is more morally acceptable (Telkamp & Anderson, 2022). Graham et al. (2012) conclude that moral foundations differ from individual to individual and underline the presence of cultural differences. Based on the research of Telkamp and Anderson (2022), a conceptual model is developed in the current study (refer to Figure 2).

Figure 2

Conceptual Model



The moral acceptability differs per algorithm and context (Kodapanakkal et al., 2020). Hence, the researcher selected a specific context during this study. Telkamp and Anderson (2022) define different dimensions of these AI algorithms used by the researcher during this study, namely organizational uses of AI, AI data and development, and AI decisions.

The first aspect is the purpose for which an algorithm is used (Telkamp & Anderson, 2022). The purpose of an algorithm determines the outcomes it can generate. When an algorithm aims to predict whether a person has cancer, the outcome is different from an algorithm that predicts whether it will rain. The outcome favorability factor influences moral acceptability (Kodapanakkal, 2020). Kodapanakkal (2020) defines outcome favorability as "how personally beneficial an outcome of a technology is for the person making the decision, irrespective of whether this outcome is unfair to others or not." (p. 2) Similarly, Martin and Waldman (2020) indicate that the outcome of an algorithm is an important determining factor for the legitimacy of the application of algorithms to decisions. For the current study, the researcher chooses one type of purpose, that is, to predict fraud.

The second aspect on which the researcher varies the type of algorithm is the data source that is used by the algorithm. According to Floridi (2012), more and more data are available. Algorithms enable companies and governments to discover correlations in these data. However, the use of specific sources generates negative effects, for example, the usage of social media data (Büchi et al., 2020). In addition, the GDPR prescribes that only data that are necessary to achieve the goal (results) may be stored, which denotes the so-called "data minimization principle" (Hoofnagle et al., 2019). Two types of data sources are used for the present research, namely data of the municipality and data of public sources (social media).

The third aspect on which the researcher varies the type of algorithm is whether a fairness metric is applied. Fairness is a principle that is referred to by many standards and legal frameworks (Jobin, 2019; Floridi & Cowls, 2019). However, Hellmann (2020) suggests that organizations must still make different choices about how to implement fairness in an algorithm. For example, training an algorithm to find the most fraudsters is possible, with the potential result that the number of erroneous predictions in a minority group is higher. Green (2020) and Ruf (2021) also raise the possibility of training an algorithm based on a fairness metric. Equalizing the number of incorrect predictions in both the minority group and the majority group is possible, but it lowers performance in terms of the purpose of the model.

In summary, the four use cases shown in Table 1 are analyzed in this study.

Table 1

Configuration of Use Cases

Use Cases	Purpose	Data Used	Fairness Metrics
Use Case 1	Predict fraud	Data of municipality	No fairness metric applied
Use Case 2	Predict fraud	Data of municipality	Fairness metric applied
Use Case 3	Predict fraud	Data of municipality + public sources (social media)	No fairness metric applied
Use Case 4	Predict fraud	Data of municipality + public sources (social media)	Fairness metric applied

1.6 Practical contribution

By demonstrating how moral foundations influence the moral acceptance of algorithms that differ on purpose, data sources used, and fairness metrics applied, governments know better what to do to increase the moral acceptability of algorithms. This approach creates advantages for both commercial companies and governments that can use these algorithms.

The application of artificial intelligence in the social security services domain results in more efficient and effective enforcement in the social security services domain (Nederlands Forensisch Instituut (2017). Ministerie van Sociale Zaken en Werkgelegenheid (2017) states that such application contributes to greater social support in society for the social security services. By gaining more insight into the influence of moral foundations on the moral acceptability of algorithms, a government will have more opportunities to adopt measures to increase the moral acceptability of algorithms and thereby improve enforcement in the social security services domain. This research focuses on the domain of social security fraud, and the approach can be used as a starting point for further research in other domains.

1.7 Scientific contribution

Various moral foundations have an impact on the moral acceptance of algorithms (Telkamp & Anderson, 2022). Regarding future investigations, Telkamp and Anderson (2022) highlight the need for new empirical research into the relationship between specific AI contexts and specific moral foundations. The objective of the present research is to make this contribution.

In the current literature, the moral principles of AI have been substantially examined. Among other topics, Green (2020), Leben (2020), and Popp Saenz (2022) have explored whether and, if so, the development of fair algorithms is possible. The present research additionally analyzes the impact of applying these fairness metrics on the relationship between moral foundations and the moral acceptability of algorithms.

1.8 Structure of the thesis

To address the main research question, Chapter 2 describes the theoretical framework and answers the first five sub-questions. Based on the theoretical framework, six hypotheses are formulated and tested. This theoretical framework forms the basis for quantitative research. Chapter 3 explains the approach to this quantitative research. Chapter 4 describes the results of the quantitative research conducted and answers the sixth sub-question. Finally, Chapter 5 contains the conclusions and an answer to the main question. It also outlines the limitations and describes the practical and academic relevance of this research.

2 Theoretical framework

This chapter provides the theoretical framework. Section 2.1 explains the meaning of moral foundation and presents its dimensions. Section 2.2 defines moral acceptability. Section 2.3 describes the theoretical framework in relation to algorithms and machine learning. Section 2.4. defines the construct of moral foundations in the context of the application of algorithms. Section 2.5 illustrates the significance of moral acceptability in the context of applying algorithms. Finally, Section 2.6 describes the hypotheses and the conceptual model developed in this research.

2.1 Moral foundations

The question "What are moral foundations, and what are their dimensions?" is initially answered in Section 2.1.1 with a general introduction to morality. Section 2.1.2 describes the development of moral foundations theory within the moral domain. Section 2.1.3. explains the different dimensions of moral foundations. Finally, Section 2.1.4 provides a critique on these moral foundations.

2.1.1 Meaning of morality

Morality is about the "right" and "wrong" way to behave (Haidt & Kesebir (2010). For instance, someone should be fair and not unfair to others. However, the ensuing question relates to how an individual determines what is right and what is wrong. Jeurissen et al. (2007) and Haidt and Kesebir (2010) classify ethics into two forms: deontology and consequentialism. On the one hand, the focus of deontology is on obligations without considering the consequences. On the other hand, the emphasis of consequentialism is on the alternative that leads to the greatest total good. Several more specific forms exist within these main forms of ethics.

Immanuel Kant is a well-known philosopher who adheres to deontology. Several universal laws govern how individuals act (Gregor & Timmerman, 2012). For Kant, "the aim is the identification and corroboration of the supreme principle of morality" (Gregor & Timmerman, 2012, p. 2). This supreme principle of morality is also described by Kant as the categorical imperative. This so-called categorical imperative assumes that each person acts in the way that he would like all other individuals to act towards all the people in the world. More simply stated, something is right when a person can apply the same standard to all individuals in this world (including oneself).

Utilitarian ethics is a well-known form of consequentialism, as described by the British philosopher Jeremey Bentham. Utilitarian ethics is a more quantitative approach (Jeurissen et al., 2007). Individuals calculate the costs and benefits of different alternatives and choose the alternative that has the most benefits and the least costs overall. More simply stated, something is right when it creates the largest positive value.

In summary, morality is about the right and wrong ways to behave. How individuals morally judge depends on their form of ethics as to how they determine the right and wrong ways to behave.

2.1.2 Moral foundations theory

Moral foundations theory, as described by Graham et al. (2011), assumes that all individuals differ based on five moral foundations. These moral foundations are care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and purity/degradation, which are further defined in the subsequent section. Individuals initially base their moral judgments on these moral foundations (Telkamp & Anderson, 2022). Thus, moral foundations influence the way people behave or make decisions. People evaluate the extent to which the outcomes of behaviors or decisions agree with or conflict with these moral foundations. Berg et al. (2022) examine the degree to which individuals with different moral foundations show different behaviors.

Haidt and Joseph (2004; 2007) define moral foundations theory by merging several studies into a new theory. Until the emergence of moral foundations theory, moral domain was mainly limited to the question of whether individuals were harming or treating people unfairly (Graham et al., 2011). Graham et al. (2012, p. 4) ask the following question: "How many basic elements are needed to represent, understand, and explain the breadth of the moral domain?" Furthermore, they identify two schools of thought, namely the monistic movement and the pluralistic movement. Researchers from the monistic movement assume that this school of thought is usually about only one element, often referred to as "justice" or "fairness." Lawrence Kohlberg is a major supporter of this theory. However, several more recent studies have shown that the moral domain is broader than fairness and justice. For example, Haidt and Kesebir (2010) provide a newer definition of moral systems.

> Moral systems are interlocking sets of values, virtues, norms, practices, identities, institutions, technologies and evolved psychological mechanisms that work together to suppress or regulate selfishness and make social life possible. (p. 800)

The foregoing description might suggest that moral pluralists did not exist before Kohlberg. However, Aristotle is already an adherent of the pluralistic approach, as he describes ethical attitudes as a "bag of virtues," including courage, generosity, friendship, and wisdom, in addition to justice (Aristotle et al., 2009).

2.1.3 Moral foundations theory and its dimensions

Haidt and Joseph (2004) distinguish five generic moral values (or the dimensions) within moral foundations theory. These moral values are care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and purity/degradation. Each moral foundation is indicated by a combination of two words. In this thesis and in the literature, only the first word is regularly used. Everyone develops these five moral foundations. Graham et al. (2011) suggest that these moral foundations help to gain insight into and reason about the moral viewpoints of an individual and/or society. As an example, Graham et al. (2011) state that a citizen with a high moral foundation often has a different political preference. Berg et al. (2022) cite another example in which citizens with different moral foundations deal with compliance with the COVID-19 rules in different ways.

The first moral foundation pertains to **care/harm**. Graham et al. (2012) underscore that all people have needs for care. Haidt (2013) cites as an example young children who need care from an older person. When people are in pain, most people dislike it. However, some individuals find this pain worse than others. A case in which virtues that belong to the care moral foundation are valued more highly in one culture than in another similarly illustrates an example. For instance, the level of care in Buddhism is higher than, for example, during the Nazi regime (Graham et al., 2012).

The second moral foundation includes **fairness/cheating**. Terms such as fair, just, and trustworthy belong to this moral foundation (Graham et al., 2012). This moral foundation is about a balance between giving and taking, or reciprocal altruism (Haidt, 2013). A key requisite is that when someone makes an agreement, he also honors it. Graham et al. (2012) argue that this moral foundation can even be about inanimate objects, for example, when one puts a euro in a soda machine and no soda comes out.

The third moral foundation pertains to **loyalty/betrayal**. Various social experiments show that people quickly identify with groups (Haidt, 2013). When identification with a group is present, an individual can stand up for the members of that group. Graham et al. (2012) contend that this same behavior can also be found in chimpanzees, which fight in groups with other groups for more power and ranking. This case is also evident in the current society, for example, among fans of sports clubs or loyalty to a specific brand. One individual finds this moral foundation more important than the other.

The fourth moral foundation comprises **authority/subversion**. The extent to which individuals have respect for parents vastly differs from culture to culture (Haidt, 2013). An example is how one addresses older people in a different way than younger people. According to Graham et al. (2012), chimpanzees also live according to certain hierarchies; for example, the oldest male is the boss. These hierarchies are similarly visible in the present society (e.g., the power that the government has according to one individual and the power that the government has according to another individual).

The fifth moral foundation relates to **purity/degradation**. All people have moral foundations that convey that some matters are sacred, whereas others are disgusting (Haidt, 2013). An extreme example of this case is cannibalism. Another example is how people deal with sexuality. A holy life is more common in culture than in any other culture. For instance, individuals with a high-purity moral foundation view the human body as a temple (Graham et al., 2012).

An overview of the five different moral foundations from Graham et al. (2012) is shown in Figure 3.

Figure 3

Moral Foundations, According to Graham et al. (2012)

Foundation:	Care/ harm	Fairness/ cheating	Loyalty/ betrayal	Authority/ subversion	Sanctity/ degradation
Adaptive challenge	Protect and care for children	Reap benefits of two-way partnerships	Form cohesive coalitions	Forge beneficial relationships within hierarchies	Avoid communicable diseases
Original triggers	Suffering, distress, or neediness expressed by one's child	Cheating, cooperation, deception	Threat or challenge to group	Signs of high and low rank	Waste products, diseased people
Current triggers	Baby seals, cute cartoon characters	Marital fidelity, broken vending machines	Sports teams, nations	Bosses, respected professionals	Immigration, deviant sexuality
Character-istic emotions	Compassion for victim; anger at perpetrator	anger, gratitude, guilt	Group pride, rage at traitors	Respect, fear	Disgust
Relevant virtues	Caring, kindness	Fairness, justice, trustworthiness	Loyalty, patriotism, self-sacrifice	Obedience, deference	Temperance, chastity, piety, cleanliness

A moral foundation must meet the following criteria (Graham et al., 2012, p. 37):

- Common in third-party normative judgments: as soon as an individual makes a moral judgment, this moral foundation is often part of such a judgment.
- Automatic affective evaluations: making a judgment based on this moral foundation is easy for an individual (e.g., "I don't think that is fair").
- Culturally widespread: the moral foundation must universally occur in individuals all over the world, not merely in a specific part of the world.
- Evidence of innate preparedness: this moral foundation occurs worldwide as an innate characteristic; it is similar to collecting water: all societies know how to perform this activity. This feature also applies to a moral foundation.
- Evolutionary model: the moral foundation has been further developed from various other theories.

As soon as a moral foundation meets these five criteria, adding a new moral foundation becomes possible. For instance, Graham et al. (2012) suggest the addition of a sixth moral foundation, namely liberty/oppression. Telkamp and Anderson (2022) also refer to this sixth moral foundation. However, substantiated empirical evidence for this moral foundation is insufficient compared to existing moral foundations. Therefore, this sixth moral foundation is not used in the current study.

The studies by Graham et al. (2012) and Nilsson and Erlandson (2015), among others, also use other concepts such as individualizing and grouping (or binding). The concept of individualizing is a combination of the notions of fairness and care, whereas the concept of grouping is a composite of the notions of loyalty, authority, and purity.

2.1.4 Other theories and critiques on moral foundations theory

The main criticism of moral foundations theory comes from the monistic perspective (Graham et al., 2012). Monists claim that a pluralistic approach does not exist because the monistic approach is the right one. For example, Gray et al. (2012) describe that individuals can make all moral considerations along the axis of whether they hurt someone else. This movement does not distinguish between the different moral foundations, as Graham et al. (2012) do. The current research uses moral foundations theory in the conceptual model and thereby adopts the pluralistic approach.

In a literature study, Ellemers et al. (2019) describe 10 other questionnaires that can be used for mapping abstract moral values. Among these questionnaires, the Moral Foundations Questionnaire is the most cited one. In addition to the Moral Foundations Questionnaire, Berg et al. (2022) also use morality as a cooperation theory; the questions from this questionnaire are comparable with the ones used for measuring moral foundations. In the current study, moral foundations theory is selected because a literature review by Ellemers et al. (2019) shows that the Moral Foundations Questionnaire is the most cited questionnaire.

2.2 Moral acceptability

Answering the question "What is moral acceptability?" involves describing three different perspectives that follow from the literature. Section 2.1.1 provides a general definition of the concept of moral acceptability. Section 2.1.2 explains the different phases related to moral behavior and, more specifically, how these process steps are related to moral acceptability. Finally, Section 2.2.3 details the different ways in which people judge morally.

2.2.1 Definition of moral acceptability

The concept of moral acceptability consists of the words "moral" and "acceptability." According to Haidt and Kesebir (2010), morality is about what is the "right" and "wrong" way of behaving oneself with respect for the values, rights, and interests of other people, for example, whether someone should be fair or dishonest. Section 2.1.1 contains a more detailed description of the construct of morality. With regard to acceptability, Poel (2016) distinguishes between acceptance and acceptability. For Poel (2016), acceptance is primarily a descriptive notion. A descriptive notion pertains to "what is, was or will be the case or is possibly the case" (Poel, 2016, p. 181). By contrast, acceptability is a normative notion, which "roughly refers to what is good or desirable and what ought to do" (Poel, 2016, p. 181). Poel (2016) considers the concepts of acceptance and acceptability as thick concepts. Thick concepts are concepts that are simultaneously normative and descriptive. In this thesis, the researcher uses the definition that something is morally acceptable if, according to an individual, it meets his values, rights, and/or the interests of other people.

2.2.2 Moral acceptability versus moral behavior

Berg et al. (2022) describe three necessary phases that precede concrete moral behavior: moral awareness, moral judgment, and moral intuition. In the moral awareness phase, an individual becomes either aware or not of the fact that a certain moral value plays a role in a decision. After an individual is aware that a moral choice must be made, he or she makes a decision about which action is the right one (without undertaking this action). This step occurs in the moral judgment phase. This phase therefore leads to moral acceptability or not. In the moral intention phase, an individual determines the consequences of a moral intention. For example, the question of whether achieving a personal goal is more important than a moral consideration made in the previous phase should be addressed. These three phases subsequently lead to moral behavior, which involves the actual execution of the moral decision. In this case, an individual failing to follow through with the moral decision is also possible, for example due to situational factors (e.g., lack of resources) or resistance from other individuals. This research specifically focuses on the moral judgment phase.

2.2.3 Moral intuition versus moral reasoning

Certain alternatives exist within the moral judgment phase, for example, moral intuition and moral reasoning (Haidt & Kesebir, 2010; Ellemers et al., 2019). Moral intuition refers to "the sudden appearance in consciousness, or at the fringe of consciousness, of an evaluative feeling (like–dislike, good–bad) about the character or actions of a person, without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion" (Haidt & Kesebir, 2010, p. 802). Haidt and Kesebir (2010) define the process of moral reasoning as a "conscious mental activity that consists of transforming given information about people (and situations) in order to reach a moral judgement."

Haidt and Joseph (2004) describe this process using modules. Suppose a module takes the conduct or character of another person as input. This module provides a signal of rejection or approval, and such signal can be strong or weak. Haidt and Joseph (2004) refer to this process as "moral intuition." The relationship between moral foundations and moral acceptance is examined in the current study; thus, these processes are also included in the investigation.

2.3 Algorithms

To answer the question, "What are algorithms and their dimensions?", Section 2.3.1 presents descriptions of big data and algorithms. Section 2.3.2 more specifically describes the concept of machine learning. Bias is a frequently cited problem, and this topic is clarified in Section 2.3.3. Section 2.3.4 explains the so-called "fairness metrics." Section 2.3.5 illustrates how data are used within these algorithms. Finally, Section 2.3.6 focuses on the principles that various companies and governments use for the development of these algorithms.

2.3.1 Algorithms

One of the factors driving the rapid rise of algorithms is the increase in the amount of data (Floridi, 2012). According to Floridi (2012), these data offer the possibility of searching for unknown patterns based on the data.

The availability of data and algorithms opens up many new application areas. Several examples of the use of artificial intelligence are found in different sectors. For instance, farmers are increasingly using smart farming applications in the agricultural sector (Mohr & Kuhl, 2021). Artificial intelligence is transforming the agricultural sector into an industry that pays more attention to individual plants and animals. This scenario is possible with data and algorithms. Jeuk et al. (2020) illustrate the use of artificial intelligence for predicting delirium in the health sector. Another widely cited application is the COMPAS algorithm, as described by Chouldechova (2017), Popp Saenz (2022), and Helman (2020). The COMPAS algorithm predicts the chance of recidivism for prisoners. The outcome of the algorithm is used for determining whether a prisoner will be released. More specifically, the Ministerie van Sociale Zaken en Werkgelegenheid (2017) also indicates that artificial intelligence presents the possibility of predicting fraud in the social assistance domain. The upshot of these examples is that AI not only uncovers new avenues for data analysis, which have not been previously possible, but also opens completely new avenues for research, monitoring, and policy development.

Different definitions of algorithms are provided in the literature. An algorithm is "a sequence of computational steps that transform inputs into outputs, similar to a recipe" (Martin, 2019, p. 837). In their literature search, Mittelstadt et al. (2015) refer to the definition of algorithm as a "mathematical construct with a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions" (Hill, 2015, p. 47). This thesis uses the latter definition, given the completeness and substantiation provided by Hill (2015).

Other concepts in relation to algorithms are mentioned in the literature, including artificial intelligence, data analytics, machine learning, and big data. Artificial intelligence pertains to "the design and study of computer programs that behave intelligently" (Spector, 2016, p. 1251). Algorithms are used within these computer programs and are therefore part of artificial intelligence. Meanwhile, data analytics is "the practice of using algorithms to make sense of streams of data" (Mittelstadt et al., 2016). Floridi (2012) argues that the term big data is often unclear, but he defines it as "large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future." Machine learning is a specific movement within artificial intelligence, and it is described in more detail in the next section.

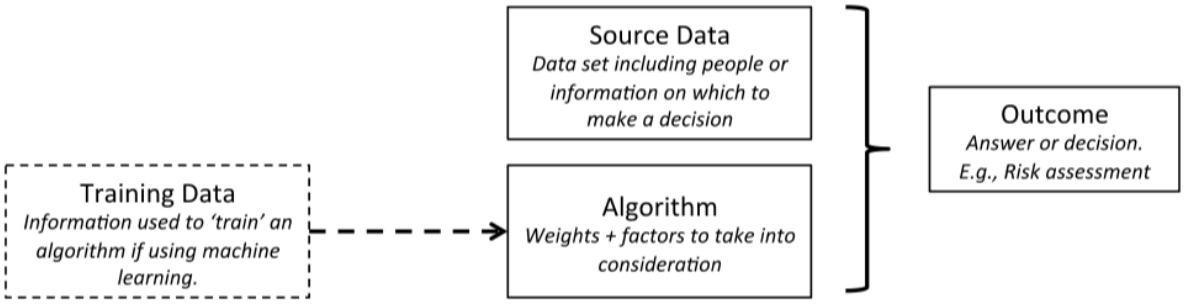
The media publishes multiple articles detailing the advantages and disadvantages of algorithms and big data (Kodapanakkal, 2020). Hill (2015) states that as algorithms are becoming increasingly important, a logical consequence is that citizens ask questions to understand the benefits and drawbacks of these algorithms.

2.3.2 Machine learning

Machine learning is "any methodology and set of techniques that can employ data to come up with novel patterns and knowledge and generate models that can be used for effective predictions about the data" (Mittelstadt et al., 2015, p. 3). Machine learning is a movement within artificial intelligence; hence, not all artificial intelligence applications are machine learning applications. Martin (2019) provides a simple model incorporating the concepts of machine learning, training data, and algorithms (see Figure 4).

Figure 4

From Training Data to Outcome (Martin, 2019)



Two forms of machine learning are found in the discipline of machine learning, namely supervised and unsupervised machine learning. According to Mittelstadt et al. (2016), supervised machine learning is the process by which an algorithm predicts a certain outcome based on labeled input (training data). By contrast, unsupervised machine learning is a process whereby an algorithm distils patterns from the data without these labeled inputs. This thesis focuses on supervised machine learning.

Leben (2020) posits a formal description of supervised machine learning. Consider a given dataset (x, y), where x is a vector of input values and y is the classifier; an example of x could be a vector with information about citizens, and the classifier y could determine whether a citizen has committed fraud. The classifier can be either 0 or 1. When the binary classifier r(X) is trained against the dataset (x, y), the binary classifier r(X) predicts the new value $\hat{y}_i$ based on a new dataset. Leben's (2020) model corresponds to Martin's (2019) model as follows:

- The dataset (x, y) corresponds to the training data in Figure 4.
- The binary classifier r(X) corresponds to the algorithm in Figure 4.
- The new dataset corresponds to the source data in Figure 4.
- The $\hat{y}_i$ corresponds to the predicted outcome in Figure 4.

The predicted outcome $\hat{y}_i$ by r(X) can generate four different outcomes as described by, among others, Ruf(2021), Leben (2020), and Chouldechova (2017), namely true positives, true negatives, false positives, and false negatives. Table 2 provides an overview of these possible outcomes. An example is as follows: when a citizen commits fraud ($y = 1$) and the binary classifier predicts that the citizen also commits fraud ($\hat{y} = 1$), the outcome is a true positive.

Table 2

Outcomes

		Predicted outcome: $\hat{y}_i$	
		1	0
Truth: y	1	True Positive (TP)	False Negative (FN)
	0	False positive (FP)	True Negative (TN)

Many different definitions are used in the machine learning literature. Table 3 presents the formulas of the different concepts, as used by the researcher in the context of this thesis.

Table 3

Formulas

Concept	**Calculation**	**Literature**
Actual Positives: P	P = FN + TP	Ruf (2021)
Actual Negatives: N	N = FP + TN	Ruf (2021)

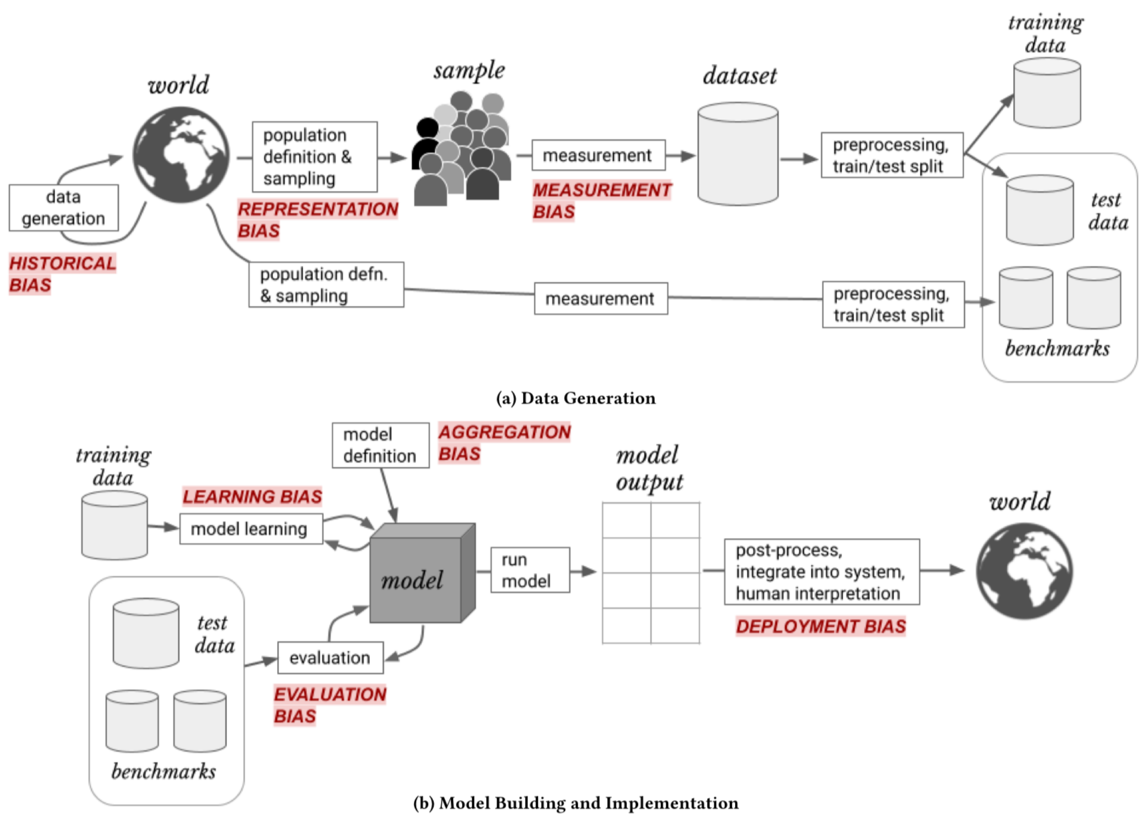
2.3.3 Bias

Bias is a frequently mentioned problem in the application of algorithms. An advantage of using algorithms is that it reduces the number of human biases (Baracos & Selbst, 2016). On the contrary, an algorithm is trained on existing data; thus, biases in the existing data will also affect the outcomes of anything that is based on it. Fahse et al. (2021) define bias as an “unintended or potentially harmful property of data that results in a deviation of algorithmic results.”

Fahse et al. (2021) and Suresh and Guttag (2022) identify various forms of bias. Figure 5 shows a schematic representation of the different forms of bias as defined by Suresh and Guttag (2022). Historical bias is a bias that already exists in society. Fahse et al. (2021) consider this form of bias as social bias. Representation bias is a type of bias in which a limited set of training data is used to train the model, for example, because data are unavailable. Measurement bias arises when labels and/or features are chosen to be used in training a model. Measurement bias occurs when the chosen labels and features are an incomplete or incorrect representation of the concept. Fahse et al. (2021) distinguish between label bias and measurement bias. Label bias is the bias that arises when a data scientist chooses a specific label. Measurement bias is the bias that emerges in existing systems; an example is a score of creditworthiness. Learning bias arises occur when the choice of a specific model increases the inequalities in different cases. An example of a learning bias is a data scientist's optimization of a model for accuracy, which results in a negative consequence for another objective. Fahse et al. (2021) refer to this form of bias as algorithmic bias. Evaluation bias arises when machine learning is evaluated against an unrepresentative group. Aggregation bias is the bias that emerges when a trained model is applied to a different set. Deployment bias is the bias that occurs when the results of the model are used incorrectly. Finally, Fahse et al. (2021) mention the so-called "feedback bias." Feedback bias arises when the outcomes of the existing model train a new model, which leads to the reinforcement of certain outcomes.

Figure 5

Forms of Bias (Suresh and Guttag, 2021)



2.3.4 Fairness metrics

Several fairness metrics can be used for avoiding historical bias. Green (2020) highlights two important concepts in the application of fairness metrics: separation and sufficiency. In this thesis, these concepts are explained using examples. A data scientist uses a dataset (x, y) to train a binary classifier r(X), where x consists of several subgroups. Consider this example: when the dataset contains the data of citizens who receive social assistance benefits, x can consist of men and women, and x can comprise citizens with a different ethnicity. The papers define separation as when r(X) is trained, such that the number of false negatives and the number of false positives are equal in the different groups. With sufficiency, the predicted outcomes are the same for all groups in the dataset.

Green (2020), Chouldechova (2017), Ruf (2021), and Pop Saenz (2022) indicate that satisfying both principles of sufficiency and separation is not possible for an algorithm. This case is denoted as fairness dilemma or the impossibility of fairness. Fairness metrics (i.e., metrics that meet the principle of sufficiency) have different forms. Leben (2020) posits several formal definitions of fairness metrics. Applying a fairness metric produces different consequences depending on the context.

2.3.5 Data and algorithms

Data are needed to make algorithms work. Floridi (2012) states that this data creates many new possibilities and advocates retaining all the data. However, the GDPR requires as little data as possible to be used according to the data minimization principle (Hoofnagle et al., 2019). These two perspectives are at odds with each other.

Data also come in different forms, including personal data and data from social media. For example, personal data that a municipality already has can be processed. Data from social media can be similarly processed. The use of these data has an impact. When using social media sources for investigative purposes, many citizens post less on social media (Büchi et al., 2019). Therefore, they feel less free after investigative authorities use these data.

Martin and Waldman (2022) argue that adding arbitrary data to a machine learning application reduces the legitimacy of these algorithms. Examples of these data are race, ethnicity, and social media use.

2.3.6 Principle-based approach

Many new principles were established between 2015 and 2019 for the application of artificial intelligence (Jobin et al., 2019). Jobin et al. (2019) reviewed 84 documents with principles from private and public parties and indicated that these principles can be reduced to five ethical principles, namely transparency, justice and fairness, non-maleficence, responsibility, and privacy. These principles have been established by both government and private initiatives. Floridi and Cowls (2019) also performed a comparable analysis, that is, only a limited number of principles are necessary, namely beneficence, non-maleficence, autonomy, fairness, and explicability. At the European level, the document Ethics Guidelines for Trustworthy AI (European Commission, 2019) is a much-cited piece in which five different principles are also mentioned.

The principles nowadays mainly describe what ethical AI is, not how ethical AI can be implemented (Floridi, 2019). Floridi (2019) also distinguishes various risks with these principles: ethics shopping, ethics blue washing, ethics lobbying, ethics dumping, and ethics shirking. Ethics shopping means that when developing an algorithm, a party chooses only those principles that make it appear that the algorithm complies with all the principles. Ethics blue washing denotes that a party uses the argument that the algorithm complies with these principles, but complying to the principles is insufficient. Ethics lobbying is used by some parties to avoid introducing legislation because principles would suffice. In ethics dumping, parties develop algorithms or other concepts in other countries that have not defined ethics criteria. Finally, ethics shirking occurs when ethical standards drop when people believe that the negative impact is smaller.

Another widely cited scientist in the fields of ethics and AI is Mittelstadt. Mittelstadt also expresses his reservations about the principle-based approach alone. He lists four objections to this approach:

1) With AI, no common aims and fiduciary duties exist. The principle-based approach originates from the medical world, and the sole goal is to promote the health of individuals. However, this case does not hold for AI applications. In addition, Mittelstadt (2019) underscores the difficulty of achieving a balance between public and private interests in the development of these algorithms.
2) Professional history and standards have yet to be developed. The AI profession is younger than the medical profession; hence, these standards are not as mature as the ones in the medical field.
3) No methods have been developed to translate the principles into practice. Floridi and Cowls (2016) mention the same objection.
4) Legal or professional accountability is lacking. Within the healthcare domain, medical specialists are responsible. This aspect does not apply to the AI domain.

The principle-based approach disregards individual preferences (Telkamp & Anderson, 2022). Telkamp and Anderson (2022) argue that depending on the moral foundation, an individual may see an algorithm as more or less morally acceptable.

2.4 Moral foundations within the context of algorithms

Concerning moral foundations in the context of algorithms, Mittelstadt (2019) and Telkamp and Anderson (2022), among other researchers, indicate that the AI domain is still in development, that is, an ample amount of research into how algorithms should be developed in the proper manner is underway. Nonetheless, little research into how individual differences influence moral decisions about algorithms has been undertaken to date.

The way that individuals assess ethical behavior (based on moral foundations) also applies to the way that they analyze AI systems (Telkamp & Anderson, 2022). In their paper, Telkamp and Anderson (2022) recognize the different dimensions of AI in which different issues play a role. For these issues, they indicate how an individual with a different moral foundation assesses this issue in a different manner. An example is the issue of data collection and privacy concerns within the AI dimension of AI data and development. An individual with a high-care moral foundation conducts the assessment as follows: collecting data may result in the misuse of these data, thereby causing harm. More generically, when an individual sees that an AI system does not meet a moral foundation, the individual characterizes an AI system as unethical. More generally, Haidt and Josesph (2004) define the same process.

2.5 Moral acceptability within the context of algorithms

The moral acceptability of algorithms has been examined in several studies. In their research, Martin and Waldman (2022) show that an important starting point for testing the moral acceptability of algorithms is that they are context-specific. Telkamp and Anderson (2022) also highlight the importance of investigating the influence of moral foundations on the acceptability of algorithms. On the contrary, several generic studies have been conducted on the moral acceptability of algorithms. For example, Smith (2018) contends that many Americans find the performance of important processes by algorithms as acceptable. However, generic conclusions about the moral acceptability of algorithms cannot be drawn (Telkamp & Anderson, 2022).

The factors that influence the moral acceptability of algorithms have been investigated in several studies. Martin and Waldman (2022) conclude that the use of arbitrary data reduces the moral acceptability of these algorithms. Social media and race data are examples of these data. In addition, Martin and Waldman (2022) find that as a choice made with an algorithm has more impact, the moral acceptability decreases. Another comparable study is the research by Kodapanakkal et al. (2020) who, like Martin and Waldman, recognize outcome favorability as an important factor in determining moral acceptability. Kodapanakkal et al. (2020) similarly recognize the data protection element. Araujo et al. (2020) also investigated the moral acceptability of algorithms.

To increase moral acceptability, several papers specifically on algorithms follow a process to develop ethical algorithms. Turilli (2007) describes a process to move from principles to the concrete implementation of ethical algorithms.

2.6 Conceptual model

This thesis investigates the influence of moral foundations on the moral acceptability of algorithms, whereby moral foundations consist of the individualizing and the grouping concepts. How the conceptual model follows from the literature is described in this section.

According to Graham et al. (2009) and Nilson and Erlandsson (2015), the construct of individualizing moral foundations consists of the care moral foundation and the fair moral foundation. The care moral foundation reflects the extent to which an individual prefers to care for others (Graham et al., 2011). Telkamp and Anderson (2022, p. 963) indicate that the care moral foundation is concerned with "a general desire to alleviate suffering and foster well-being." By contrast, the fair moral foundation denotes the extent to which an individual finds fairness and justice important. The focus of the fair moral foundation is on the issue of "whether parties are treated fairly, justly, or equally", and the fair moral foundation is "sensitive to evidence of cheating and cooperation" (Telkamp & Anderson, 2022, p. 963).

Second, regarding the moral acceptability of algorithms, an algorithm is morally acceptable when it is good or desirable and does what it should do (Poel, 2016). This characteristic is normative.

Another key point is the relationship between moral foundations and the moral acceptability of algorithms. Individuals with a high care moral foundation consider algorithms good if they do not cause pain based on the definition of the care moral foundation; hence, they classify these algorithms as morally acceptable. By contrast, individuals with a high fair moral foundation characterize algorithms as good when such algorithms create a situation in which parties are treated fair and justly; thus, they delineate these algorithms as morally acceptable.

The ensuing question relates to how this works more specifically for algorithms within the social assistance domain. Answering this question requires an examination of the outcomes of the application of algorithms in the social assistance domain. Leben (2020) and Ruf (2021), among other researchers, argue that algorithms generate four possible outcomes: true positives, true negatives, false positives, and false negatives. From the perspective of the care and fairness moral foundation, false positives and false negatives particularly result in harm/unfairness. In the case of a false positive result, the citizen gets an investigation, but the citizen does not commit fraud. In the case of a false negative result, the citizen does not receive an investigation, but the chance that the citizen will have to pay a higher fine in the future is possible. Individuals with a higher moral foundation of care and a higher moral foundation of fairness view this as less morally acceptable. Individuals with a lower care and/or fair moral foundation find this more morally acceptable. In the case of true positives and true negatives, the algorithm predicts the correct result. On the contrary, for true positives and true negatives, no effects of either care moral foundation or fair moral foundation on their moral acceptability are expected.

As per the theories of Graham et al. (2009), Telkamp and Anderson (2022), and Leben (2022), this can be aggregated into stipulating a negative relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms. Based on this, the first hypothesis (H1) is proposed:

H1: Moral foundations in terms of individualizing have a negative relationship with the moral acceptability of algorithms.

The grouping moral foundation consists of the authority, grouping, and purity moral foundations (Graham et al., 2009; Nilson & Erlandsson, 2015). The grouping moral foundation indicates the extent to which an individual considers a group as important. The authority moral foundation signifies the extent to which an individual regards authority as important. The purity moral foundation denotes the extent to which purity is important.

As with H1, an algorithm can generate four types of outcomes: false negative, false positive, true positive, and true negative (Leben, 2020). Every outcome has an impact on individual citizens. The relationship between these three moral foundations is further substantiated below.

As for the authority moral foundation, Telkamp and Anderson (2022) indicate that it "concerns managing and maintaining effective status hierarchies, order, legitimacy, and direction." With each outcome, the algorithm supports an employee of the municipality, thereby making his work easier. The algorithms also contribute to maintaining order (i.e., lawful use of social assistance benefits). In this thesis, the researcher therefore hypothesizes that citizens with a higher authority moral foundation classify algorithms as more morally acceptable. Hence, a positive relationship exists between moral foundations in terms of authority and the moral acceptability of algorithms.

As for the loyalty moral foundation, Telkamp and Anderson (2022) state that it "concerns the need for individuals to form cohesive coalitions that can compete against other coalitions." The researcher mentions several groups that assess these algorithms, such as individuals on social assistance benefits and individuals without social assistance benefits. Individuals without social security benefits (the majority) find the application of algorithms (even if it has negative consequences for another group) more morally acceptable. Thus, a positive relationship exists between moral foundations in terms of loyalty and the moral acceptability of algorithms.

Regarding the purity moral foundation, Telkamp and Anderson (2022) indicate that it "concerns physical and spiritual cleanliness, and avoiding pathogens, parasites, diseases, and 'disgusting' people or objects." The researcher states that the application classifies individuals with high spiritual cleanliness as disgusting fraudsters who deliberately commit fraud. The application of these algorithms reduces the number of fraudsters, and it is more morally acceptable to individuals who consider it important. Hence, a positive relationship exists between moral foundations in terms of purity and the moral acceptability of algorithms.

As per the theories of Graham et al. (2009), Telkamp and Anderson (2022), and Leben (2022, the previous expectations can be aggregated into the expectation that individuals with a high grouping moral foundation view algorithms as more morally acceptable. By contrast, individuals with a low grouping moral foundation classify algorithms as less morally acceptable. Thus, hypothetically a positive relationship occurs between grouping moral foundation and the moral acceptability of algorithms in the welfare domain. Based on the preceding considerations, the second hypothesis (H2) is proposed:

H2: Moral foundations in terms of grouping have a positive relationship with the moral acceptability of algorithms.

Many researchers, including Mittelstadt et al. (2016), Barocas and Selbst (2016), and Hellmann (2020), identify discrimination as a risk of the application of algorithms. As a potential result, the negative impact for minority groups can be greater due to, for example, more incorrect predictions. This issue can be solved by training an algorithm with a fairness metric, in which the number of incorrect predictions is the same in all groups (also the minority groups), thereby reducing the risk of discrimination.

Individuals with a high care moral foundation intend to avoid negative effects, including discrimination (Graham et al., 2016). Applying a fairness metric decreases the chance of these negative effects, according to Green (2020) and Hellman (2020), among researchers. Meanwhile, in this thesis, the researcher hypothesizes that a higher care moral foundation has a positive effect on the moral acceptability of algorithms that are trained with a fairness metric.

Individuals with a high fair moral foundation strive for all parties to be treated equally and fairly (Graham et al., 2016). As Green (2020) and Hellman (2020) underscore, applying a fairness metric results in the equal distribution of the number of incorrect predictions among all groups, thereby constituting a more equal treatment of all groups. Therefore, in this thesis, the researcher hypothesizes that individuals with a high fair moral foundation classify algorithms where a fairness metric is applied as more morally acceptable.

For both individuals with a high fair moral foundation and individuals with a high care moral foundation, algorithms where a fairness metric has been applied are classified as more morally acceptable. This postulation also signifies that individuals with a high individualizing moral foundation (consisting of the care and fair moral foundation) regard the algorithms to which a fairness metric has been applied as more morally acceptable. Hence, the negative relationship as established in H1 will be weaker when a fairness metric is applied. Based on the above considerations, the third hypothesis (H3) is proposed:

H3: The relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms is moderated by fairness metrics in such a way that the negative relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms will be weaker when a fairness metric is applied.

According to Green (2020), Ruf (2021), and Hellman (2020), applying the fairness metric leads to the identification of fewer fraudsters. An advantage of applying a fairness metric is that the number of incorrect and/or correct predictions is the same in all groups.

Social order, obedience, and respect are important characteristics of individuals with a high authority moral foundation (Telkamp & Anderson, 2022). In this thesis, the researcher assumes that individuals with a high authority moral foundation prevails that more citizens are found committing fraud than preventing the negative effects of the application of algorithms. Therefore, applying a fairness metric to individuals with a high moral foundation result in a weaker relationship.

Loyalty pertains to “the need for individuals to form cohesive coalitions that can compete against other coalitions” (Telkamp & Anderson, 2022). At this juncture, the issue relates to what individuals with a high loyalty moral foundation think of applying a fairness metric. The largest group comprises those individuals who do not receive social assistance benefits. This group finds more importance in the idea that the negative consequences for this group are limited and that more fraudsters are identified. This group thus finds the application of a fairness metric as less important. Therefore, in this thesis, the researcher hypothesizes that applying a fairness metric results in a less morally acceptable algorithm for individuals with a high loyalty moral foundation.

Purity is concerned with "physical and spiritual cleanliness and avoiding pathogens, parasites, diseases, and 'disgusting' people or objects" (Telkamp & Anderson, 2022). Applying a fairness metric does not result in fewer fraudsters but in a fairer distribution of the false negatives. As with H2, the researcher states that individuals with a high purity moral foundation deem the absence of fraudsters as important. Therefore, the researcher hypothesizes that applying a fairness metric results in a less morally acceptable algorithm for individuals with a high purity moral foundation.

Individuals with a higher authority moral foundation, a higher loyalty moral foundation, and a higher purity moral foundation classify algorithms that apply a fairness metric as less morally acceptable; hence, the researcher hypothesizes that the relationship between moral foundations and the moral acceptability of algorithms becomes weaker. Based on the preceding considerations, the fourth hypothesis (H4) is proposed:

H4: The relationship between moral foundations in terms of grouping and the moral acceptability of algorithms is moderated by the fairness metrics applied in such a way that the positive relationship between moral foundations in terms of grouping and the moral acceptability of algorithms will be weaker when a fairness metric is applied.

The combination of multiple sources, including social media, produces several negative effects (Büchi et al., 2020). As Floridi (2012) suggests, collecting and combining more sources create more possibilities but result in less privacy (negative effect). The data minimization principle of the GDPR assumes that only the data that are necessary to achieve the goal may be collected (Hoofnagle et al., 2019).

Individuals with a high care moral foundation intend to avoid harm (Graham et al., 2011). Büchi et al. (2020) identify several negative effects of using social media data, or the so-called "chilling effects." Thus, in this thesis, the researcher hypothesizes that when social media is used, individuals with a high moral foundation view these algorithms as less morally acceptable.

Furthermore, individuals with a high fair moral foundation desire to be treated equally (Graham et al., 2011). Individuals are all treated equally with this algorithm, without any exceptions. However, individuals on social assistance benefit feel less free to post something on social media when the effects are considered (Büchi et al., 2020). Thus, in this thesis, the researcher argues that when social media is used, individuals with a high fair moral foundation view these algorithms as less morally acceptable.

In summary, this means that the relationship between the individualizing moral foundation and the moral acceptability of algorithms is stronger when public sources are used. Based on the foregoing considerations, the fifth hypothesis (H5) is proposed:

H5: The relationship between moral foundation in terms of individualizing and the moral acceptability of algorithms is moderated by the public data sources used in such a way that the negative relationship between moral foundations in terms of individualizing and the moral

acceptability of algorithms will be stronger when public data sources are used.

Leben (2020) and Floridi (2012) assert that adding more sources or data to the algorithm increases the chance of finding possible fraudsters. However, the disadvantage of adding these sources is that welfare recipients are less likely to post something on social media (Büchi et al., 2020).

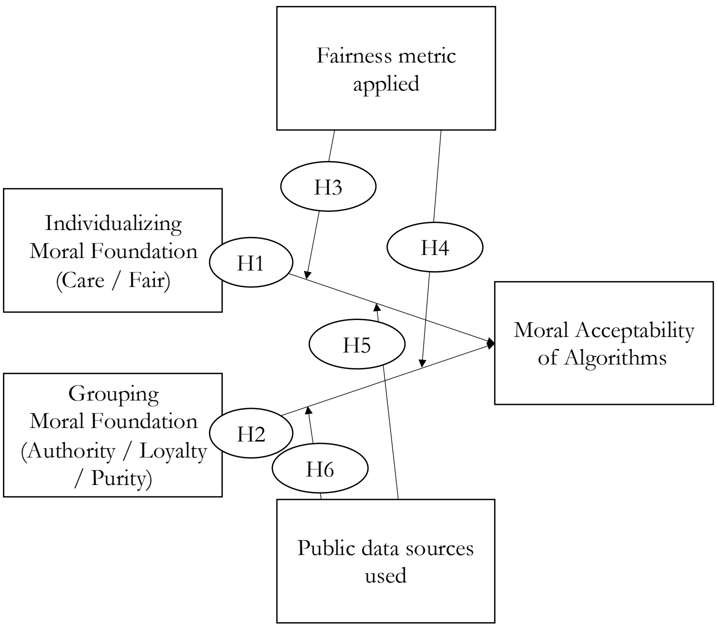
The argumentation for the sixth hypothesis (H6) is comparable to the support for the argumentation of H4. Individuals with high authority, purity, and loyalty moral foundations consider the identification of more fraudsters as more important than the possible negative consequences that such act entails. Adding a source that results in the finding of more fraudsters is therefore regarded as more morally acceptable by individuals with authority, purity, and loyalty moral foundations (and thus the moral foundation). Thus, in this thesis, the researcher hypothesizes that adding public sources (including social media) results in a more morally acceptable algorithm. Based on the above considerations, H6 is proposed:

H6: The relationship between moral foundations in terms of grouping and the moral acceptability of algorithms is moderated by the public data sources used in such a way that the positive relationship between moral foundations in terms of grouping and the moral acceptability of algorithms will be stronger when public data sources are used.

Based on the proposed hypothesis, the researcher developed a conceptual model as illustrated in Figure 6. This conceptual model is based on the initial conceptual model described in the introduction.

Figure 6

Conceptual Model with Hypothesis



3 Method

The focus of this chapter is on the research method. Section 3.1 and Section 3.2 explain the procedure and the research population, respectively. Section 3.3 details the questionnaire used in this study and the operationalization of the constructs from the research question. Section 3.4 and Section 3.5 describe the characteristics of the respondents in the dataset and the validation of the data, respectively. Finally, Section 3.6 presents the analysis of the data to answer the research question.

3.1 Procedure

The aim of this research is to investigate the relationship between moral foundations and the moral acceptability of the application of algorithms. A quantitative approach is adopted. A quantitative study is usually selected to validate a theory through data (Saunders et al., 2019). In addition, the goal of this research is to draw conclusions based on the results about a larger population: the Dutch population. Thus, a quantitative approach is adopted. Quantitative research often involves a deductive approach, which is also the case for this study. The researcher examines the relationship between moral foundations and moral acceptability based on a statistical approach. This research uses two constructs: moral foundations and the moral acceptability of algorithms to detect fraud. The four phases comprising the empirical research design are as follows: construction of the questionnaire; validation of the questionnaire with a sample of approximately 100 respondents; data collection with a sample of approximately 1,000 respondents; and analysis of the data.

3.2 Research population

The research is about the moral acceptability of algorithms in the social assistance domain that is specific to the Netherlands; thus, the population of the study includes Dutch citizens (>= 18 years old). A group of citizens older than 18 is selected because they are also entitled to vote in the Netherlands. The results of this research can have an impact on policymakers and politicians, hence this age limit. The same reason underlies the decision to obtain a sample of Dutch citizens who are older than 18. In 2017, 13,701,285 citizens comprised the Dutch population (Centraal Bureau voor de Statistiek, 2021). Given the size of the population, interviewing the entire population with a questionnaire is impossible. A sample from this population is therefore obtained. The selection of a random sample is not feasible either because the researcher does not have all the data on all citizens in the population. Thus, the researcher uses an online research panel through a research agency.

3.3 Questionnaire

The two constructs measured in this study are moral foundations and moral acceptability. Appendix A contains a detailed description of this questionnaire and its operationalization.

3.3.1 Construct of moral foundations

The construct of moral foundations is used as an independent variable in this study. As indicated by Graham et al. (2011; 2012) and further described in the theoretical framework, individuals differ in their moral foundations. To measure this construct, the researcher uses the validated Moral Foundations

Questionnaire (see Moral Foundations (2017). In their literature review of the Moral Foundations Questionnaire, Ellemers et al. (2019) indicate that this questionnaire is the most cited one for the measurement of moral foundations theory. In addition, these moral foundations play a role in assessing the moral acceptability of algorithms (Telkamp & Anderson, 2022). Therefore, this questionnaire is chosen for the current research.

The Moral Foundations Questionnaire comprises six questions for the five different moral foundations: care, fairness, loyalty, authority, and purity (Graham et al., 2011). These six questions consist of two categories. In the first category of relevance questions, the respondents are asked for three items per moral foundation to respond to this query: "When you decide whether something is good or bad, to what extent are the following considerations important for your judgment?" The respondents score these items on a Likert scale. The scale values include "not very relevant," "slightly relevant," "somewhat relevant," "very relevant," and "extremely relevant." In the second category of relevance judgments questions (Graham et al., 2011), the respondents are asked to score the following query: "To what extent do you agree or disagree with the following statements?" The respondents also score the statements based on a Likert scale from strongly disagree to strongly agree. Various researchers have used and validated this questionnaire in practice, including Van Leeuwen and Park (2009), Nilsson and Erlandsson (2022), Berg et al. (2022), and Graham et al. (2011).

The questionnaire has been translated from English into Dutch by a native English speaker and native Dutch speaker (see Moral Foundation, 2017). After inspection, the researcher concluded that this translation was good.

3.3.2 Construct of moral acceptability

In this study, the construct of moral acceptability is measured by presenting four different cases (vignettes) to the respondents. For each case, the questionnaire asks how morally acceptable the respondent evaluates this case on a scale of 0 to 100. This approach is comparable to the study by Kodapanakkal et al. (2020). In terms of research design, the researcher opts for a factorial vignette survey. Factorial vignette surveys are mainly used for obtaining better insights into respondents' judgments. This approach results in a high internal and external validity of the research compared to a standard questionnaire (Auspurg & Hinz, 2015). As a complex theme with a broad target group (citizens) is investigated in the current study, this method is appropriate. Auspurg and Hinz (2015) define a vignette as a brief, carefully composed description of a person, object, or situation. In the present research, each case (vignette) is varied on two aspects: first, whether public data are used and second, whether a fairness metric is applied. The moderators from the conceptual model include the fairness metrics used and the public data sources used. Both moderators can have a value of yes or no. Each vignette contains a description of where it is possible to change the value of the moderators. An example of a vignette is shown in Table 4. The usage of vignettes facilitates the measurement of the moral acceptability for the four use cases. Thus, four vignettes are utilized in this study.

Table 4

Example of a Vignette

A municipality uses an algorithm for predicting fraud in the social assistance domain.
The municipality uses the data from [alternative data sources used] for predicting fraud in the social assistance domain.
The algorithm is optimized to [alternative fairness metric applied].

3.3.3 Validation and development of the questionnaire

The development and validation of the questionnaire consist of two phases: programming of the questionnaire and test run of the questionnaire with 100 respondents. In the first phase, the researcher created a draft questionnaire. In defining this questionnaire, the researcher followed the recommendations of Baarda (2021): among other steps, registering the completion time, ensuring that the questionnaire can be completed within 10 minutes, and checking for language that everyone understands. During this phase, the researcher conducted the steps outlined in Table 5.

Table 5

Validation of the Questionnaire

#	Step	Received Feedback	Improvements
1	Academic and practice supervisors conduct an initial review of the questionnaire.	Various linguistic improvements are required. A validation run with multiple respondents should be performed.	In the next steps, the researcher performs multiple validation runs.
2	Ten respondents fill in the concept questionnaire with Qualtrics.	The 12 vignettes utilized for defining the moral acceptability of a use case are difficult to distinguish.	The number of vignettes is limited to four.
3	Five respondents complete the concept questionnaire with Qualtrics. The researcher is present when the respondent completes the questionnaire. The researcher asks the respondent to think aloud while completing the questionnaire.	The respondents ask many questions about the definitions in the questionnaire (e.g., what an algorithm is, what predictions are).	Verhagen et al. (2020) provide the starting point that a simpler questionnaire improves the completion rate of questionnaires. Hence, all the questionnaires were tested for simple language use (B1 language use). In addition, a film was created to introduce the vignettes.
4	The academic and practical supervisors review the questionnaire.	No new comments	No modifications
5	The market research agency programs the questionnaire, and the researcher asks for feedback on the questionnaire.	No new comments	No modifications

After the validation of the questionnaire as described in Table 5, the researcher performed a test run. A market research agency obtained a random sample from the market research dataset comprising 140,000 Dutch people. One hundred respondents filled in the questionnaire. A detailed description of this test run and the analysis is provided in Appendix C. Based on this test run, some adjustments were made to the questionnaire, such as measuring the time that a respondent needs to complete a question in the questionnaire and adding gender to the questionnaire.

3.4 Dataset

During this research, a market research agency conducted the data collection. This market research agency has a database of 140,000 Dutch people. The market research bureau derived a random sample from the market research dataset; 2,375 respondents completed the questionnaire. Table 6 presents an overview of the various characteristics of the respondents. The number of respondents per group in the dataset, the percentage of respondents per group in the dataset, and the percentage in the population (society) are also shown in the table. The analysis reveals that the dataset used for this study consists of 73.2% of respondents between 60 and 80 years old, and this age group represents 27.2% of the Dutch population

Table 6

Characteristics of Respondents

Characteristics of Respondents		**Dataset Used for the Final Questionnaire**		**Society**
Characteristic	**Level**	**N**	**% in sample**	**% in society**
Education level	University	147	13.1%	16.0% (according to Malowski, 2020)
	Higher vocational education (HBO)	405	36.2%	25.0% (according to Malowski, 2020)
	Senior general secondary education (HAVO), pre-university education (VWO), senior secondary vocational education 2–4 (MBO 2–4)	332	29.7%	39.0% (according to Malowski, 2020)
	Secondary vocational education (VBO), Prevocational secondary education (VMBO or MAVO), senior	226	20.2%	14.0% (according to Malowski, 2020)

	secondary vocational education 1 (MBO 1)			
	Only primary school	8	0.7%	6.0% (according to Malowski, 2020)
Age	18–20	2	0.2%	3.2% (according to Centraal Bureau voor de Statistiek, 2022)
	20–30	9	0.8%	18.1%
	30–40	17	1.5%	14.2%
	40–50	48	4.3%	15.8%
	50–60	157	14.0%	16.3%
	60–70	416	37.2%	15.6%
	70–80	402	36.0%	11.6%
	80–90	65	5.8%	5.1%
	90–100	2	0.2%	0.1%
Political preference	VVD	164	14.7%	21.9% (according to Kiesraad, 2021)
	D66	77	6.9%	15.0%
	PVV	78	7.0%	10.8%
	CDA	66	5.9%	9.5%
	SP	117	10.5%	6.0%
	PvdA	80	7.2%	5.7%
	Groenlinks	80	7.2%	5.2%
	PvdD	62	5.5%	3.8%
	Christenunie	61	5.5%	3.4%
	FvD	5	0.4%	5.0%
	Ja21	53	4.7%	2.4%
	SGP	16	1.4%	2.1%
	Denk	2	0.2%	2.0%
	Volt	43	3.8%	2.4%
	BBB	90	8.1%	1.0%
	Bij1	2	0.2%	0.8%
	Other	122	10.9%	3.0%

Gender	Male	761	68.1%	49.5% (according to Centraal Bureau voor de Statistiek, 2022)
	Female	357	31.9%	50.5%

3.5 Data preparation

To prepare the data and then answer the research questions, various analyses were performed with SPSS Statistics and Microsoft Excel. As described in the procedure, the researcher initially conducted a test run with 100 respondents and then collected the data with 2,375 respondents. These runs consisted of the following steps: removal of incorrect data, factor analysis, reliability analysis, and verification of the normality of the constructs.

3.5.1 Removal of incorrect data

In the final data collection, 2,375 respondents completed the survey. The market research agency paid the respondents for completing this survey. To ensure that the correct data had been analyzed, the researcher removed the following respondents from the dataset:

1. The Moral Foundations Questionnaire comprised two control variables. When a respondent gave an illogical answer to this control variable, the researcher removed these respondents from the dataset (a description of these control variables is provided in Appendix A). In this step, the researcher excluded 289 respondents.
2. In the questionnaire, an instruction film lasting 2 minutes was shown. To ensure that all respondents had the same knowledge about the theme, the researcher removed the respondents who watched the instruction film for less than 2 minutes. In this step, the researcher excluded 941 respondents.
3. In the questionnaire, 16 different questions were shown twice on one screen, where the respondent must select an answer from a Likert scale ranging from 1 to 7. When a respondent gave the same answers to all 16 questions, the researcher removed this respondent from the answer set. In this step, the researcher excluded 239 respondents.
4. Finally, the researcher removed the respondents who took less than 90 seconds to answer the on-screen questions for moral foundations. A measurement showed that someone who quickly filled out the questionnaire needed at least 90 seconds to complete the questionnaire. With this step, the researcher removed 189 respondents.

After the data cleaning, the dataset consisted of 1,118 respondents.

3.5.2 Confirmative factor analysis

Confirmatory factor analysis is used for testing the extent to which a theoretical construct appears in the dataset (Hair et al., 2019). Therefore, the investigator performed a confirmatory factor analysis.

For a comprehensive and detailed description of the confirmatory factor analyses performed, refer to Appendix F.

The researcher performed the confirmatory factor analysis based on the step-by-step plan of Hair et al. (2019). Furthermore, the model fit for different models as described by Graham et al. (2011) and Nilson and Erlandson (2014) was determined, namely the one-factor model, the two-factor model, the five-factor model, and the hierarchical factor model. In the one-factor model, all the items of the Moral Foundations Questionnaire load on one concept. In the two-factor model, all the items labelled with care and fairness load on the individualizing concept, and all the items labelled with authority, loyalty, and purity load on the grouping concept. In the five-factor model, all the items labelled with care load on the concept of care; all the items labelled with fairness load on the concept of fairness; all the items labelled with the loyalty concept load on the concept of loyalty; and all the items labelled with purity load on the concept of purity. The hierarchical model is similar to the five-factor model, but the concepts of fairness and care load on the concept of individualizing, and the concepts of loyalty, authority, and purity load on the concept of grouping. A visual representation is shown in Appendix F.

The researcher performed these models for the related items from the questionnaire, the judgment items and the relevance and judgement items (see Appendix A for a detailed description of the questionnaire). The researcher used different datasets (see also Appendix B for a detailed description of the datasets): the cleaned dataset, the dataset with the higher educated respondents, and the dataset with the lower educated respondents. The researcher distinguished between these datasets to test whether the results of the lower educated respondents lead to a lower model fit, as these respondents might not understand the questionnaire.

Several model-fit measures were used for determining the model fit for these models, including Comparative Fit Index (CFI), Goodness of Fit Index (GFI), Root Mean Square Error of Approximation (RMSEA), and CMIN/df. Scientific research indicates the ranges for these model fit measures to establish the model fit. Ridgon (1996) compares different model fit measures (RMSEA and CFI) and states that different researchers use various ranges for the model fit measures. Regarding the CFI, a minimum value of .95 is required for model fit (Bentler, 1990). Concerning the RMSEA, a maximum value of .06 to .10 is required for model fit (Hu & Bentler, 1999; Ridgon, 1996). As for the GFI, Hair et al. (2019) suggests a value above .95. Finally, for the CMIN/df, if this value is above >= 5, then a reasonable fit is obtained (Marsh & Hocevar, 1985).

The confirmatory factor analysis performed shows that no model with a specific dataset fit within the ranges of the model fit measures GFI, CFI, and RMSEA for an adequate model fit. However, the analysis does show that the five-factor model has the best model fit. In addition, only the relevance items apparently have a better model fit. The model with the best model fit is the five-factor model with only the relevance items with only the cleaned data (CMIN/df = 16.1, GFI = .849, CFI = .770, RMSEA = .770).

After solving the sources of misfit in the two-factor model with only the relevance questions, the researcher improved the model fit: CMIN/df = 10.3, GFI = .956, CFI = .941, RMSEA = .091 (90% CI [.080, .103]). The model after the improvements is presented in Table 7. To further improve the model fit, the researcher started with the two-factor model with only the relevance items. The researcher selected this model based on three arguments: (1) the models with only the relevance items have a better model fit than the models with all items; (2) the concepts of individualizing and grouping are used in the research; and (3) when the five-factor model with three items is selected and an item with a low factor loading is removed, insufficient items (<=3) per factor remain. The researcher omitted the items MF_FAIR_Q2, MF_FAIR_Q3, MF_AUTH_Q1, MF_AUTH_Q2, MF_LOY_Q1, MF_LOY_Q2, and MF_PUR_Q3 because they had an excessively low factor loading (<.55). Furthermore, the researcher deleted MF_CARE_Q3 because it resulted in an extremely low discriminant validity. The removal of MF_CARE_Q3 also improved the convergent validity of the individualizing concept.

Table 7

Identified Model

Construct	Items	Questions
Individualizing	MF_CARE_Q1	Whether or not someone suffered emotionally
	MF_CARE_Q2	Whether or not someone cared for someone weak or vulnerable
	MF_FAIR_Q1	Whether or not some people were treated differently from others
Grouping	MF_AUTH_Q3	Whether or not an action caused chaos or disorder
	MF_LOY_Q3	Whether or not someone showed a lack of loyalty
	MF_PUR_Q1	Whether or not someone violated the standards of purity and decency
	MF_PUR_Q2	Whether or not someone did something disgusting

In addition to the model fit measures, the construct reliability, convergent validity, and discriminant validity are also tested. Construct reliability is tested with the Cronbach alpha and the composite reliability. Hair et al. (2019) assume that a Cronbach's alpha of >= .70 is sufficient. In the current research, the Cronbach alpha values of the individualizing construct and the grouping construct are .609 and .768, respectively. Although the Cronbach alpha of the individualizing construct is lower than the required .70, it is consistent with the studies by Graham et al. (2011) and Nilson and Erlandson (2014), which also report a lower Cronbach alpha. The composite reliability values of the individualizing construct and the grouping construct are .644 and .729, respectively. These values are higher than the minimum thresholds. Convergent validity is tested using the average variance extracted (AVE). The AVE is higher than the acceptance level of .50 prescribed by Fornell and Larcker (1981). Table 8 shows the factor loadings, Cronbach alpha, composite reliability, AVE for the construct individualizing and grouping.

Table 8

Results of the Confirmatory Factor Analysis

Construct	Items	Loadings	Cronbach Alpha	Composite Reliability	AVE
Individualizing	MF_CARE_Q1	.80	.609	.644	.603
	MF_CARE_Q2	.46			
	MF_FAIR_Q1	.55			
Grouping	MF_AUTH_Q3	.60	.768	.729	.673
	MF_LOY_Q3	.76			
	MF_PUR_Q1	.57			
	MF_PUR_Q2	.76			

Finally, discriminant validity is tested. Discriminant validity occurs when the square root of the AVE of a given construct is greater than the correlation of the other items with that same construct (Fornell & Larcker, 1981). These figures (square root of AVE per construct) and the correlation between the constructs of individualizing and grouping) are shown in Table 9; discriminant validity is thus achieved.

Table 9

Discriminant Validity Check

	Individualizing	Grouping
Individualizing	**.776**	.274
Grouping	.274	**.820**

3.5.3 Normality analysis

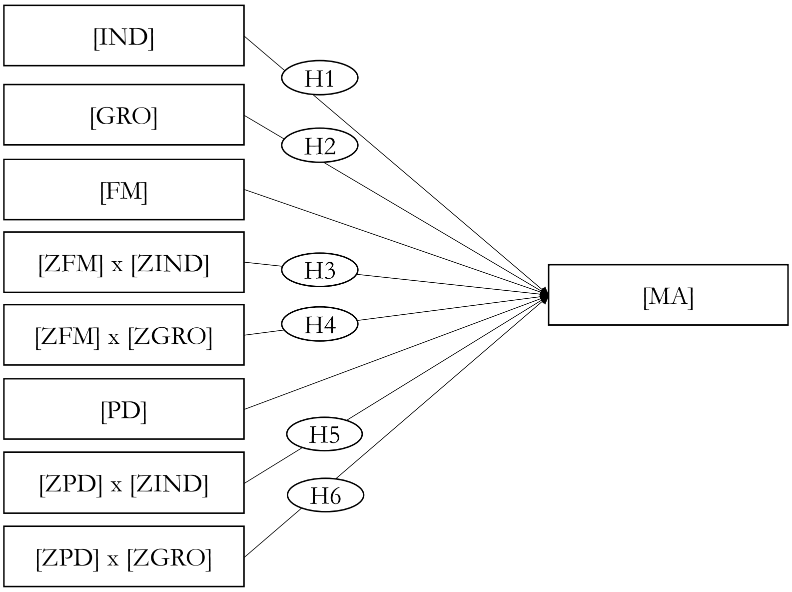
For the different measured constructs, a normality analysis is performed in three different ways: (a) by visually checking whether a normal distribution is visible on the histogram per scale; (b) by analyzing the kurtosis; and (c) by assessing the skewness. A detailed description of the normality analysis is provided in Appendix F. Skewness value of +/- 2.58 and kurtosis value of +/- 1.96 are required (Hair et al., 2019). Only the skewness is outside the boundaries; a visual inspection reveals that the histograms look like a normal distribution. The assumption of normality is the least important assumption in linear regression (Hayes, 2022). Hence, the normality analysis indicates that these scales are normally distributed.

3.6 Data analysis strategy

To answer these hypotheses, the researcher uses the statistics program SPSS 28.0.0.0. The researcher performs hierarchical multiple regression analysis to test whether the hypotheses are supported. To simplify the results section, the researcher uses the abbreviations shown between brackets and in italics throughout this thesis as depicted in Figure 7. IND is the individualizing concept. GRO is the grouping concept. FM is whether a fairness metric is applied. ZFM x ZIND is the interaction effect of the z-score of FM on the z-score of FM. ZFM x ZGRO is the interaction effect of the z-score of FM on the z-score of GRO. PD is whether public data sources are used. ZPD x ZIND is the interaction effect of the z-score of PD on the z-score of FM. ZPD x ZGRO is the interaction effect of the z-score of PD on the z-score of GRO. Finally, MA is the moral acceptability of the algorithm.

Figure 7

Statistical Model



4 Results

Chapter 4 presents the results based on the data analysis strategy described in the previous chapter. The stated hypotheses are tested through a multiple regression analysis.

4.1 Descriptive statistics and correlation

The conceptual model shown in Figure 7 uses several variables. The mean and standard deviation of these variables are presented in Table 10.

Table 10

Descriptive Statistics

Variable	Abbreviation	Mean	Standard Deviation
Moral acceptability	MA	60.910	32.882
Individualizing	IND	3.698	1.027
Grouping	GRO	4.046	1.098
Fairness metrics used	FM	0.510	0.500
Interaction term IND to FM	IND x FM	0.030	0.999
Interaction term GRO to FM	GRO x FM	0.004	1.000
Public data sources used	PD	0.516	0.500
Interaction term IND to PD	IND x PD	0.021	1.000
Interaction term GRO to PD	GRO x PD	0.014	1.001

The correlation matrix of the variables used without the interaction terms is shown in Table 11.

Table 11

Correlation Matrix

	MA	IND	GRO	FM	PD
MA	1.000				
IND	-.183**	1.000			
GRO	.073*	.243**	1.000		
FM	.281**	.030	.004	1.000	
PD	-.134**	.021	.014	.032	1.000

Note: Significance is denoted as ** significant at the .01 level and * significant at the .05 level.

4.2 Hierarchical multiple regression analysis

To test the hypotheses, a hierarchical multiple regression analysis is performed. A comprehensive description of the hierarchical multiple regression analysis is included in Appendix I. Table 12 shows the results of the hierarchical multiple regression analysis. During the study, a test is conducted to determine the presence of multicollinearity. The statistics are presented in Appendix I.

Table 12

Regression Output MA

Model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	60.423***	82.538***	71.015***	62.507***	62.421***	62.562***	67.264***	67.783***	67.627***
	(6.240)	(7.094)	(7.569)	(7.284)	(7.278)	(7.267)	(7.240)	(7.233)	(7.233)
Control variable									
AGE	.007	.001	.000	-0.003	.000	-.002	-0.009	-0.012	-.012
	(.093)	(.091)	(.091)	(.087)	(.087)	(0.087)	(0.086)	(0.086)	(0.086)
Independent variable									
IND		-5.860***	-6.837***	-7.118***	-7.144***	-7.123***	-7.045***	-7.225***	-7.289***
		(.943)	(.965)	(.923)	(.923)	(.921)	.911	.914	0.915
GRO			3.753***	3.782***	3.770***	3.749***	3.787***	3.867***	3.948***
			(.903)	(.863)	(.862)	(.861)	(.851)	(.851)	(.854)
Moderating variable									
FM				18.880***	18.886***	18.885***	19.179***	19.377***	19.385***
				(1.838)	(1.837)	(1.834)	(1.814)	(1.813)	1.813
PD							-9.447***	-9.453***	-9.449***
							(1.816)	(1.813)	(1.813)
Interaction variable									
IND x FM					1.600*	2.085**	2.319**	2.301**	2.268**
					(.920)	(.947)	(.937)	(.935)	(.936)
GRO x FM						-1.997**	-2.010**	-1.945**	-1.945**
						(.946)	(.935)	(.934)	(.934)
IND x PD								-1.895**	-2.115**
								(.909)	(.938)
GRO x PD									1.049
									(.935)
Observations	1118	1118	1118	1118	1118	1118	1118	1118	1118
R-squared	.000	.033	.048	.131	.133	.136	.157	.160	.161
Adjusted R-squared	-.001	.032	.046	.128	.129	.132	.152	.154	.154
Δ Adjusted R-squared	.000	.033	0.15	.082	.002	.003	.021	.003	.001
F-statistics	.006	19.309	18.824	41.812	34.116	29.262	29.535	26.465	23.670
Prob > F	.937	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Note: N = 1118. The standard errors are indicated in parentheses and the significance levels are determined and denoted with *** $p < .01$, ** $p < .05$, and * $p < .10$.

4.3 Hypotheses

The variables IND, GRO, FM, IND x FM, GRO x FM, PD, IND x PD, and GRO x PD significantly predict the moral acceptability of algorithms, $F(9, 1108) = 23.670$, $p < .001$, indicating that the eight factors investigated in this study have a significant impact on the moral acceptability of algorithms. $R^2 = .161$ denotes that 16.1% of the moral acceptability of algorithms is predicted by these eight variables. The results of the multiple regression analysis are shown in Table 20. In more detail, the hypotheses can also be answered.

H1: Moral foundations in terms of individualizing have a negative relationship with the moral acceptability of algorithms.

The first hypothesis (Moral foundations in terms of individualizing have a negative relationship with the moral acceptability of algorithms) is tested. The multiple regression analysis (see Table 12) shows that moral foundations in terms of individualizing have a negative relationship with the moral acceptability of algorithms ($B = -7.289$, $t = -7.964$, $p < 0.001$). Thus, H1 is supported.

H2: Moral foundations in terms of grouping have a positive relationship with the moral acceptability of algorithms.

The second hypothesis (Moral foundations in terms of grouping have a positive relationship with the moral acceptability of algorithms) is likewise tested. The multiple regression analysis (see Table 12) reveals that moral foundations in terms of grouping have a positive relationship with the moral acceptability of algorithms ($B = 3.948$, $t = 4.624$, $p < 0.001$). Hence, H2 is supported 2.

H3: The relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms is moderated by fairness metrics in such a way that the negative relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms will be weaker when a fairness metric is applied.

The third hypothesis (Moral foundations in terms of individualizing the moral acceptability of algorithms are moderated by the fairness metrics applied) is tested. The multiple regression analysis shows that when the fairness metric = 1, the relationship between individualizing and moral acceptability is less negative ($B = 2.268$, $t = 2.424$, $p = .016$), thereby resulting in a weaker relationship. Thus, H3 is supported.

H4: The relationship between moral foundations in terms of grouping and the moral acceptability of algorithms is moderated by the fairness metrics applied in such a way that the positive relationship between moral foundations in terms of grouping and the moral acceptability of algorithms will be weaker when a fairness metric is applied.

The fourth hypothesis (Moral foundations in terms of grouping on the moral acceptability of algorithms are moderated by the fairness metrics applied) is tested. The multiple regression analysis indicates that when the fairness metric = 1, the relationship between individualizing and moral acceptability is less negative (B =-1.945, t = -2.083, p = .037), thereby resulting in a weaker relationship. Hence, H4 is supported.

H5: The relationship between moral foundation in terms of individualizing and the moral acceptability of algorithms is moderated by the public data sources used in such a way that the negative relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms will be stronger when public data sources are used.

The fifth hypothesis (Moral foundations in terms of individualizing the moral acceptability of algorithms are moderated by the public data sources used) is tested. The multiple regression analysis reveals that when the public data sources used = 1, the relationship between individualizing and moral acceptability is more negative (B = -2.155, t = -2.298, p = .022), thereby resulting in a stronger relationship. Thus, H5 is supported.

H6: **The relationship between moral foundations in terms of grouping and the moral acceptability of algorithms is moderated by the public data sources used in such a way that the positive relationship between moral foundations in terms of grouping and the moral acceptability of algorithms will be stronger when public data sources are used.**

The sixth hypothesis (Moral foundations in terms of grouping and the moral acceptability of algorithms are moderated by the data sources used) is tested. The multiple regression analysis denotes that when public data sources used = 1, the relationship between grouping and moral acceptability is more positive (B = 1.049, t = 1.122, p = .262), thereby resulting in a stronger relationship; however, the relationship is not significant. Hence, H6 is not supported.

4.1. Conclusions

Based on this results chapter, several conclusions are drawn for the six hypotheses stated in the theoretical framework. The results of the hypothesis testing are summarized in Table 13.

Table 13

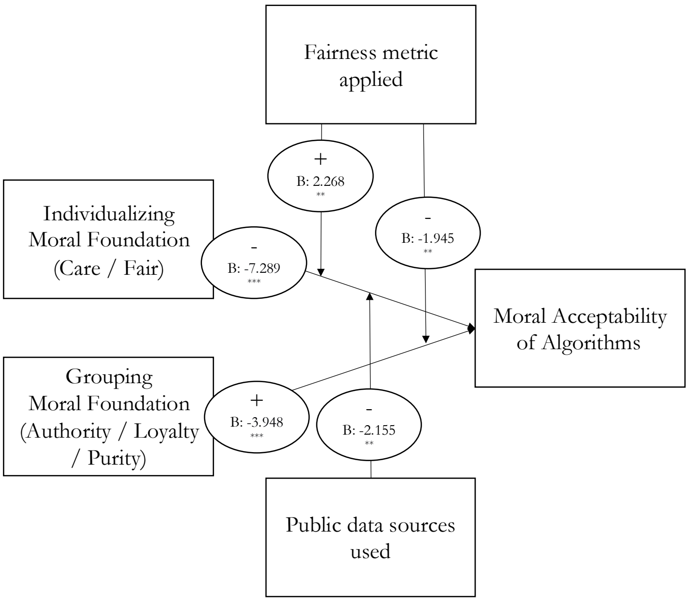
Results of the Hypothesis Testing

Hypothesis	Result
H1: Moral foundations in terms of individualizing have a negative relationship with the moral acceptability of algorithms.	Supported
H2: Moral foundations in terms of grouping have a positive relationship with the moral acceptability of algorithms.	Supported
H3: The relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms is moderated by fairness metrics in such a way that the negative relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms will be weaker when a fairness metric is applied.	Supported
H4: The relationship between moral foundations in terms of grouping and the moral acceptability of algorithms is moderated by the fairness metrics applied in such a way that the positive relationship between moral foundations in terms of grouping and the moral acceptability of algorithms will be weaker when a fairness metric is applied.	Supported
H5: The relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms is moderated by the public data sources used in such a way that the negative relationship between moral foundations in terms of individualizing and the moral acceptability of algorithms will be stronger when public data sources are used.	Supported
H6: **The relationship between moral foundations in terms of grouping and the moral acceptability of algorithms is moderated by the public data sources used in such a way that the positive relationship between moral foundations in terms of grouping and the moral acceptability of algorithms will be stronger when public data sources are used.**	Not supported

Based on the outcomes of the empirical testing of the conceptual model, an empirical model can now be constructed, which is shown in Figure 8.

Figure 8

Empirical Model



5 Conclusion, discussion, and recommendations

This chapter comprises three sections. Section 5.1 presents the main conclusion of this research. Section 5.2 provides the discussion. Finally, Section 5.3 outlines the limitations of this study and several recommendations for follow-up research.

5.1 Main conclusion

The research question, "To what extent do moral foundations influence the moral acceptability of algorithms in the social assistance benefits domain in the Netherlands?", is addressed in this thesis. The answer to the main question based on quantitative research (n = 1,118) is that a negative relationship exists between moral foundations in terms of individualizing and the moral acceptability of algorithms. This negative relationship will be weaker when a fairness metric is applied and will be stronger when public data sources are used. A positive relationship similarly exists between the grouping moral foundation and the moral acceptability of algorithms. Furthermore, the positive relationship between moral foundations in terms of grouping and the moral acceptability of algorithms will be weaker when a fairness metric is applied. This positive relationship between moral foundations in terms of grouping and moral acceptability is not stronger (or weaker) when public data sources are used. Approximately 16% of the variance of the moral acceptability is predicted moral foundations in terms of grouping, moral foundations in terms of individualizing, fairness metrics used when public data sources used.

5.2 Discussion

The use of generic principles for the development of artificial intelligence poses certain risks (Mittelstadt, 2020). Generic sets of principles for the development of artificial intelligence have been proposed in many studies. These principles can be used for assessing whether algorithms are ethical (Mittelstadt et al., 2015). Jobin et al. (2019) inventoried 84 documents with these principles. Floridi and Cowls (2019) identified the five most frequently quoted principles. Immanuel Kant introduced the concept of the categorical imperative (Gregor & Timmerman, 2012). The categorical imperative assumes that people act in the way that they would like all other people to act towards all individuals in the world. According to Kant, a neglect of the moral views of individuals is highly problematic. This quantitative research indicates that individuals with different moral foundations judge algorithms differently. This result implies that research based on generic principles, without considering these individual moral views, can create a situation in which less societal support for algorithms emerges. The use of generic principles for AI also has several limitations (Mittelstadt, 2020). In addition, Mittelstadt (2020) and the researcher of this study raise the same question, that is, whether the use of generic principles is prudent in the context of AI.

This research uses the construct of moral foundations from moral foundations theory as established by Graham et al. (2012). The current study shows the lack of model fit based on the thresholds determined by Hu and Bentler (1999) by performing the confirmatory factor analysis. Meanwhile, the Moral Foundations Questionnaire is the most cited in the literature, as shown by a literature review by Ellemers et al. (2019). Similar conclusions have been drawn from several other studies, including that of Nilsson and Erlandson (2015). Based on this research, one question in the questionnaire of Graham et al. (2011) results in a lower discriminant validity, and several questions have an excessively low factor loading. The present research uses refined notions of the individualizing and grouping concepts, which lead to an appropriate model fit. This research provides sufficient grounds for at least a follow-up investigation into the applicability of this questionnaire and proposes a new questionnaire.

The use of algorithms has both supporters and opponents (Barocas & Selbst, 2016). Fairness metrics provide an opportunity to limit the negative effects of algorithms, including discrimination (Green, 2020). However, the present research shows that fairness metrics also have advantages and disadvantages. The application of a fairness metric depends on the context (Popp Saenz, 2022). However, the current study adds that applying a fairness metric affects the relationship between moral foundations and the moral acceptability of algorithms.

5.3 Limitations and recommendations

This section tackles two topics: the limitations of the study and several recommendations at the academic and practice levels.

5.3.1 Limitations

The researcher used a market research agency for the data collection. The sample used in this study mainly comprises Dutch people between 60 and 80 years old (73.2%), which is not a representative sample of society (27.2% of Dutch people are between 60 and 80 years old). Thus, the results of this study are based on a dataset that primarily consists of Dutch people between 60 and 80 years old. On the contrary, the sample constitutes Dutch people with different political preferences.

During this research, individuals were informed beforehand about what algorithms are via a simple video with the advantages and disadvantages based on the literature. This implies that all respondents have, albeit limited, knowledge about these algorithms. During the validation of the questionnaire, many respondents were found to have no knowledge of algorithms and were therefore unable to provide a substantiated answer. However, the number of citizens with insufficient knowledge about this theme was not investigated. Thus, these results relate to Dutch people who have been informed in advance.

Moral foundations predict moral behavior in daily life to a limited extent (Berg et al., 2022). According to Berg et al. (2022), this reason explains why various other factors influence moral behavior, for example, whether a citizen has experienced a certain situation in everyday life. This finding of Berg et al. (2022) potentially leads to a different moral behavior than one would expect based on the moral foundation. However, other factors that impact the moral acceptability of algorithms are not identified in this research.

5.3.2 Recommendations

5.3.2.1 Academic recommendations

As this research has highlighted, moral foundations influence the moral acceptability of algorithms in terms of predicting fraud within the social assistance benefits domain. The extent to which an artificial intelligence algorithm is accepted strongly depends on the context (Martin & Waldman, 2022). The present research is limited to the welfare domain in the Netherlands. Hence, conducting this research in several other domains is recommended. Examples of these domains are insurance business, advertising, and music industry (Martin & Waldman, 2022).

The current conceptual model explains 16% of the difference in the moral acceptability of algorithms. Research is needed to expand the current conceptual model by including additional factors. Examples of these factors have been indicated by, among others, Martin and Waldman (2022): data protection, purpose for which it is used. Moreover, Berg et al. (2022) suggest that personal situations can have a strong influence on moral acceptability. The investigation of additional factors in the conceptual model is therefore recommended, such that a larger proportion of moral acceptability is predicted.

Within this research, the construct of moral foundations is used for exposing the differences between individuals. In their study, Ellemers et al. (2019) show that various questionnaires render the visibility of these differences. The questionnaire on moral foundations has been validated in diverse contexts, including in different parts of the world. This questionnaire has also been validated in a Swedish context by Nilson and Erlandson (2015). However, the confirmatory factor analysis of the present study indicates that the Moral Foundations Questionnaire does not meet the minimum thresholds for use, including the CFI, GFI, and RMSEA. Although an improved model for moral foundations in terms of individualizing and grouping has been proposed in this research, it has not been validated outside Dutch society. Thus, the recommendation is a follow-up research into the extent to which the improved conceptualizations can be applied in other studies and at other levels. One of the properties established by Graham et al. (2011) is that the moral foundation should be universally applicable.

5.3.2.2 Practical recommendations

AI is increasingly used in practice. The AI market is predicted to continually grow at a faster rate, from 14.4% between 2021 and 2022 to 31.1% between 2024 and 2025 (Woodward et al., 2021). Floridi (2012) attributes the increase in data to better infrastructure (computing and storage). Hence, a vast amount of artificial intelligence is expected to be used in the coming years. This research shows that individuals have different moral foundations and that these moral foundations influence how individuals judge algorithms as morally acceptable. Therefore, organizations and governments that use artificial intelligence should be aware of these influences, such that they also recognize the social impact of artificial intelligence.

In addition, the usage of fairness metrics or public data apparently influences this relationship. In other words, people with individualizing moral foundations positively view the use of fairness metrics.

The opposite case is true for people with a high grouping moral foundation. By contrast, people with a high individualizing moral foundation negatively rate the addition of public data. No relationship was measured here for people with a grouping moral foundation. This insight provides organizations and governments with the opportunity to increase social support among groups by adopting measures. However, undertaking such measures creates effects on other groups.

As also described by Floridi and Cowls (2019), European Commission (2019), and Jobin et al. (2019), governments and companies make algorithms comply with generic principles. Examples are privacy (including data minimization), justice, fairness, and equity. This research shows that people with diverse moral foundations think differently about how an algorithm should comply with these principles, for example, about whether to apply a fairness metric or use public data. In concrete terms, this research provides policymakers and companies with insight into how a society thinks differently about algorithms, thereby helping to stimulate a more informed debate about these algorithms. When establishing these principles, the researcher advises governments and organizations to constantly ask the question about the extent to which moral foundations influence them.

Literature

Aristotle, Brown, L., & Ross, D. (2009). *The Nicomachean Ethics.* Oxford University Press.

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Atzmuller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology*, *6*(3), 128–138. https://doi.org/10.1027/1614-2241/a000014

Auspurg, K. & Hinz, T. (2014). *Factorial Survey Experiments.* SAGE Publications.

Baarda, B., Bakker, E., Fischer, T., Julsing, M., & van Vianen, R. (2021). *Basisboek Methoden en Technieken, kwantitatief praktijkgericht onderzoek op wetenschappelijke basis.* Noordhoff uitgevers.

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, *104*(671), 671–732. https://doi.org/10.2139/ssrn.2477899

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Berg, T. G. C., Kroesen, M., & Chorus, C. G. (2022). Why Are General Moral Values Poor Predictors of Concrete Moral Behavior in Everyday Life? A Conceptual Analysis and Empirical Study. *Frontiers in Psychology*, *13*, 1–20. https://doi.org/10.3389/fpsyg.2022.817860

Büchi, M., Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A., Velidi, S., & Viljoen, S. (2020). The chilling effects of algorithmic profiling: mapping the issues *Computer Law & Security Review*, 36. 1-14. https://doi.org/10.1016/j.clsr.2019.105367

Carlson, M., & Mulaik, S. A. (1993). Trait Ratings from Descriptions of Behavior As Mediated by Components of Meaning. *Multivariate Behavioral Research*, *28*(1), 111–159. https://doi.org/10.1207/s15327906mbr2801_7

Centraal Bureau voor de Statistiek. (2021). *Bevolking; geslacht, leeftijd en burgerlijke staat, 1 januari.* Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/7461BEV/table?fromstatweb

Centraal Bureau voor de Statistiek. (2022). *Bevolking op 1 januari en gemiddeld; geslacht, leeftijd en regio.* Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/#/CBS/nl/dataset/03759ned/table?dl=39E0B

Centraal Bureau voor de Statistiek. (2023a). *Personen met bijstand; persoonskenmerken*. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82016NED/table?ts=1615283585713

Centraal Bureau voor de Statistiek. (2023b). *Bijstand; bijstandsvorderingen naar ontstaansgrond en regio*. Centraal Bureau voor de Statistiek. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82020NED/table?ts=1679982677706

Chouldechova, A. (2017). *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments.*

College Bescherming Persoonsgegevens. (2010). *Bestandskoppelingen door de SIOD voor de ontwikkeling van risicoprofielen.* College Bescherming Persoonsgegevens.

Cook, W., & Kuhn, K. M. (2021). Off-Duty Deviance in the Eye of the Beholder: Implications of Moral Foundations Theory in the Age of Social Media. *Journal of Business Ethics*, *172*, 605–620. https://doi.org/10.1007/s10551-020-04501-9

Dameski, A. (2020). *Foundations of an Ethical Framework for AI Entities: The Ethics of Systems* l'Universite du Luxembourg

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated Moral Reasoning. *Moral judgment and decision making*, *50*, 307–334. https://doi.org/10.1016/S0079-7421(08)00410-6

Egorov, M., Kalshoven, K., Verdorfer, A. P., & Peus, C. (2020). It's a Match: Moralization and the Effects of Moral Foundation Congruence on Ethical and Unethical Leadership Perception *Journal of Business Ethics, 167*, 707–723. https://doi.org/10.1007/s10551-019-04178-9

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017. *Personality and Social Psychology Review, 23*(4), 332–366. https://doi.org/10.1177/1088868318811759

European Commission. (2019). *Ethics guidelines for trustworthy AI.* European Commission. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Europese Commissie. (2021). *Verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie (wet op de artificiële intelligentie) en tot wijziging van de bepaalde wetgevingshandelingen van de unie*. Europese Commissie.

Fahse, T., Huber, V., & van Giffen, B. (2021). *Managing Bias in Machine Learning Projects.* International Conference on Wirtschaftsinformatik. https://doi.org/10.1007/978-3-030-86797-3_7

Floridi, L. (2012). Big Data and Their Epistemological Challenge. *Philos. Technol., 25*, 435–437. https://doi.org/10.1007/s13347-012-0093-4

Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology, 32*, 185–193. https://doi.org/10.1007/s13347-019-00354-x

Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1. https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, *14*, 349-379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d

Fornell, C., & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error Algebra and Statistics. *Journal of Marketing Research, 18*(3), 382–388. https://doi.org/10.2307/3150980

Glikson, E. (2020). Human Trust in Artificial Intelligence: A Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Graham, J., Nosek, B. A., & Haidt, J. (2011). Mapping the Moral Domain. *Journal of Personality and Social Psychology*, *101*(2). 1–38. https://doi.org/10.1037/a0021847

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2012). *Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism.* Advances in Experimental Social Psychology, Forthcoming.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*, 101–124. https://doi.org/10.1080/1047840X.2012.651387

Green, B. (2020). *Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness.* University of Michigan.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55–66.

Haidt, J., & Joseph, C. (2008). The Moral Mind 367–391. https://doi.org/10.1093/acprof:oso/9780195332834.003.0019

Haidt, J., & Kesebir, S. (2010). Handbook of Social Psychology (5th ed.). Wiley.

Hayes, A. F., Glynn, C. J., & Huge, M. E. (2012). Cautions Regarding the Interpretation of Regression Coefficients and Hypothesis Tests in Linear Models with Interactions *Communication Methods and Measures*, *6*. 1–11. https://doi.org/10.1080/19312458.2012.651415

Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (3th ed.). The Guilford Press

Hellman, D. (2020). Measuring algorithmic fairness. *Virginia Law Review*, *106*(4), 811–866.

Hill, R. K. (2016). What an Algorithm Is. *Philos. Technol., 29*, 35–59. https://doi.org/10.1007/s13347-014-0184-5

Hoofnagle, C. J., van der Sloot, B., & Borgesius, F. Z. (2019). The European Union's general data protection regulation: what it is and what it means *Information & Communications Technology Law, 28*(1), 65–98. https://doi.org/10.1080/13600834.2019.1573501

Howard, M. C. (2016). A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? *Intl. Journal of Human–Computer Interaction*, *32*, 51–62. https://doi.org/10.1080/10447318.2015.1087664

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jauk, S., Kramer, D., Avian, A., Berghold, A., Leodolter, W., & Schulz, S. (2021). Technology Acceptance of a Machine Learning Algorithm Predicting Delirium in a Clinical Setting: A Mixed-Methods Study. *Journal of Medical Systems, 45*(48), 1–8. https://doi.org/10.1007/s10916-021-01727-6

Jasso, G. (2006). Factorial Survey Methods for Studying Beliefs and Judgments. *Sociological Methods & Research, 34*(3), 334–423. https://doi.org/10.1177/0049124105283121

Jobin, A., Ienca, M., & Vayena, E. (2019). *Artificial Intelligence: The Global Landscape of Ethics Guidelines*

Kamiran, F., Zliobaite, I., & Calders, T. (2012). *Quantifying explainable discrimination and removing illegal discrimination in automated decision-making.*

Kiesraad. (2021). *Proces-verbaal van de verkiezingsuitslag van de Tweede Kamer.* Kiesraad. https://www.kiesraad.nl/binaries/kiesraad/documenten/proces-verbalen/2021/03/26/uitslag-tweede-kamerverkiezing-17-maart-2021/Proces-verbaal+verkiezingsuitslag+Tweede+kamer+2021.pdf

Kodapanakkal, R. I., Brandt, M. J., Kogler, C., & van Beest, I. (2020). Self-interest and data protection drive the adoption and moral acceptability of big data technologies: a conjoint analysis approach. *Computers in Human Behavior, 108*. 1-13. https://doi.org/10.1016/j.chb.2020.106303

Kohlberg, L., & Hersh, R. H. (1977). Moral development: a review of theory *Theory into Practice, 16*(2), 53–59. https://doi.org/10.1080/00405847709542675

Kraemer, F., van Overveld, K., & Peterson, M. (2010). Is there an ethics of algorithms? *Ethics and Information Technology*, *13*(3), 251–260. https://doi.org/10.1007/s10676-010-9233-7

Leben, D. (2020). *Normative Principles for Evaluating Fairness in Machine Learning.* AIES, New York.

Lee, M. K., & Baykal, S. (2017). *Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division.* ACM Conference on Computer Supported Cooperative Work and Social Computing. 1035-1048. http://dx.doi.org/10.1145/2998181.2998230

Logg, J. M. (2017). *Theory of Machines: When Do People Rely on Algorithms?* Harvard Business School.

Lotto, L., Manfrinati, A., & Sarlo, M. (2014). A New Set of Moral Dilemmas: Norms for Moral Acceptability, Decision Times, and Emotional Salience. *Journal of Behavioral Decision Making*, 27, 57–65. https://doi.org/10.1002/bdm.1782

Maninger, T., & Shank, D. B. (2022). Perceptions of violations by artificial and human actors across moral foundations *Computers in Human Behavior Reports*, *5*, 1–11. https://doi.org/10.1016/j.chbr.2021.100154

Marsh, H. W., & Hocevar, D. (1985). Application of Confirmatory Factor Analysis to the Study of Self-Concept: First- and Higher-Order Factor Models and Their Invariance Across Groups. *Psychological Bulletin, 97*(3), 532–582.

Martin, K. (2018). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*, 835–850. https://doi.org/10.1007/s10551-018-3921-3

Martin, K., & Nissenbaum, H. (2017). Privacy interest in public records: An emperical investigation. *Harvard Journal of Law & Technology, 31*, 111–143.

Martin, K., Shilton, K., & Smith, J. (2019). Business and the Ethical Implications of Technology: An Introduction to the Symposium. *Journal of Business Ethics, 160*, 307–317. https://doi.org/10.1007/s10551-019-04213-9

Martin, K., & Waldman, A. (2022). Are algorithmic decisions legitimate? The Effect of Process and Outcomes on Perceptions of the Legitimacy of AI Decisions. *Journal of Business Ethics, 183*. 653-670. https://doi.org/10.1007/s10551-021-05032-7

Maslowski, R. (2020, September). *De sociale staat van Nederland: onderwijs.* Sociaal Cultureel Planbureau. https://digitaal.scp.nl/ssn2020/onderwijs/

Ministerie van Sociale Zaken en Werkgelegenheid. (2017). *Evaluatie handhavingsbeleid Sociale Zaken en Werkgelegenheid.* Ministerie van Sociale Zaken en Werkgelegenheid.

Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*. 501-507. https://doi.org/10.1038/s42256-019-0114-4

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: mapping the debate *Big Data & Society*, *3*(2). 1-21. https://doi.org/10.1177/2053951716679679

Moral Foundations. (2017). *Moral Foundations Questionnairre*. https://moralfoundations.org

Mohr, S., & Kuhl, R. (2021). Acceptance of artificial intelligence in German agriculture: an application of the technology acceptance model and the theory of planned behavior *Precision Agriculture*, *22*, 1816–1844. https://doi.org/10.1007/s11119-021-09814-x

Nederlands Forensisch Instituut. (2017). *Risicoanalyse toelaatbaarheid bestandskoppeling.*

Nilsson, A., & Erlandsson, A. (2015). The Moral Foundations taxonomy: Structural validity and relation to political ideology in Sweden. *Personality and Individual Differences*, *76*, 28–32. http://dx.doi.org/10.1016/j.paid.2014.11.049

Olsthoorn, P. (2014). *Verdacht door Data.* iBestuur Magazine. https://ibestuur.nl/magazine/verdacht-door-data

Pirson, M., Martin, K., & Parmar, B. (2017). Formation of Stakeholder Trust in Business and the Role of Personal Values. *Journal of Business Ethics*, *145*, 1–20. https://doi.org/10.1007/s10551-015-2839-2

Poel, I. (2016). A Coherentist View on the Relation Between Social Acceptance and Moral Acceptability of Technology. *Philosophy of Engineering and Technology*, *23*. https://doi.org/10.1007/978-3-319-33717-3_11

Popp Saenz, A. (2022). *The Fairness Dilemma.* Vrije Universiteit.

PWC. (2013). *Extend of fraud.* https://www.consultancy.nl/nieuws/7513/pwc-fraude-kost-nederlander-600-euro-per-jaar

Rechtbank Den Haag (2020). *SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens.* https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:865&showbutton=true&keyword=den+haag+fraude

Ridgon, E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling, 3*(4), 369–379. https://doi.org/10.1080/10705519609540052

Ruf, B. (2021). *Towards the Right Kind of Fairness in AI*

Schellevis, J. (2021). *Fraude opsporen of gevaar van discriminatie? Gemeenten gebruiken "slimme" algoritmes.* NOS.nl. https://nos.nl/artikel/2366864-fraude-opsporen-of-gevaar-van-discriminatie-gemeenten-gebruiken-slimme-algoritmes

Schermer, B. (2011). The limits of privacy in automated profiling and data mining. *Computer Law & Security Review*, *27*, 45–52. https://doi.org/doi:10.1016/j.clsr.2010.11.009

Shank, D. B., & Gott, A. (2020). Exposed by AIs! People Personally Witness Artificial Intelligence Exposing Personal Information and Exposing People to Undesirable Content. *International Journal of Human-Computer Interaction*, *36*, 1636–1645. https://doi.org/10.1080/10447318.2020.1768674

Smith, A. (2018). *Public attitudes towards computer algorithms.* Pew Research Center.

Song, Y. W. (2019). *User Acceptance of an Artificial Intelligence (AI) Virtual Assistant: An Extension of the Technology Acceptance Model.* The University of Texas

Spector, L. (2006). Evolution of artificial intelligence. *Artificial Intelligence, 170*, 1251–1253. https://doi.org/10.1016/j.artint.2006.10.009

Suresh, H., & Guttag, J. (2021). *A Framework for Understanding Sources of Harm Throughout the Machine Learning Life Cycle* EAAMO.

Telkamp, J. B., & Anderson, M. H. (2022). The Implications of Diverse Human Moral Foundations for Assessing the Ethicality of Artificial Intelligence *Journal of Business Ethics, 178*, 961-976. https://doi.org/10.1007/s10551-022-05057-6

Turilli, M. (2007). Ethical protocols design. *Ethics and Information Technology*, *9*(1), 49–62. https://doi.org/10.1007/s10676-006-9128-9

Woodward, A., Griffen, A., Priestly, A., Hare, J., Hunter, E., & Quinn, K. (2021). *Forecast Analysis: Artificial Intelligence Software, Worldwide*. Gartner.

Appendix A: Questionnaire

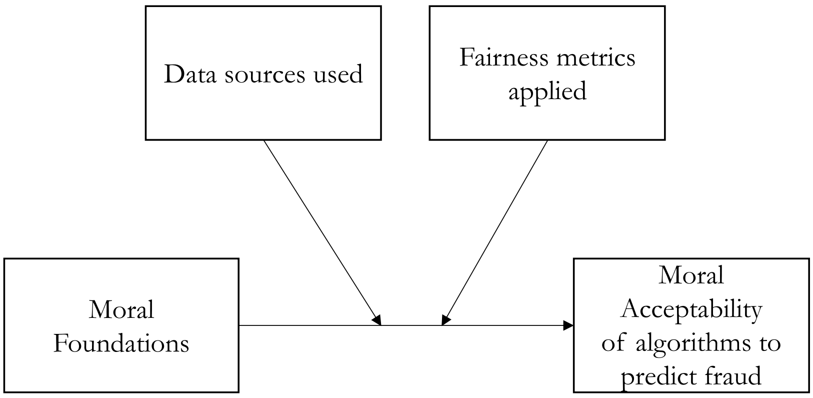
During the research, the researcher uses a questionnaire. Section A1 of this appendix describes how the researcher operationalized the questionnaire from the conceptual model. Section A2 describes how the researcher validated this questionnaire. Section A3 contains the description of two questions that are used as a control technique to determine the quality of the answers given. Section A4 contains the questionnaire. Section A5 the instruction that the researcher sent to the market research agency. Finally, section 0 contains the screenshots of the questionnairre.

A1 Operationalize questionnaire

Baarda et al. (2021) describe how it is possible to operationalize a conceptual model based on four steps: a) defining a concept and definition, b) distinguishing dimensions, c) devising indicators and d) devising items. The conceptual model is shown in Figure 9.

Figure 9

Conceptual Model



The conceptual model consists of four different concepts: moral foundations, data sources used, fairness metrics applied, moral acceptability of algorithms to predict fraud. The definition of these terms follows from the literature (see the theoretical framework for a detailed explanation).

Table 14 provides an overview of how the conceptual model has been operationalized into a questionnaire based on the step-by-step plan of Baarda et al (2021).

Table 14

Operationalization conceptual model to questionnaire

Definition	Dimension	Indicator	Item	Scale
Moral Foundations	Care	Care score (see Moral Foundations Theory (2017)). Number between 0 and 30. The care score per respondent is calculated as follows. Each respondent answers the questions associated with this dimension. Each answer has a number from 0 to 5 (see the scale column). The care score follows by adding up the values associated with the answers.	1. Whether or not someone suffered emotionally 7. Whether or not someone cared for someone weak or vulnerable 12. Whether or not someone was cruel	[0] = not at all relevant (This consideration has nothing to do with my judgments of right and wrong) [1] = not very relevant [2] = slightly relevant [3] = somewhat relevant [4] = very relevant [5] = extremely relevant (This is one of the most important factors when I judge right and wrong)
			17. Compassion for those who are suffering is the most crucial virtue. 23. One of the worst things a person could do is hurt a defenseless animal 28. It can never be right to kill a human being.	[0] Strongly disagree [1] Moderately disagree [2] Slightly disagree [3] Slightly agree [4] Moderately agree [5] Strongly agree
	Fairness	Fairness score (see Moral Foundations Theory (2017)). Number between 0 and 30. The fairness score is calculated similarly to the care score.	2. Whether or not some people were treated differently than others 8. Whether or not someone acted unfairly 13. Whether or not someone was denied his or her rights	[0] = not at all relevant (This consideration has nothing to do with my judgments of right and wrong) [1] = not very relevant [2] = slightly relevant [3] = somewhat relevant [4] = very relevant [5] = extremely relevant (This is one of the most

			important factors when I judge right and wrong)
		18. When the government makes laws, the number one principle should be ensuring that everyone is treated fairly. 24. Justice is the most important requirement for a society. 29. I think it's morally wrong that rich children inherit a lot of money while poor children inherit nothing.	[0] Strongly disagree [1] Moderately disagree [2] Slightly disagree [3] Slightly agree [4] Moderately agree [5] Strongly agree
Loyalty	Loyalty score (see Moral Foundations Theory (2017)). Number between 0 and 30. The loyalty score is calculated similarly to the care score.	3. Whether or not someone's action showed love for his or her country 9. Whether or not someone did something to betray his or her group 14. Whether or not someone showed a lack of loyalty	[0] = not at all relevant (This consideration has nothing to do with my judgments of right and wrong) [1] = not very relevant [2] = slightly relevant [3] = somewhat relevant [4] = very relevant [5] = extremely relevant (This is one of the most important factors when I judge right and wrong)
		19. I am proud of my country's history. 25. People should be loyal to their family members, even when	[0] Strongly disagree [1] Moderately disagree [2] Slightly disagree [3] Slightly agree [4] Moderately agree

		they have done something wrong. 30. It is more important to be a team player than to express oneself.	[5] Strongly agree
Authority	Authority score (see Moral Foundations Theory (2017)). Number between 0 and 30. The authority score is calculated similarly to the care score.	4. Whether or not someone showed a lack of respect for authority 10. Whether or not someone conformed to the traditions of society 15. Whether or not an action caused chaos or disorder	[0] = not at all relevant (This consideration has nothing to do with my judgments of right and wrong) [1] = not very relevant [2] = slightly relevant [3] = somewhat relevant [4] = very relevant [5] = extremely relevant (This is one of the most important factors when I judge right and wrong)
		20. Respect for authority is something all children need to learn. 26. Men and women each have different roles to play in society. 31. If I were a soldier and disagreed with my commanding officer's orders, I would obey anyway because that is my duty.	[0] Strongly disagree [1] Moderately disagree [2] Slightly disagree [3] Slightly agree [4] Moderately agree [5] Strongly agree
Sancity	Sanctity score (see Moral Foundations Theory (2017)). Number between 0 and 30.	5. Whether or not someone violated standards of purity and decency 11. Whether or not someone did something disgusting	[0] = not at all relevant (This consideration has nothing to do with my judgments of right and wrong) [1] = not very relevant

		The sancity score is calculated similarly to the care score.	16. Whether or not someone acted in a way that God would approve of	[2] = slightly relevant [3] = somewhat relevant [4] = very relevant [5] = extremely relevant (This is one of the most important factors when I judge right and wrong)
			21. People should not do things that are disgusting, even if no one is harmed. 27. I would call some acts wrong on the grounds that they are unnatural. 32. Chastity is an important and valuable virtue	[0] Strongly disagree [1] Moderately disagree [2] Slightly disagree [3] Slightly agree [4] Moderately agree [5] Strongly agree
Data sources used	Municipality data only	Yes / No	The municipality only uses the data of the municipality.	Yes / No
	Municipal data and public data	Yes / No	The municipality uses the data of the municipality and public data (including Social Media).	Yes / No
Fairness metrics applied?	Fairness metrics applied	Yes / No	The algorithm is trained to find the most fraudsters. A risk is that the algorithm more often gives an incorrect prediction for minority groups.	Yes / No
	No fairness metrics applied	Yes / No	The algorithm is trained so that the number of false and correct predictions is equal in all groups (including the minority groups). A	Yes / No

			risk is that this algorithm finds fewer fraudsters.	
Moral acceptability of algorithms in the fraud	Moral Acceptability	Moral acceptability (see Kodapanakkal (2020)).	How would you morally evaluate this application of an algorithm?	0 is morally unacceptable and 100 is morally acceptable
prediction domain	Domain	Domain where the algorithm is applied.	In which domain is the algorithm applied?	Fraud prediction domain

A2 Validation of questionnaire

To improve the quality of the questionnaire, the questionnaire was tested in various ways before it was sent to the respondents as described in Table 15.

Table 15

Validation of questionnaire

Validation step	Description	How processed in questionnaire?
1. External validation: define questionnaire based on existing questionnaire (Moral foundations and moral acceptability)	- The researcher uses the Moral Foundations Questionnaire (Moral Foundations Theory, 2017). This is a standard and validated questionnaire. Graham et al (2012) describes the relation of this questionnairre to the Moral Foundations Theory. - The researcher uses the moral acceptability as used in the research of Kodapanakkal (2020).	The questionnaire was copied 1-on-1.
2. Review of questionnaire by academic supervisors and practical supervisor.	- The initial version of the questionnaire has been reviewed by the practical supervisor and academic supervisors.	The vignettes are described more simply. The texts used are simplified. To avoid common method bias, the vignettes are shown in random order.

3. Test run of the questionnaire using a sent questionnaire in Qualtrics	- The version of the questionnaire that was reviewed in step 2 is modelled in Qualtrics (on-line survey engine). The survey is sent to more than 20 respondents and the survey is completed by 13 respondents. These were highly educated respondents. - The researcher called 4 respondents with the following questions: What did you think of the questionnaire? To what extent was the questionnaire understandable? - Feedback received is: o Difficult to distinguish the different vignettes. o What is an algorithm? o What is training an algorithm?	The initial questionnaire distinguished between 3 sources (data from the municipality, data from the municipality and the government and public data). This has been reduced to 2 sources. This reduces the number of vignettes to 8. In addition, the initial questionnaire distinguishes between 2 types of use cases (predict fraud and predict need). This is limited to predict fraud only. This limits the number of vignettes to 4. In addition, a more extensive text has been added to define algorithms.
4. Test run of the questionnaire by several respondents. The researcher asks the respondents to tell what they think while they fill in the questionnaire.	- During the test run, the respondents mentioned the following questions: o What is an algorithm? o What is training of an algorithm? o What is public data? o How can I go back to the previous question? o One respondent say that an algorithm is morally not acceptable and answers that the	The researcher rewrites the questionnaire to simple Dutch (B1 level). Simplifying through B1 language use leads, according to Verhagen et al. (2020), to a better response to more complex questions at lower educational levels. In addition, a video instead of text has been added. Adding a movie

		algorithm is morally acceptable.	prevents readers clicking to fast to the next question. It has also been added that respondents can click back and forth in the questionnaire.
5.	Test run with 100 respondents at a research agency	The researcher identified that only the time it takes a respondent to complete the vignettes was measured. It is not clear whether the respondent fills in questions about the moral foundations very quickly. The researcher also found that the gender was not registered.	The researcher made the following improvements: - Add gender to questionnaire. - Measure time that it takes to answer each questionnairre.

A3 Data quality check moral foundations questionnaire

The moral foundations questionnaire consists of two questions to identify whether the respondents give logical answers.

The first question is: when you decide whether something is right or wrong, to what extent is **whether** someone was good at math relevant to your thinking? Please rate each statement using this scale: When a respondent answers this question with: [3] = somewhat relevant, [4] = very relevant, or [5] = extremely relevant, the researcher removes this respondent from the dataset.

The second question is: Please read the following sentences: 'It is better to do good than to do bad.' and indicate your agreement or disagreement. When a respondent answers this question with: "strongly, moderately, or slightly disagree," the researcher removes this respondent from the dataset.

A4 Questionnaire

This section presents the questions as shown to the respondents. Because the research was conducted in Dutch, this section is in Dutch.

A4.1 Introductie

Wij doen wetenschappelijk onderzoek naar de mening van Nederlanders over een specifiek onderwerp. In de hiernavolgende schermen volgen verschillende vragen.

A4.2 Vignettes

[In filmpje waarin de volgende tekst wordt getoond]

In Nederland krijgen ongeveer 400.000 burgers geld van de overheid omdat ze zelf geen geld hebben.
Uit onderzoeken blijkt dat ongeveer 10% van de burgers fraudeert.
Gemeenten gebruiken computers voor het opsporen van deze frauderende burgers.
Maar hoe werkt dat?
De computer herkent bepaalde patronen uit de gegevens van mensen.
Bijvoorbeeld, wanneer u vaak een filmpje van dieren kijkt, denkt YouTube dat u deze filmpjes vaker wil kijken.
Een ander voorbeeld: uit gegevens blijkt dat frauderende burgers vaker op vakantie gaan.
De computer voorspelt dan dat burgers die vaker op vakantie gaan en bijstand ontvangen waarschijnlijk frauderen.
Welke keuzes kan een gemeente maken?
Welke gegevens gebruikt de gemeente daarbij?
De gemeente kan bijvoorbeeld alleen de gegevens van de gemeente gebruiken.
Of de gemeente kan ook publieke gegevens gebruiken. Bijvoorbeeld van social media (Facebook).
Wanneer de gemeente alleen de gegevens van de gemeente gebruikt, zijn burgers niet bang om op iets social media te plaatsen. Een nadeel is dat de gemeente minder fraudeurs opspoort.
Wanneer de gemeente ook de gegevens van social media gebruikt, spoort de gemeente meer fraudeurs op. Een nadeel is dat burgers banger zijn om iets op social media te plaatsen.
Computers geven net als mensen soms ook verkeerde voorspellingen. De computer denkt dan dat iemand fraudeert, terwijl dat niet zo is.
Soms komen deze fouten vaker voor bij burgers in minderheidsgroepen, bijvoorbeeld bij Nederlanders met een buitenlandse afkomst.
Hoe kan de gemeente daar mee omgaan?
De gemeente zorgt dat het aantal foute voorspellingen gelijk is in alle groepen. Een risico is dat de gemeente minder fraudeurs opspoort.
De gemeente zorgt er voor dat de computer de meeste fraudeurs vindt. Een risico is dat er een kans dat er meer fouten optreden bij minderheidsgroepen.
Wij geven u vier voorbeelden, waarbij wij u vragen hoe goed u dit voorbeeld van opsporing van fraudeurs vindt.
[New page]

Use Case 1:
Een gemeente gebruikt machine learning om fraudeurs te vinden.
De gemeente gebruikt daarbij alleen de gegevens van gemeente.

De gemeente zorgt er voor dat machine learning de meeste fraudeurs vindt. Een risico is dat er een kans dat er meer fouten optreden bij minderheidsgroepen.

Hoe moreel aanvaardbaar vindt u deze toepassing van een algoritme? (0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)

[New page]

Use Case 2:
Een gemeente gebruikt machine learning om fraudeurs te vinden.
De gemeente gebruikt daarbij alleen de gegevens van de gemeente zelf.
De gemeente zorgt dat het aantal foute voorspellingen gelijk is in alle groepen. Een risico is dat de gemeente minder fraudeurs opspoort.

Hoe moreel aanvaardbaar vindt u deze toepassing van een algoritme? (0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)

[New page]

Use Case 3:
Een gemeente gebruikt machine learning om fraudeurs te vinden.
De gemeente gebruikt daarbij de gegevens van gemeente, maar daarnaast ook nog publieke gegevens (waaronder Social Media).
De gemeente zorgt er voor dat machine learning de meeste fraudeurs vindt. Een risico is dat er een kans is dat er meer fouten optreden bij minderheidsgroepen.

Hoe moreel aanvaardbaar vindt u deze toepassing van een algoritme? (0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)

[New page]

Use Case 4:
Een gemeente gebruikt machine learning om fraudeurs te vinden.
De gemeente gebruikt daarbij de gegevens van gemeente en publieke gegevens (waaronder Social Media).
De gemeente zorgt dat het aantal foute voorspellingen gelijk is in alle groepen. Een risico is dat de gemeente minder fraudeurs opspoort.

Hoe moreel aanvaardbaar vindt u deze toepassing van een algoritme? (0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)

[New page]

A4.3 Moral Foundations

Wanneer je besluit of iets goed of slecht is, in welke mate zijn de volgende overwegingen dan van belang voor jouw oordeel? Scoor elke uitspraak op de volgende schaal:

[0] = Helemaal niet belangrijk (Deze overweging heeft niets te maken met mijn besluit over goed en slecht)

[1] = Niet erg relevant

[2] = Enigszins relevant

[3] = Redelijk relevant

[4] = Erg relevant

[5] = Heel erg relevant (Dit is een van de belangrijkste factoren wanneer ik oordeel over goed en slecht)

___ 1. Of iemand emotioneel heeft geleden

___ 2. Of sommige mensen anders behandeld werden dan anderen

___ 3. Of iemands daden liefde toonden voor zijn of haar land

___ 4. Of iemand te weinig respect voor autoriteit heeft getoond

___ 5. Of iemand standaarden van puurheid en fatsoenlijkheid geschonden heeft

___ 6. Of iemand goed was in wiskunde

___ 7. Of iemand zorgde voor een zwak of kwetsbaar iemand

___ 8. Of iemand oneerlijk heeft gehandeld

___ 9. Of iemand zijn of haar groep verraden heeft

___ 10. Of iemand zich conformeerde aan de tradities van de maatschappij

___ 11. Of iemand iets walgelijks heeft gedaan

___ 12. Of iemand wreed was

___ 13. Of iemands rechten zijn ontzegt

___ 14. Of iemand te weinig loyaliteit heeft getoond

___ 15. Of iemands actie chaos of wanorde veroorzaakte

___ 16. Of iemand zich gedroeg op een wijze die God zou goedkeuren

[New page]

A4.4 Moral foundations

Zou je voor de volgende stellingen aan willen geven in welke mate je het ermee eens of oneens bent.

[0]	[1]	[2]	[3]	[4]	[5]
Zeer mee oneens	redelijk mee oneens	enigszins mee oneens	enigszins mee eens	redelijk mee eens	zeer mee eens

____ 17. Medeleven met degenen die lijden, is de belangrijkste deugd.

____ 18. Wanneer de overheid wetten maakt, dan moet de garantie dat iedereen eerlijk behandeld wordt het belangrijkste principe zijn.

____ 19. Ik ben trots op de geschiedenis van mijn land.

____ 20. Respect voor autoriteit is iets dat alle kinderen moeten leren.

____ 21. Mensen behoren geen walgelijke dingen te doen, zelfs wanneer er niemand schade berokkend wordt.

____ 22. Het is beter iets goeds te doen dan iets slechts.

____ 23. Een van de ergste dingen die een mens kan doen is een weerloos dier pijn doen.

____ 24. Rechtvaardigheid is de belangrijkste behoefte voor een maatschappij.

____ 25. Mensen behoren loyaal te zijn aan hun familieleden, zelfs wanneer zij iets slechts hebben gedaan

____ 26. Mannen en vrouwen hebben elk verschillende rollen in de maatschappij.

____ 27. Ik vind sommige daden slecht, omdat zij onnatuurlijk zijn.

____ 28. Het kan nooit goed zijn om een mens te doden.

____ 29. Ik vind dat het moreel onjuist is dat rijke kinderen een heleboel geld erven, terwijl arme kinderen niets erven.

____ 30. Het is belangrijker om een teamspeler te zijn dan om jezelf te uiten.

____ 31. Als ik een soldaat was en ik was het oneens met de orders van mijn leidinggevende, dan zou ik toch gehoorzamen omdat dit mijn plicht is.

____ 32. Kuisheid is een belangrijke en waardevolle deugd.

[New page]

A4.5 Algemene vragen

Hieronder volgen nog enkele algemene vragen:

1. Wat is uw leeftijd?
2. Welke landelijke politieke partij heeft uw politieke voorkeur?
   a. VVD
   b. D66
   c. PVV

d. CDA
e. SP
f. PvdA
g. Groenlinks
h. Partij voor de Dieren
i. Christenunie
j. Forum voor Democratie
k. Ja21
l. SGP
m. Denk
n. Volt
o. BBB
p. Bij1
q. Overig

3. Wat is uw hoogste opleidingsniveau?
   a. WO
   b. HBO
   c. Havo / Vwo / mbo 2-4
   d. Vbo / mavo / vmbo / mbo-1
   e. Basisschoolniveau

[New page]

Bedankt

Bedankt voor het invullen van deze vragenlijst. Uw gegevens worden anoniem verwerkt.

A5 Questionnaire: instructions to the research agency

The researcher gives the research agency the following instructions when collecting data:

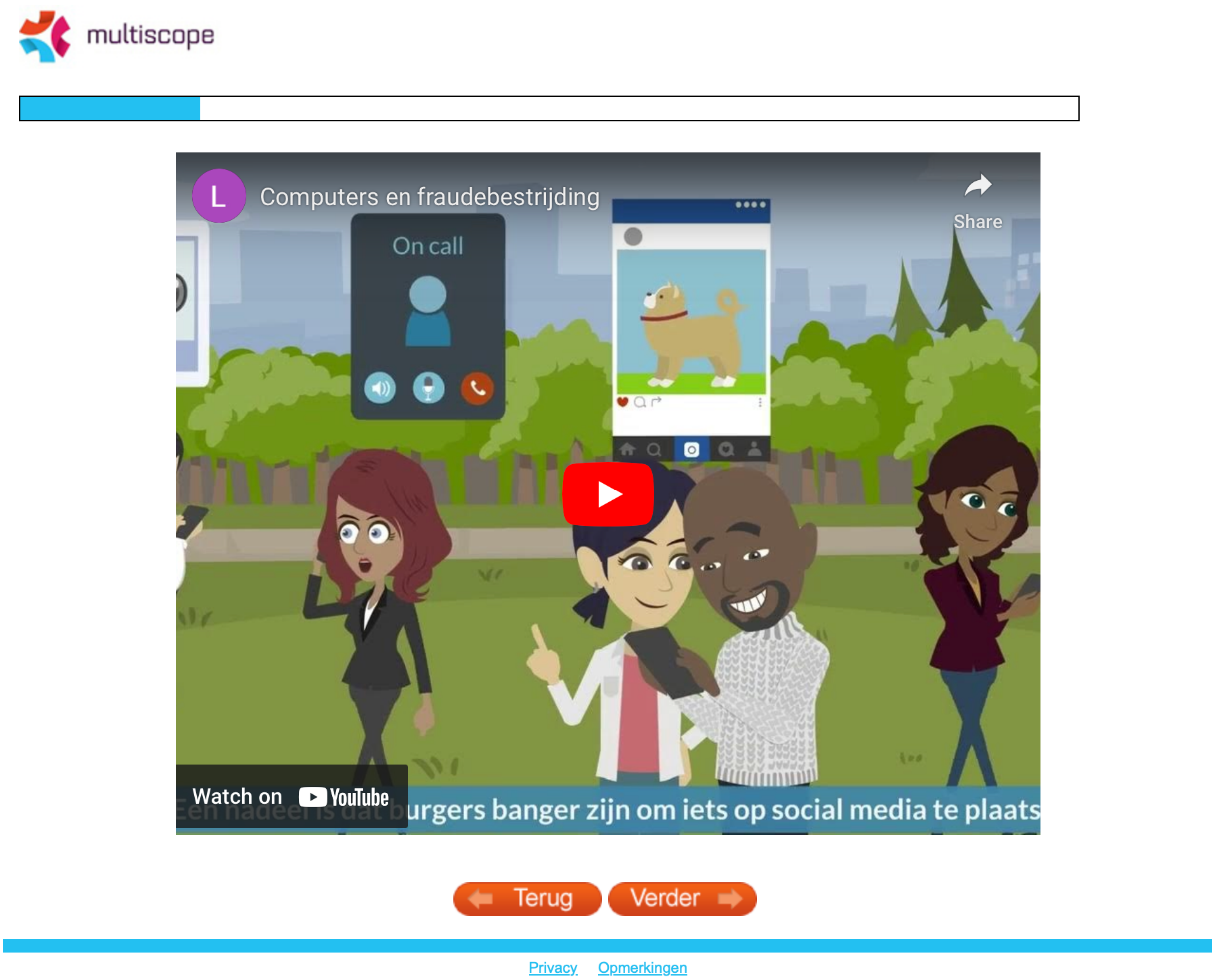
- Measure how much time a respondent needs to answer a question.
- Show the vignettes as mentioned under 1.3.2. in random order.
- Make sure that the respondents can go back to the previous question.
- Follow the new page instructions as depicted in 1.3.2;
- Can you validate the questionnaire and give feedback to the researcher?
- The researcher checks the questionnaire before it is distributed to the respondents.
- Start a test run with 100 respondents.

A6 Screenshots questionnaire

The following images in show the screenshots of the questionnaire.

Figure 10

Screenshots questionnaire



multiscope

Wij laten nu vier voorbeelden zien. Deze voorbeelden verschillen van elkaar. Let daarbij goed op de **vet gedrukte woorden**.

Terug Verder

Privacy Opmerkingen

multiscope

Voorbeeld

Een gemeente gebruikt een computer om fraudeurs te vinden.
De gemeente gebruikt daarbij de gegevens van **de gemeente, maar daarnaast ook publieke gegevens (waaronder social media)**.
De gemeente zorgt dat het aantal **fout voorspelde fraudeurs gelijk is bij alle groepen burgers**. Een risico is dat de gemeente **minder fraudeurs opspoort**.

Hoe moreel aanvaardbaar vind jij deze toepassing om fraudeurs te vinden?

(0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)

Moreel onaanvaardbaar 71 Moreel aanvaardbaar

Terug Verder

Privacy Opmerkingen

multiscope

Voorbeeld

Een gemeente gebruikt een computer om fraudeurs te vinden.
De gemeente gebruikt daarbij **de gegevens van de gemeente, maar daarnaast ook publieke gegevens (waaronder social media)**.
De gemeente zorgt ervoor dat de computer **de meeste fraudeurs vindt**. Een risico is dat er **meer fouten optreden bij minderheidsgroepen**.

Hoe moreel aanvaardbaar vind jij deze toepassing om fraudeurs te vinden?

(0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)

Moreel onaanvaardbaar 70 Moreel aanvaardbaar

Terug Verder

Privacy Opmerkingen

multiscope

Voorbeeld

Een gemeente gebruikt een computer om fraudeurs te vinden.
De gemeente gebruikt daarbij **alleen de gegevens van de gemeente zelf**.
De gemeente zorgt ervoor dat de computer **de meeste fraudeurs vindt**. Een risico is dat er **meer fouten optreden bij minderheidsgroepen**.

Hoe moreel aanvaardbaar vind jij deze toepassing om fraudeurs te vinden?

(0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)

Moreel onaanvaardbaar 75 Moreel aanvaardbaar

Terug Verder

Privacy Opmerkingen

multiscope

Voorbeeld

Een gemeente gebruikt een computer om fraudeurs te vinden.
De gemeente gebruikt daarbij **alleen de gegevens van de gemeente zelf**.
De gemeente zorgt dat het aantal **fout voorspelde frauders gelijk is bij alle groepen burgers**. Een risico is dat de gemeente **minder fraudeurs opspoort**.

Hoe moreel aanvaardbaar vind jij deze toepassing om fraudeurs te vinden?

(0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)

Moreel onaanvaardbaar 69 Moreel aanvaardbaar

Terug Verder

Privacy Opmerkingen

multiscope

Wanneer je besluit of iets **goed of slecht** is, in welke mate zijn de volgende overwegingen dan van belang voor jouw oordeel? Scoor elke uitspraak op de volgende schaal:

[0] = Helemaal niet belangrijk (Deze overweging heeft niets te maken met mijn besluit over goed en slecht)

[1] = Niet erg relevant

[2] = Enigszins relevant

[3] = Redelijk relevant

[4] = Erg relevant

Of iemand goed was in wiskunde - Erg relevant

[5] = Heel erg relevant (Dit is een van de belangrijkste factoren wanneer ik oordeel over goed en slecht)

	Helemaal niet belangrijk	Niet erg relevant	Enigszins relevant	Redelijk relevant	Erg relevant	Heel erg relevant
Of iemand emotioneel heeft geleden	○	○	●	○	○	○
Of sommige mensen anders behandeld werden dan anderen	○	○	●	○	○	○
Of iemands daden liefde toonden voor zijn of haar land	○	○	●	○	○	○
Of iemand te weinig respect voor autoriteit heeft getoond	○	○	●	○	○	○
Of iemand standaarden van puurheid en fatsoenlijkheid geschonden heeft	○	○	●	○	○	○
Of iemand goed was in wiskunde	○	○	●	○	○	○
Of iemand zorgde voor een zwak of kwetsbaar iemand	○	○	●	○	○	○
Of iemand oneerlijk heeft gehandeld	○	○	●	○	○	○

	Helemaal niet belangrijk	Niet erg relevant	Enigszins relevant	Redelijk relevant	Erg relevant	Heel erg relevant
Of iemand zijn of haar groep verraden heeft	○	○	●	○	○	○
Of iemand zich conformeerde aan de tradities van de maatschappij	○	○	●	○	○	○
Of iemand iets walgelijks heeft gedaan	○	○	●	○	○	○
Of iemand wreed was	○	○	●	○	○	○
Of iemands rechten zijn ontzegd	○	○	●	○	○	○
Of iemand te weinig loyaliteit heeft getoond	○	○	●	○	○	○
Of iemands actie chaos of wanorde veroorzaakte	○	○	●	○	○	○
Of iemand zich gedroeg op een wijze die God zou goedkeuren	○	○	●	○	○	○

Terug Verder

Geef voor de volgende stellingen aan in welke mate je het ermee eens of oneens bent.

	Zeer mee oneens	Redelijk mee oneens	Enigszins mee oneens	Enigszins mee eens	Redelijk mee eens	Zeer mee eens
Medeleven met degenen die lijden, is de belangrijkste deugd.						
Wanneer de overheid wetten maakt, dan moet de garantie dat iedereen eerlijk behandeld wordt het belangrijkste principe zijn.						
Ik ben trots op de geschiedenis van mijn land.						
Respect voor autoriteit is iets dat alle kinderen moeten leren.						
Mensen behoren geen walgelijke dingen te doen, zelfs wanneer er niemand schade berokkend wordt.						
Het is beter iets goeds te doen dan iets slechts.						
Een van de ergste dingen die een mens kan doen is een weerloos dier pijn doen.						
Rechtvaardigheid is de belangrijkste behoefte voor een maatschappij.						
	Zeer mee oneens	Redelijk mee oneens	Enigszins mee oneens	Enigszins mee eens	Redelijk mee eens	Zeer mee eens
Mensen behoren loyaal te zijn aan hun familieleden, zelfs wanneer zij iets slechts hebben gedaan						
Mannen en vrouwen hebben elk verschillende rollen in de maatschappij.						
Ik vind sommige daden slecht, omdat zij onnatuurlijk zijn.						
Het kan nooit goed zijn om een mens te doden.						
Ik vind dat het moreel onjuist is dat rijke kinderen een heleboel geld erven, terwijl arme kinderen niets erven.						
Het is belangrijker om een teamspeler te zijn dan om jezelf te uiten.						
Als ik een soldaat was en ik was het oneens met de orders van mijn leidinggevende, dan zou ik toch gehoorzamen omdat dit mijn plicht is.						
Kuisheid is een belangrijke en waardevolle deugd.						

Terug Verder

Privacy Opmerkingen

multiscope

Wat is jouw leeftijd?

Jaar

Welke landelijke politieke partij heeft jouw politieke voorkeur?

- VVD
- D66
- PVV
- CDA
- SP
- PvdA
- Groenlinks
- Partij voor de Dieren
- Christenunie
- Forum voor Democratie
- Ja21
- SGP
- Denk
- Volt
- BBB
- Bij1
- Overig

Wat is jouw hoogste voltooide opleiding?

- WO
- HBO
- Havo / vwo / mbo 2-4
- Vbo / mavo / vmbo / mbo-1
- Basisschool

Terug Verder

Privacy Opmerkingen

multiscope

Bedankt voor het invullen van deze vragenlijst. Jouw antwoorden worden anoniem verwerkt.

Privacy Opmerkingen

Appendix B: Datasets

This appendix describes the datasets used in section B1 and the fields within this dataset in section B2.

B1 Datasets

During the research, the researcher uses different datasets. The researcher refers to each dataset with a unique name. The attached Table 16 describes in detail what the characteristics of which dataset are. The next section describes the fields in the dataset.

Table 16

Datasets used

Identifier	Description	N
TestDataset0	Testset	94
FinalDataset1	Complete dataset of the real run.	2375
FinalDataset2	FinalDataset1, with the addition that: the moral foundation questionnaire consists of two control variables. When a respondent gave an illogical answer to this control variable, the researcher removed these respondents from the dataset.	2085
FinalDataset3	FinalDataset2, with the addition that: In the questionnaire, two times 16 different questions are shown on one screen, where the respondent must select an answer from a likert scale of 7. When a respondent gives the same answers on all the 16 questions, the researcher removes this respondent from the answer set.	2023
FinalDataset4	FinalDataset3 with the addition that: in the questionnairre an instruction film is shown that lasts 2 minutes. To ensure that all respondents have the same knowledge about the theme, the researcher removes respondents who watched the instruction film for less than 2 minutes.	
FinalDataset5	FinalDataset4 with the addition that: the researcher removes the respondents who take less than 90	1118

	seconds to answer the questions on screens for the moral foundations. A measurement shows that someone who fills in the questionnaire quickly needs at least 90 seconds to complete the questionnaire.	
FinalDataset6	Only the respondents with of FinalDataset4 with RESP_EDU = 'WO' and RESP_EDU = 'HBO'	552
FinalDataset7	Only the respondents of FinalDataset4 with RESP_EDU = ' Havo / vwo / mbo 2-4', RESP_EDU = 'Vbo / mavo / vmbo / mbo-1', RESP_EDU = 'Basisschool'	566

B2 Fields in dataset

The datasets have fields as shown in Table 17.

Table 17

Fields in datasets

Label	Description (in Dutch)
ID	Het nummer van de respondent.
MA_CMB_FM	Dit veld geeft 0 wanneer voor het voorbeeld waarvan de score is weergegeven in MA_CMB_SCORE geen fairness metric is toegepast en 1 wanneer er wel een = fairness metric is toegepast.
MA_CMB_PD	Dit veld geeft 0 wanneer voor het voorbeeld waarvan de score is weergegeven in MA_CMB_SCORE geen publieke databron is gebruikt en 1 wanneer er wel publieke databron is gebruikt.
MA_CMB_SCORE	De score van het voorbeeld dat als eerste wordt getoond (MA_SEQ_FIRST)
MA_FIRST	Een gemeente gebruikt een computer om fraudeurs te vinden. De gemeente gebruikt daarbij alleen de gegevens van de gemeente zelf. De gemeente zorgt ervoor dat de computer de meeste fraudeurs vindt. Een risico is dat er meer fouten optreden bij minderheidsgroepen. Hoe moreel aanvaardbaar vindt u deze toepassing om fraudeurs te vinden? (0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)
MA_FOURTH	Een gemeente gebruikt een computer om fraudeurs te vinden. De gemeente gebruikt daarbij de gegevens van de gemeente, maar daarnaast ook publieke gegevens (waaronder social media). De gemeente zorgt dat het aantal foute voorspellingen gelijk is bij alle groepen burgers. Een risico is dat de gemeente minder fraudeurs opspoort. Hoe moreel aanvaardbaar vindt u deze toepassing om fraudeurs te vinden? (0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)
MA_SECOND	Een gemeente gebruikt een computer om fraudeurs te vinden. De gemeente gebruikt daarbij alleen de gegevens van de gemeente zelf. De gemeente zorgt dat het aantal foute voorspellingen gelijk is bij alle groepen burgers. Een risico is dat de gemeente minder fraudeurs opspoort. Hoe moreel aanvaardbaar vindt u deze toepassing om fraudeurs te vinden? (0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)
MA_SEQ	Volgorde waarin voorbeelden V004B, V005B, V006B en V007B zijn getoond. In de vorm van 1|2|3|4
MA_SEQ_FIRST	Het voorbeeld dat is als eerste wordt getoond.

MA_THIRD	Een gemeente gebruikt een computer om fraudeurs te vinden. De gemeente gebruikt daarbij de gegevens van de gemeente, maar daarnaast ook publieke gegevens (waaronder social media). De gemeente zorgt ervoor dat de computer de meeste fraudeurs vindt. Een risico is dat er meer fouten optreden bij minderheidsgroepen. Hoe moreel aanvaardbaar vindt u deze toepassing om fraudeurs te vinden? (0 is moreel onaanvaardbaar en 100 is moreel aanvaardbaar)
MF_AUTH_Q1	Of iemand te weinig respect voor autoriteit heeft getoond
MF_AUTH_Q2	Of iemand zich conformeerde aan de tradities van de maatschappij
MF_AUTH_Q3	Of iemands actie chaos of wanorde veroorzaakte
MF_AUTH_Q4	Respect voor autoriteit is iets dat alle kinderen moeten leren.
MF_AUTH_Q5	Mannen en vrouwen hebben elk verschillende rollen in de maatschappij.
MF_AUTH_Q6	Als ik een soldaat was en ik was het oneens met de orders van mijn leidinggevende, dan zou ik toch gehoorzamen omdat dit mijn plicht is.
MF_CARE_Q1	Of iemand emotioneel heeft geleden
MF_CARE_Q2	Of iemand zorgde voor een zwak of kwetsbaar iemand
MF_CARE_Q3	Of iemand wreed was
MF_CARE_Q4	Medeleven met degenen die lijden, is de belangrijkste deugd
MF_CARE_Q5	Een van de ergste dingen die een mens kan doen is een weerloos dier pijn doen.
MF_CARE_Q6	Het kan nooit goed zijn om een mens te doden.
MF_FAIR_Q1	Of sommige mensen anders behandeld werden dan anderen
MF_FAIR_Q2	Of iemand oneerlijk heeft gehandeld
MF_FAIR_Q3	Of iemands rechten zijn ontzegd
MF_FAIR_Q4	Wanneer de overheid wetten maakt, dan moet de garantie dat iedereen eerlijk behandeld wordt het belangrijkste principe zijn.
MF_FAIR_Q5	Rechtvaardigheid is de belangrijkste behoefte voor een maatschappij.
MF_FAIR_Q6	Ik vind dat het moreel onjuist is dat rijke kinderen een heleboel geld erven, terwijl arme kinderen niets erven.
MF_LOY_Q1	Of iemands daden liefde toonden voor zijn of haar land
MF_LOY_Q2	Of iemand zijn of haar groep verraden heeft
MF_LOY_Q3	Of iemand te weinig loyaliteit heeft getoond
MF_LOY_Q4	Ik ben trots op de geschiedenis van mijn land.
MF_LOY_Q5	Mensen behoren loyaal te zijn aan hun familieleden, zelfs wanneer zij iets slechts hebben gedaan.
MF_LOY_Q6	Het is belangrijker om een teamspeler te zijn dan om jezelf te uiten.
MF_NA_Q1	Of iemand goed was in wiskunde
MF_NA_Q2	Het is beter iets goeds te doen dan iets slechts.
MF_PUR_Q1	Of iemand standaarden van puurheid en fatsoenlijkheid geschonden heeft
MF_PUR_Q2	Of iemand iets walgelijks heeft gedaan
MF_PUR_Q3	Of iemand zich gedroeg op een wijze die God zou goedkeuren
MF_PUR_Q4	Mensen behoren geen walgelijke dingen te doen, zelfs wanneer er niemand schade berokkend wordt.
MF_PUR_Q5	Ik vind sommige daden slecht, omdat zij onnatuurlijk zijn.
MF_PUR_Q6	Kuisheid is een belangrijke en waardevolle deugd.
RESP_AGE	Wat is jouw leeftijd?
RESP_EDU	Wat is jouw hoogste voltooide opleiding?

	1	WO
	2	HBO
	3	Havo / vwo / mbo 2-4
	4	Vbo / mavo / vmbo / mbo-1
	5	Basisschool
RESP_PP	Welke landelijke politieke partij heeft jouw politieke voorkeur?	
	1	VVD
	2	D66
	3	PVV
	4	CDA
	5	SP
	6	PvdA
	7	Groenlinks
	8	Partij voor de Dieren
	9	Christenunie
	10	Forum voor Democratie
	11	Ja21
	12	SGP
	13	Denk
	14	Volt
	15	BBB
	16	Bij1
	17	Overig
RESP_SEX	Het geslacht.	
TIME_INST	Hoe lang iemand naar het filmpje heeft gekeken dat twee minuten duurt.	
TIME_MF_2	Hoe lang iemand nodig had de tweede 16 vragen van de moral foundations questionnairre te beantwoorden.	
TIME_MF1	Hoe lang iemand nodig had de eerste 16 vragen van de moral foundations questionnairre te beantwoorden.	
TIME_S1	Hoe lang iemand nodig had om de eerste stelling te beantwoorden.	
TIME_S2	Hoe lang iemand nodig had om de tweede stelling te beantwoorden.	
TIME_S3	Hoe lang iemand nodig had om de derde stelling te beantwoorden.	
TIME_S4	Hoe lang iemand nodig had om de vierde stelling te beantwoorden.	

Appendix C: Data analysis test run

During the test run with 100 respondents, the researcher performed the steps as described in the following paragraphs and adjusted the questionnaire/method of data collection.

C1 Remove incorrect data

The data collection during this research was carried out by a research agency. Respondents in this survey are paid for conducting a survey. There is a risk that respondents complete this questionnaire too quickly (without reading the questions) which reduces the reliability of the data.

In two places, the survey shows 16 questions from the Moral Foundations Questionnaire on one page. The survey asks the respondents twice to score sixteen statements on a Likert scale consisting of 6 levels. Out of the n=100 respondents, 5 answered a set of statements at least once in the same way. The researcher removes these respondents from the research data.

During the test run, the researcher found that the time it takes to complete the questionnaire was not measured. This makes it impossible to determine whether a respondent has not read the questionnaire. It has therefore been proposed to measure, per page (see also the screenshots in section A6) how long it takes a respondent to complete the questions on a page.

C2 Factor analysis

Hair et al. (2019) describe two multivariate data analysis techniques to validate questionnaires: exploratory factor analysis and confirmatory factor analysis. The researcher uses exploratory factor analysis for the test run.

C2.1 Exploratory factor analysis

With an explorative factor analysis, according to Hair et al. (2019), it is possible to summarize information from different variables into different new composite variables (also called factors) with minimal information loss. The researcher performs an explorative factor analysis of the moral foundations. Graham et al. (2011) also describe the results of an exploratory factor analysis. It is therefore possible to compare both results.

For the explorative factor analysis, the researcher makes several choices. Hair et al. (2019) describe a step-by-step plan to perform an explorative factor analysis that is followed by the researcher. During this exploratory factor analysis, the researcher uses the oblimin rotation method. According to Hair et al. (2019), the oblimin rotation method best fits constructs that correlate with each other. This is the case with moral foundations. For example, Graham et al. (2012) describe that an individual with a high moral foundation often also scores higher on the fair moral foundation.

Table 18 gives the results of this explorative factor analysis. Hair et al. (2019) recommend removing the variables that do not map to a factor (with a factor loading <= .30) from the factor analysis.

Therefore, we removed the 6 variables that have factor loading of = .30. Costello and Osborn (2005) recommend in a literature review to remove the variables with cross loadings, for which the loading factor is >= 0.32 for two constructs. Based on this recommendation, the researcher removed 6 variables with cross-loading. Therefore, this table does not display 12 variables.

This factor analysis shows that two variables belonging to the grouping construct are now covered by the individualizing construct, namely MF_PUR_Q1 and MF_PUR_Q2. This is inconsistent with the study by Graham et al. (2011). When creating the questionnaire, the researcher received feedback that the moral foundations questionnaire is complex. The researcher sees this as a possible explanation for this result. That is why the researcher carries out the explorative factor analysis for only the higher educated (university and higher professional education). Table 19 shows the results of this explorative factor analysis. For this analysis, the researcher removed 4 variables that have no factor loadings < .30 and 5 variables with cross-loadings (MF_PUR_Q1, MF_PUR_Q2, MF_PUR_Q3, MF_PUR_Q5, MF_AUTH_Q3, MF_AUTH_Q5, MF_FAIR_Q1, MF_FAIR_Q6, MF_LOY_Q5). The KMO of this explorative factor analysis is 0.508 and is therefore too low in accordance with Howard (2016). However, a possible explanation is that there were only 54 respondents in this set.

Table 18

Results exploratory factor analysis

Item	Factor 1 (Individualizing)	Factor 2 (Grouping)
MF_CARE_Q1	.520	
MF_CARE_Q2	.555	
MF_CARE_Q3	.728	
MF_CARE_Q4	.598	
MF_FAIR_Q2	.527	
MF_FAIR_Q3	.586	
MF_FAIR_Q6	.383	
MF_LOY_Q3		.464
MF_LOY_Q4		.720
MF_LOY_Q6		.641
MF_AUTH_Q2		.569
MF_AUTH_Q4		.645
MF_AUTH_Q6		.545
MF_PUR_Q1	.685	
MF_PUR_Q2	.665	
MF_PUR_Q4		.643
MF_PUR_Q5		.430

MF_PUR_Q6		.512
Explained variance (%)	22.00	15.14
Cronbach's alpha	.769	.755

Note: Factor loadings <= .30 are not shown.

Table 19

Explorative factor analysis (only high edecuated)

Item	Factor 1 (Grouping)	Factor 2 (Individualizing)
MF_CARE_Q1		.556
MF_CARE_Q2		.486
MF_CARE_Q3		.464
MF_CARE_Q4		.760
MF_CARE_Q5		.640
MF_CARE_Q6		.388
MF_FAIR_Q2		.523
MF_FAIR_Q3		.457
MF_FAIR_Q4		.610
MF_FAIR_Q5		.528
MF_LOY_Q1	.364	
MF_LOY_Q2	.358	
MF_LOY_Q3	.655	
MF_LOY_Q4	.793	
MF_LOY_Q6	.601	
MF_AUTH_Q1	.677	
MF_AUTH_Q2	.679	
MF_AUTH_Q4	.642	
MF_AUTH_Q6	.506	
MF_PUR_Q4	.450	
MF_PUR_Q6	.460	
Explained variance (%)	21.033	13.198
Cronbach's alpha	.797	.741

Note: Factor loadings <= .30 are not shown.

C2.2 Confirmative factor analyse

With a confirmatory factor analysis, according to Hair et al. (2019), it is possible to test to what extent a certain theoretical construct can be found in the data. This research uses the concept of moral foundations.

Based on Quran (2016), a confirmatory factor analysis with 6 factors and 6 items each requires a minimum of 266 respondents. A confirmatory factor analysis can only be performed after the collection of data from more than 266 respondents.

The results of this confirmatory factory analysis can be compared with the research done by Nilsson and Erlandson (2015). The research by Nilsson and Erlandson (2015) is an investigation into, among other things, how the items of the Moral Foundations Questionnaire load into the constructs that are also used in this research. The research by Nilsson and Erlandson (2015) only relates to the Swedish research group, not the Dutch. For the output of the confirmatory factor analysis, the researcher uses SPSS Amos.

C2.3 Reliability analysis

During the study, a reliability analysis of the constructs used in this study was performed. Tables 6 and 7 show Cronbach's alpha for the concepts found in the exploratory factor analyses. Hair et al. (2019) assume that a Cronbach's alpha of >= .70 is sufficient. All values found in Cronbach's alpha are >= .70.

C2.4 Normality analysis

For the different measured constructs, a normality analysis was performed in three different ways: a) by visually checking whether a normal distribution is visible on the histogram per scale; b) by analyzing the kurtosis; and c) by analyzing the skewness. Appendix E provides a detailed description of this normality analysis. Hair et al. (2019) give the requirement that the zskewness be +/- 2.58 and the zkurtosis +/- 1.96. This is the case for both constructs. It also follows from the visual inspection that the histograms look like a normal distribution. So, it follows from the normality analysis that these are normally distributed scales.

Appendix D: Exploratory Factor Analysis test run

During the test run (first 100 respondents) the researcher performs an explorative factor analysis. When performing the explorative factor analysis, the researcher uses the step-by-step plan of Hair et al. (2019). This step-by-step plan contains various choices that must be made. This appendix describes the choices made.

Stage 1: Objectives of factor analysis

In stage 1, the goal of the factor analysis must be determined. The objective of this factor analysis is to verify whether scientifically defined concepts follow from the data. Based on Hair et al. (2019), the researcher should then perform a confirmatory factor analysis (CFA). During the trial run, there were another 95 respondents. A minimum of 266 respondents are required for a CFA, according to the Koran (2016). That is why the researcher opts for an EFA for the test run.

Based on Hair et al. (2019), the researcher chooses an R-type factor analysis (a study based on variables and not based on respondents).

Stage 2: Designing an exploratory factor analysis

As input, the researcher takes the fourth version of the Moral Foundations Questionnaire as defined by moral foundations theory (2017). This questionnaire consists of 30 items and was answered by 95 respondents. Hair et al. (2019) give the starting point that a sample for an EFA should consist of at least 50 respondents. As the sample gets smaller, the communality should be higher. Hair et al. (2019) recommend that communalities should fall in the range of.40 .40 and .70 for a sample of >= 200. For a sample = 200, a communality of >= .70 applies.

Stage 3: Assumptions in exploratory factor analysis

During stage 3, the researcher tests two different tests: the Barlett's test of spericity and the Kaiser Maiser Olkin test. Prior to running the EFA, the researcher determined that KMO equals .679 and the Barlets test of spericity is significant (<0.001). With this, the preconditions for the EFA have been tested.

Stage 4: Deriving factors and assessing overall fit

In stage 4, the researcher determines how the exploratory factor analysis is performed. Based on Hair et al. (2019), the researcher chooses the principal component analysis model. This model is mainly used when new constructs must be derived from different variables.

In addition, the researcher chooses two factors as a stopping rule for the exploratory factor analysis. From the study by Graham et al. (2011), the researcher expects that two constructs can be derived from the data: the grouping and the indivualizing moral foundation.

Stage 5: Interpreting the factors

During step 5, the researcher interprets the results of the factor analysis. Based on Hair et al. (2019), the researcher opts for an oblique factor rotation method. An oblique factor rotation method is chosen when the factors to be derived can correlate with each other. Studies by Graham et al. (2011) show that respondents with a high fair moral foundation probably also have a high care moral foundation. It is therefore likely that the constructs found correlate with each other, which is why the researcher opts for the oblique factor rotation method.

Hair et al. (2019) show that the factor pattern matrix is most often used to analyze the explorative factor analysis. Therefore, the researcher also uses this analysis.

The researcher performed an explorative factor analysis for all data, the results of which are shown in Table 20. Based on Hair et al. (2019), it is possible to remove variables that do not have factor loadings on the identified constructs. In addition, Costello and Osborn (2005) recommend, based on a literature review, to remove the variables with crossloadings, for which the loading factor is >= 0.32 for two constructs. Table shows the results of the exploratory factor analysis for the variables without factor loading and the variables with cross-loading removed and the results after this transformation.

Table 20

Results exploratory factor analysis

Part	Before removing cross loadings and variables with no factor loadings	After removing cross loadings and variables with no factor loadings
KMO and Bartlett's test	**KMO and Bartlett's Test** Kaiser-Meyer-Olkin Measure of Sampling Adequacy. .679 Bartlett's Test of Sphericity: Approx. Chi-Square 1118.798; df 435; Sig. <.001	**KMO and Bartlett's Test** Kaiser-Meyer-Olkin Measure of Sampling Adequacy. .682 Bartlett's Test of Sphericity: Approx. Chi-Square 544.246; df 153; Sig. <.001

Total variance explained

Total Variance Explained

Component	Initial Eigenvalues Total	Initial Eigenvalues % of Variance	Initial Eigenvalues Cumulative %	Rotation Sums of Squared Loadings[a] Total
1	5.863	19.543	19.543	4.908
2	3.159	10.531	30.074	4.547
3	2.349	7.831	37.905	
4	1.929	6.429	44.334	
5	1.847	6.156	50.490	
6	1.456	4.852	55.342	
7	1.211	4.037	59.380	
8	1.101	3.669	63.049	
9	1.048	3.494	66.543	
10	1.044	3.481	70.024	
11	.860	2.865	72.889	
12	.842	2.806	75.695	
13	.776	2.586	78.281	
14	.759	2.532	80.813	
15	.630	2.099	82.912	
16	.586	1.953	84.865	
17	.574	1.914	86.779	
18	.550	1.833	88.611	
19	.498	1.662	90.273	
20	.447	1.489	91.762	
21	.412	1.374	93.136	
22	.377	1.256	94.392	
23	.319	1.064	95.456	
24	.284	.948	96.404	
25	.268	.893	97.297	
26	.246	.819	98.116	
27	.175	.584	98.700	
28	.166	.555	99.255	
29	.122	.407	99.662	
30	.101	.338	100.000	

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Total Variance Explained

Component	Initial Eigenvalues Total	Initial Eigenvalues % of Variance	Initial Eigenvalues Cumulative %	Rotation Sums of Squared Loadings[a] Total
1	3.961	22.006	22.006	3.484
2	2.724	15.134	37.140	3.300
3	1.693	9.405	46.544	
4	1.323	7.351	53.896	
5	1.146	6.364	60.260	
6	.969	5.384	65.644	
7	.925	5.141	70.785	
8	.870	4.832	75.617	
9	.748	4.153	79.770	
10	.635	3.527	83.297	
11	.602	3.343	86.640	
12	.548	3.047	89.687	
13	.424	2.355	92.042	
14	.410	2.277	94.319	
15	.379	2.108	96.426	
16	.290	1.611	98.037	
17	.234	1.299	99.336	
18	.120	.664	100.000	

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Pattern matrix (toont alleen de factor ladingen >= .30)

Pattern Matrix[a]

	Component 1	Component 2
MF_CARE_Q1	.578	
MF_CARE_Q2	.468	
MF_CARE_Q3	.703	
MF_CARE_Q4	.600	
MF_CARE_Q5		.362
MF_CARE_Q6		
MF_FAIR_Q1	.601	-.379
MF_FAIR_Q2	.499	
MF_FAIR_Q3	.548	
MF_FAIR_Q4		
MF_FAIR_Q5	.406	
MF_FAIR_Q6	.359	
MF_LOY_Q1		.370
MF_LOY_Q2	.401	.372
MF_LOY_Q3		.467
MF_LOY_Q4		.704
MF_LOY_Q5		.312
MF_LOY_Q6		.619
MF_AUTH_Q1	.390	.450
MF_AUTH_Q2		.546
MF_AUTH_Q3	.500	.382
MF_AUTH_Q4		.581
MF_AUTH_Q5		
MF_AUTH_Q6	-.307	.524
MF_PUR_Q1	.675	
MF_PUR_Q2	.623	
MF_PUR_Q3		.334
MF_PUR_Q4		.575
MF_PUR_Q5		.479
MF_PUR_Q6		.583

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 12 iterations.

Pattern Matrix[a]

	Component 1	Component 2
MF_CARE_Q1	.520	
MF_CARE_Q2	.555	
MF_CARE_Q3	.728	
MF_CARE_Q4	.598	
MF_FAIR_Q2	.527	
MF_FAIR_Q3	.586	
MF_FAIR_Q6	.383	
MF_LOY_Q3		.464
MF_LOY_Q4		.720
MF_LOY_Q6		.641
MF_AUTH_Q2		.569
MF_AUTH_Q4		.645
MF_AUTH_Q6		.545
MF_PUR_Q1	.685	
MF_PUR_Q2	.665	
MF_PUR_Q4		.643
MF_PUR_Q5		.430
MF_PUR_Q6		.512

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 7 iterations.

After performing this explorative factor analysis for MF_PUR_1 and MF_PUR_2 it appears that they load on the grouping concept for all data. This does not match expectations. The researcher therefore repeated the exploratory factor for only the higher educated. The result of this researcher is as shown in Table 21. The SME for this analysis is .508. In accordance with the advice of Howard (2016), this SME is bad, but this is to be expected given the limited number of respondents. The Barlett's Test of Spericity is significant (< 0.001).

Table 21

Exploratory Factor Analysis

Part	Before removing cross loadings and variables with no factor loadings	After removing cross loadings and variables with no factor loadings

KMO

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.507
Bartlett's Test of Sphericity	Approx. Chi-Square	826.979
	df	435
	Sig.	<.001

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.508
Bartlett's Test of Sphericity	Approx. Chi-Square	402.147
	df	210
	Sig.	<.001

Total variance explained

Total Variance Explained

Component	Initial Eigenvalues Total	% of Variance	Cumulative %	Rotation Sums of Squared Loadings[a] Total
1	5.607	18.689	18.689	4.983
2	3.539	11.796	30.485	4.429
3	2.586	8.620	39.105	
4	2.383	7.945	47.050	
5	2.098	6.994	54.044	
6	1.824	6.082	60.126	
7	1.506	5.020	65.146	
8	1.190	3.967	69.113	
9	1.128	3.759	72.872	
10	1.051	3.503	76.375	
11	.868	2.892	79.266	
12	.807	2.691	81.958	
13	.731	2.436	84.394	
14	.602	2.008	86.402	
15	.555	1.850	88.253	
16	.477	1.589	89.842	
17	.463	1.542	91.384	
18	.400	1.335	92.719	
19	.383	1.277	93.995	
20	.338	1.128	95.124	
21	.298	.993	96.116	
22	.235	.782	96.898	
23	.207	.690	97.588	
24	.170	.566	98.155	
25	.130	.434	98.588	
26	.126	.419	99.008	
27	.111	.369	99.377	
28	.084	.280	99.657	
29	.059	.198	99.855	
30	.044	.145	100.000	

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Total Variance Explained

Component	Initial Eigenvalues Total	% of Variance	Cumulative %	Rotation Sums of Squared Loadings[a] Total
1	4.417	21.033	21.033	4.036
2	2.772	13.198	34.231	3.396
3	2.124	10.114	44.345	
4	1.553	7.393	51.739	
5	1.486	7.078	58.817	
6	1.173	5.586	64.403	
7	1.018	4.849	69.251	
8	.916	4.360	73.611	
9	.828	3.940	77.551	
10	.780	3.715	81.267	
11	.677	3.223	84.489	
12	.570	2.715	87.204	
13	.494	2.354	89.557	
14	.451	2.145	91.703	
15	.376	1.788	93.491	
16	.333	1.584	95.075	
17	.298	1.418	96.492	
18	.269	1.280	97.772	
19	.221	1.051	98.823	
20	.161	.769	99.592	
21	.086	.408	100.000	

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Pattern matrix

Pattern Matrix[a]

	Component 1	Component 2
MF_CARE_Q1		.538
MF_CARE_Q2		.396
MF_CARE_Q3		.521
MF_CARE_Q4		.676
MF_CARE_Q5		.584
MF_CARE_Q6		.381
MF_FAIR_Q1	-.405	.673
MF_FAIR_Q2		.558
MF_FAIR_Q3		.522
MF_FAIR_Q4		.630
MF_FAIR_Q5		.480
MF_FAIR_Q6	-.467	.383
MF_LOY_Q1	.344	
MF_LOY_Q2	.345	
MF_LOY_Q3	.618	
MF_LOY_Q4	.786	
MF_LOY_Q5		
MF_LOY_Q6	.559	
MF_AUTH_Q1	.681	
MF_AUTH_Q2	.614	
MF_AUTH_Q3	.590	.350
MF_AUTH_Q4	.637	
MF_AUTH_Q5		
MF_AUTH_Q6	.461	
MF_PUR_Q1	.408	.415
MF_PUR_Q2	.396	.441
MF_PUR_Q3		
MF_PUR_Q4	.415	
MF_PUR_Q5		
MF_PUR_Q6	.511	

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 10 iterations.

Pattern Matrix[a]

	Component 1	Component 2
MF_CARE_Q1		.556
MF_CARE_Q2		.486
MF_CARE_Q3		.464
MF_CARE_Q4		.760
MF_CARE_Q5		.640
MF_CARE_Q6		.388
MF_FAIR_Q2		.523
MF_FAIR_Q3		.457
MF_FAIR_Q4		.610
MF_FAIR_Q5		.528
MF_LOY_Q1	.364	
MF_LOY_Q2	.358	
MF_LOY_Q3	.655	
MF_LOY_Q4	.793	
MF_LOY_Q6	.601	
MF_AUTH_Q1	.677	
MF_AUTH_Q2	.679	
MF_AUTH_Q4	.642	
MF_AUTH_Q6	.506	
MF_PUR_Q4	.450	
MF_PUR_Q6	.460	

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Appendix E: Normality analysis – test run

The researcher performed a normality analysis for the constructs as found in the exploratory factor analysis. The first construct is the Individualizing moral foundations, consisting of the following items: MF_CARE_Q1 MF_CARE_Q2 MF_CARE_Q3 MF_CARE_Q4 MF_CARE_Q5 MF_CARE_Q6 MF_FAIR_Q2 MF_FAIR_Q3 MF_FAIR_Q4 MF_FAIR_Q5). The second construct is the Grouping moral foundation, consisting of the items: MF_LOY_Q1 MF_LOY_Q2 MF_LOY_Q3 MF_LOY_Q4 MF_LOY_Q6 MF_AUTH_Q1 MF_AUTH_2 MF_AUTH_Q4 MF_AUTH_Q6 MF_PUR_Q4 MF_PUR_Q6.

Figure 11 below provides an overview of the descriptive statistics, where IN_L is the Individualizing moral foundation and GRO_L is the Grouping moral foundation. Hair et al. (2019) provides formulas to calculate the $z_{skewness}$ and the $z_{kurtosis}$. These are listed in Table 22. For a normality distribution the $z_{skewness}$ should be <= 2.58 and the $z_{kurtosis}$ <= 1.96. This is the case for both constructs. Visually, the histograms also indicate that there is a normal distribution for both histograms (see Figure 12 and Figure 13)

Figure 11

Descriptive statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
IN_L	53	2.50	5.70	4.4811	.69005	-.805	.327	.735	.644
GRO_L	53	1.45	4.55	3.3122	.77293	-.437	.327	-.413	.644
Valid N (listwise)	53								

Table 22

$z_{kurtosis}$ and $z_{skewness}$

Item	Individualizing moral foundation	Grouping moral foundation
$z_{skewness}$	-2,3925344	-1,2988044
$z_{kurtosis}$	-1,092244	-0,6137371

Figure 12

Histogram Individualizing construct

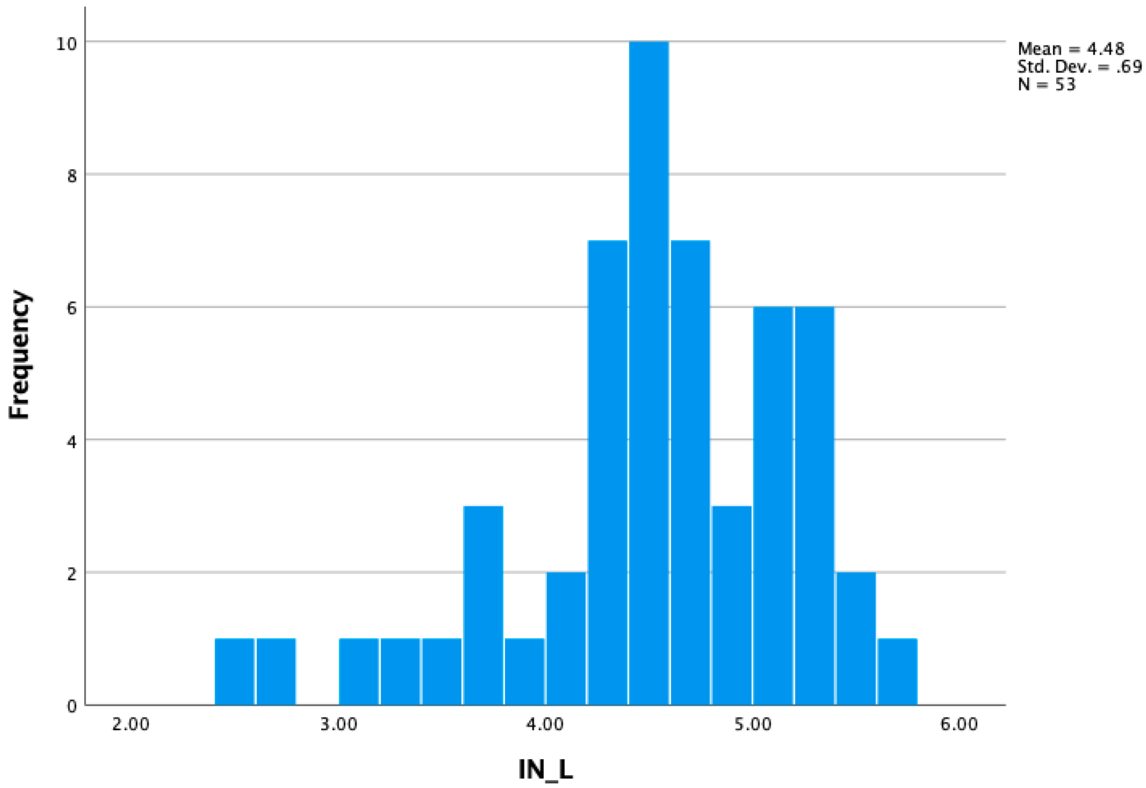
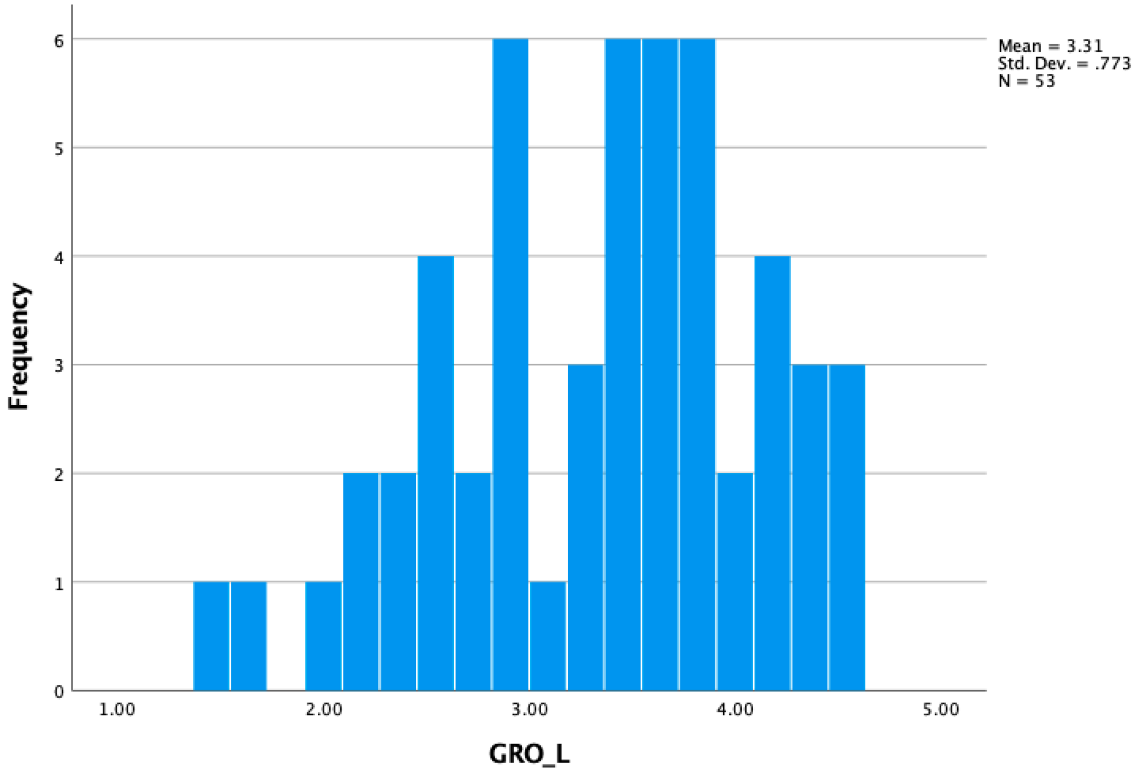


Figure 13

Histogram Grouping construct



Appendix F: Confirmative factor analysis – moral foundations questionnaire

Hair et al. (2019) provide detailed instructions on how to perform a confirmative factor analysis (hereafter CFA). This detailed instruction consists of four different steps that are described in more detail in this appendix.

Step 1: Defining individual constructs

In the first step, the researcher defines the individual constructs. The CFA tests the extent to which these individual constructs occur in the dataset. These constructs follow from the literature. This research is based on the moral foundations of Graham et al. (2011). Graham et al. (2011) performed a CFA for six different models, namely: a single factor model, a two factor model, a three factor model, a five factor model, a six factor model, and a hierarchical model. The research by Nilson and Erlandson (2015) performed a CFA for three models: the five-factor model, the three-factor model, and the hierarchical model. Table 23 gives an overview of the different constructs per model and how the different items of the questionnaire (see Appendix A) load on the constructs.

Table 23

Defining constructs

Items in questionnairre	Single factor model	Second factor model	Three factor model	Five factor model	Six factor model
MF_CARE_Q1	Factor 1	Individualizing	Individualizing	Care	Care
MF_CARE_Q2					
MF_CARE_Q3					
MF_CARE_Q4					
MF_CARE_Q5					
MF_CARE_Q6					
MF_FAIR_Q1				Fair	Fair
MF_FAIR_Q2					
MF_FAIR_Q3					
MF_FAIR_Q4					
MF_FAIR_Q5					
MF_FAIR_Q6					
MF_LOY_Q1		Grouping	Grouping and Authority	Loyalty	Loyalty
MF_LOY_Q2					
MF_LOY_Q3					
MF_LOY_Q4					
MF_LOY_Q5					
MF_LOY_Q6					

MF_AUTH_Q1		Authority	Authority
MF_AUTH_Q2			Tradition
MF_AUTH_Q3			
MF_AUTH_Q4			Authority
MF_AUTH_Q5			Tradition
MF_AUTH_Q6			Authority
MF_PUR_Q1	Purity	Purity	Purity
MF_PUR_Q2			
MF_PUR_Q3			
MF_PUR_Q4			
MF_PUR_Q5			
MF_PUR_Q6			

Step 2: Developing the overall measurement model

In the second step, the researcher creates the model in SPSS AMOS, v. 29.0. When defining the model, the researcher does not assume covariance between the items of the constructs. In addition, the researcher follows the definition of items per construct as given by Graham et al. (2011) and followed by Nilson and Erlandson (2015). The definition of the models is given in Table 21.

Step 3: Designing a study to produce empirical results.

In the third step, the researcher performs the CFA in SPSS and investigates if there are identification or estimation problems. According to Hair et al. (2019), the model may fail due to negative factor loadings or Heywood cases (an error variance estimate less than zero). After running the model, the researcher tests whether this is the case. This was not the case in the CFA analyses performed in this study.

The researcher performs the CFA with different datasets, namely Dataset 5 (after data cleaning), Dataset 6 (only highly educated people) and Dataset 7 (only lower educated people). See Appendix X for a detailed description of the datasets.

Step 4: Assessing measurement model validity

In the fourth step, the model validity is tested. Table 24 gives the results of the performed CFA analyzes for 4 different models for all data. It follows from this analysis that the model fit is lower than the prescribed norms from the literature as shown in Table 24. When the researcher performs the same analysis for the data of the higher educated (Dataset6) and the lower educated (Dataset7), comparable results are obtained.

Table 24

Results Confirmative Factor Analysis (All data)

Item	Baseline according literature	1-factor model with all data	2-factor model with all data	5-factor with all data	Hierarchical model with all data	2-factor model (first-3-items)	5-factor-model (first 3-items)
Model		1-factor model	2-factor model	5-factor model	Hierarchical model	2-factor model (first 3 items)	5-factor model (first 3 items)
χ^2 and degrees of freedom	N.a.	χ^2 (405) = 4818	χ^2 (404) = 4687	χ^2 (395) = 4265	χ^2 (399) = 4420	χ^2 (89) = 1658	χ^2 (80) = 1289
χ^2 / df	>= 5 is a reasonable fit according to Marsh and Hocevar (1985).	11.9	11.6	10.8	11.1	18.6	16.1
Goodness of Fit Index	GFI >= .90 according to Hair et al. (2019)	.696	.723	.733	.725	.808	.849
Comparative Fit Index (CFI)	CFI above .90 according to Bentler (1990) and Carlson and Mulaik (1993)	.477	.492	.541	.523	.701	.770
Root Mean Square Error of Approximation (RMSEA)	Value of .06 to .10 is required for model fit according to Hu & Bentler (1999) and Ridgon (1996).	.099 (90% CI [.096, .101])	.097 (90% CI [.095, .100])	.094 (90% CI [.091, .096])	.095 (90% CI [.092, .098])	.126 (90% CI [.120, .131])	.116 (90% CI [.111, .122])
Factor loadings		See Figure	See Figure	See Figure	See	See Figure	See Figure

The figures below show the factor loadings for the six different models.

Figure 14

Factor loadings, one factor-model, all data

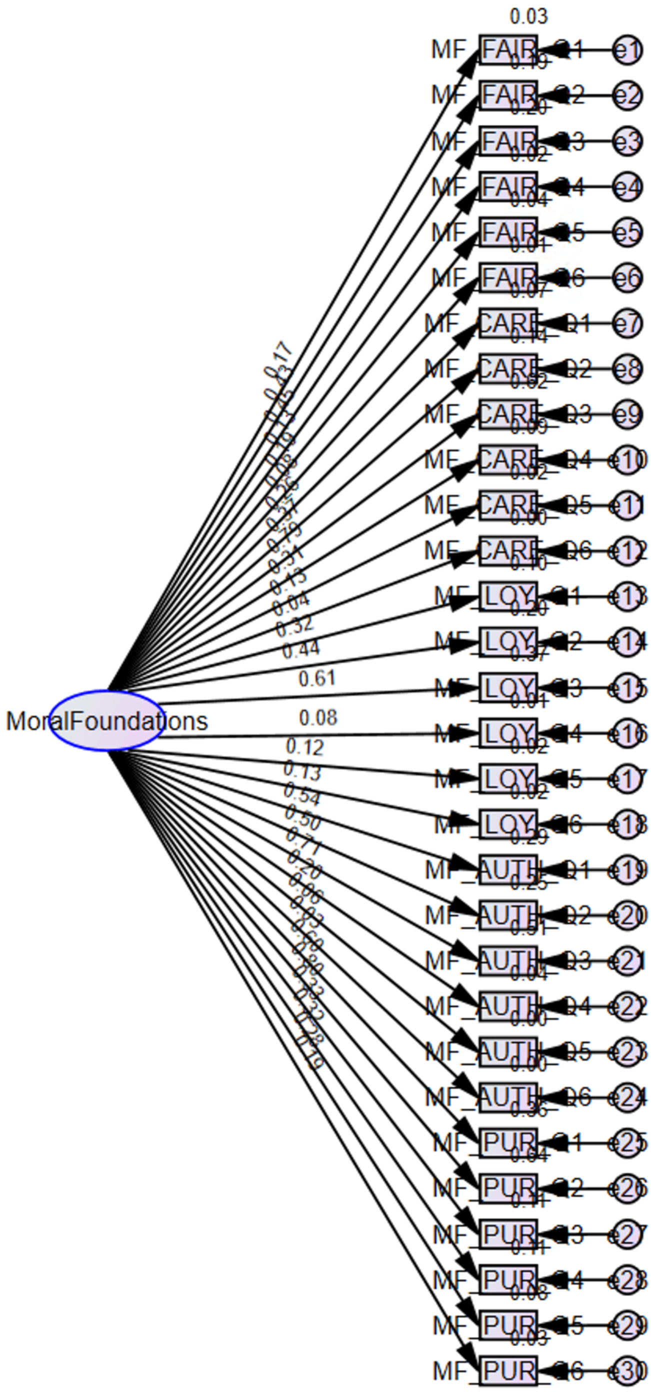


Figure 15

Factor loadings, 2-factor model, all data

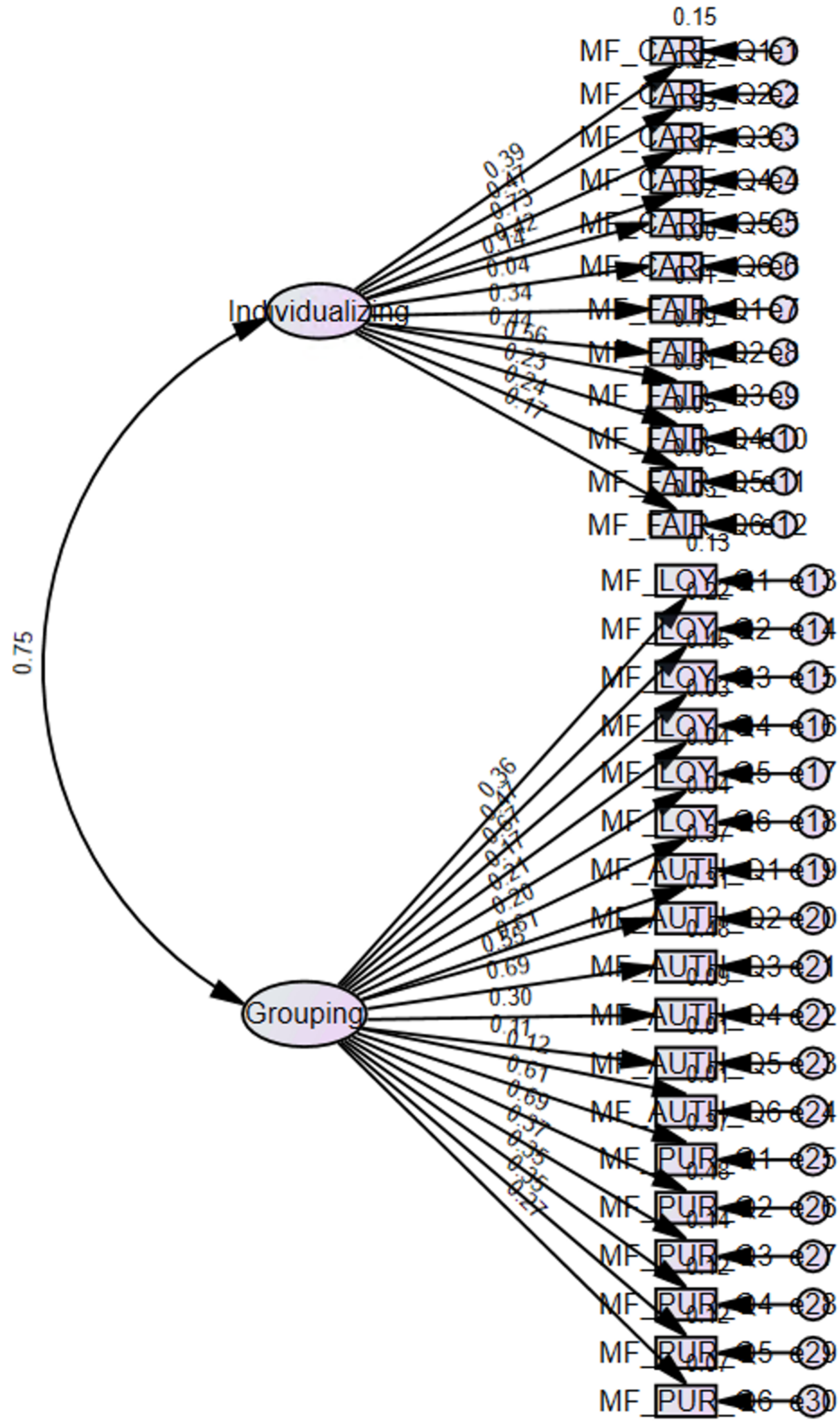


Figure 16

Factor loadings, 5-factor model, all data

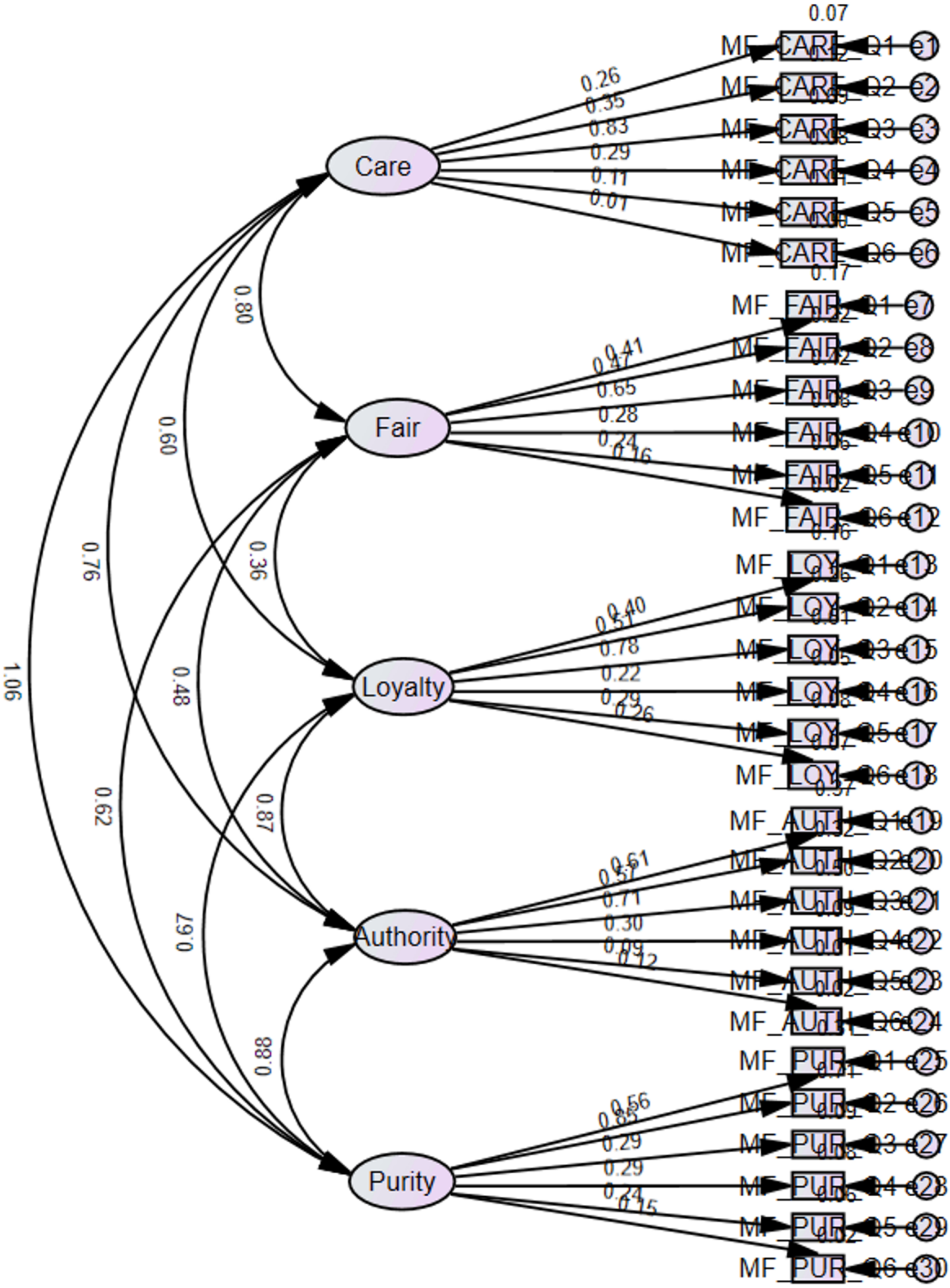


Figure 17

Factor loadings, hierarchical model

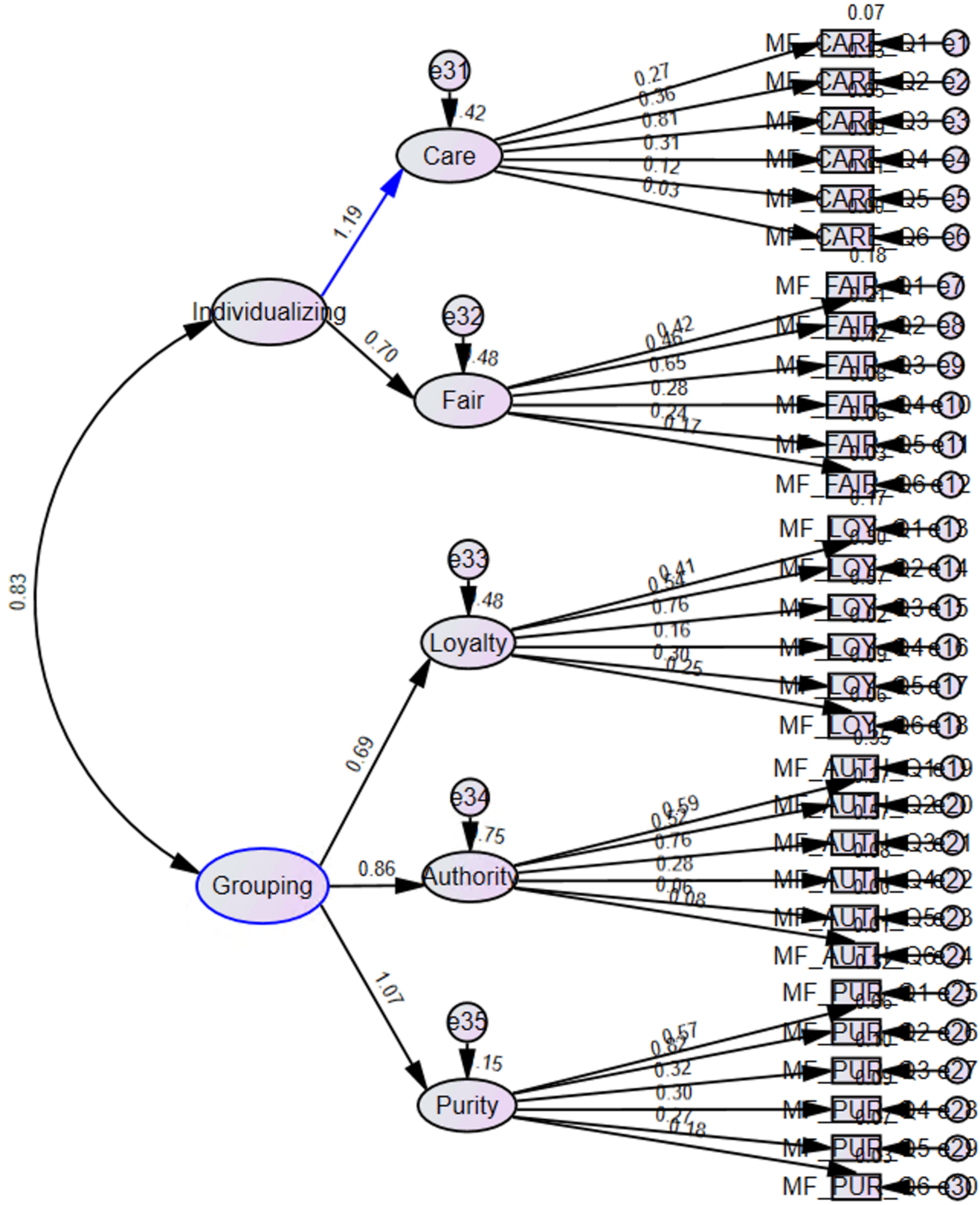


Figure 18

Factor loadings, 2-factor model (three items)

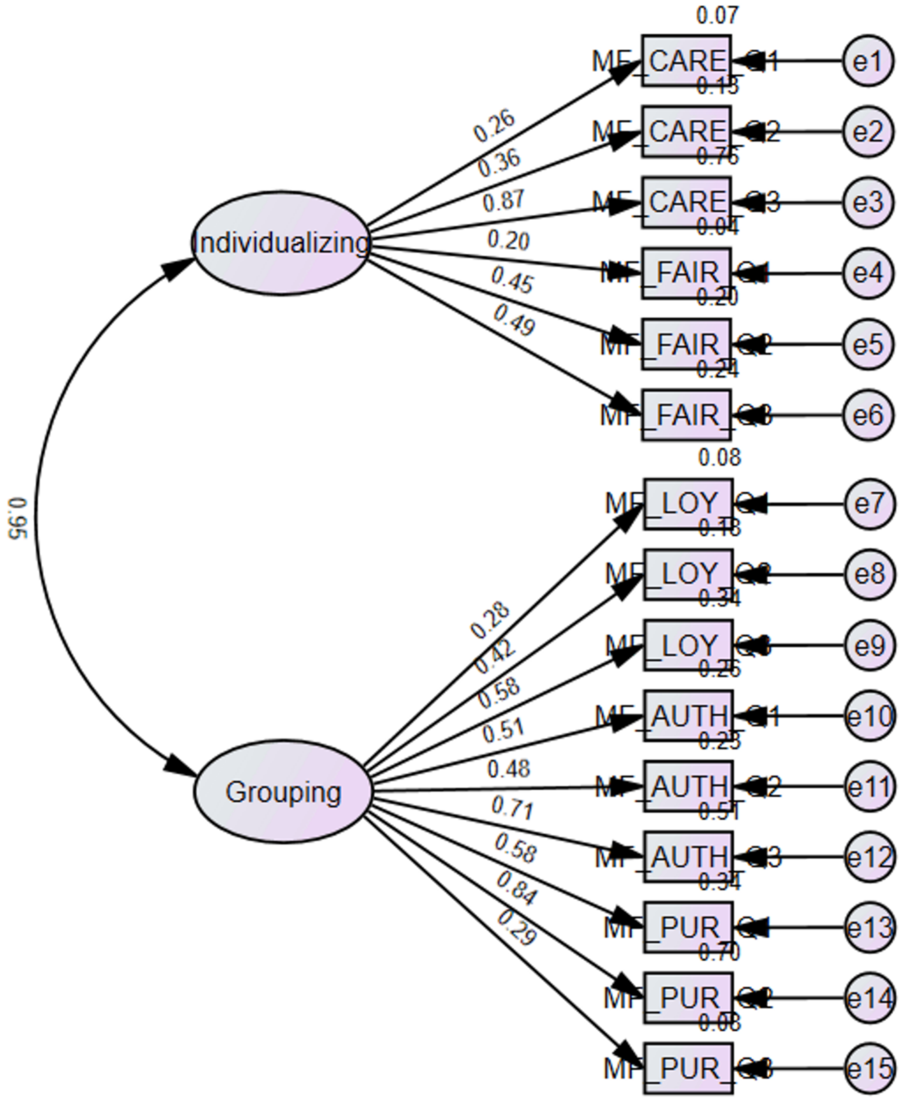
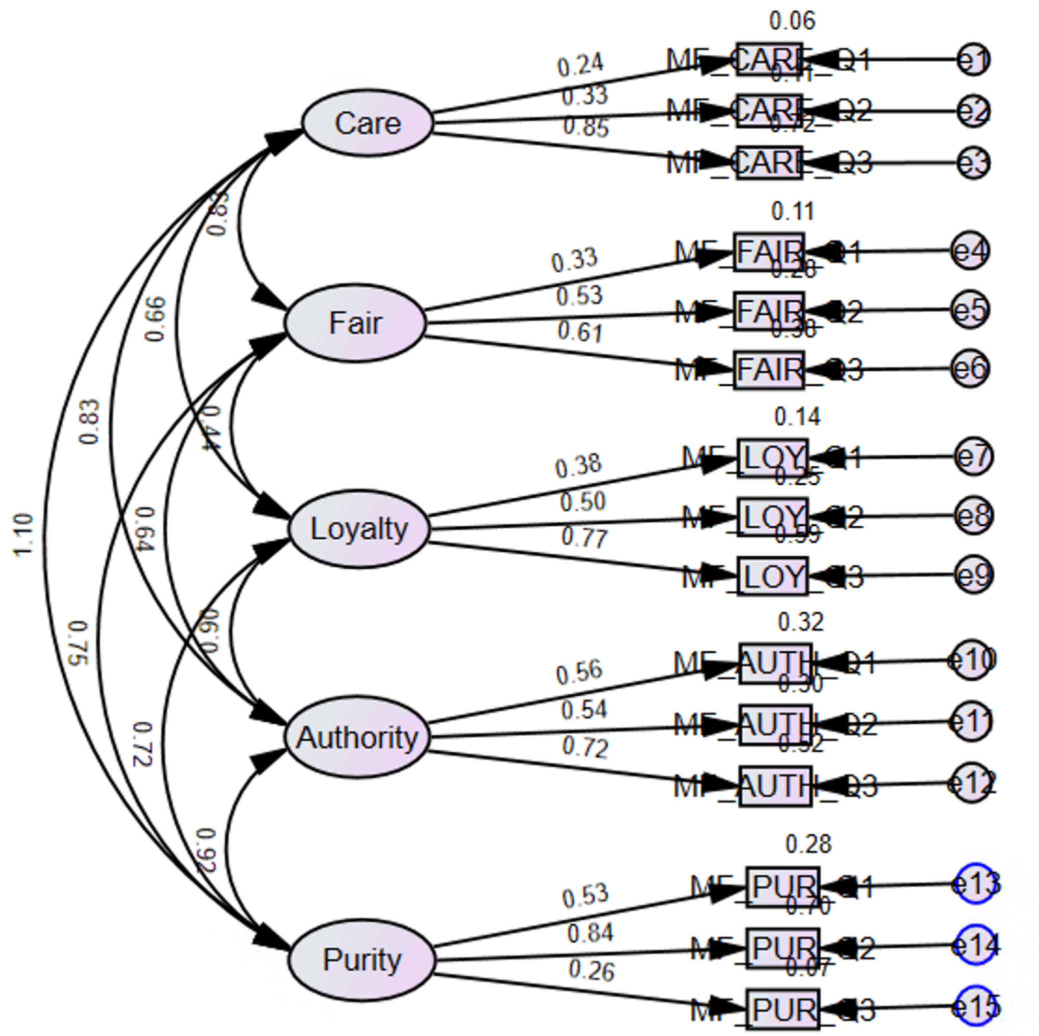


Figure 19

Factor loadings, 5-factor model (three items)



The researcher compares these results with the results found by Nilson and Erlandson (2015) and Graham et al. (2011) as shown in Table 25. From this comparison, it follows that the model fit in this study is lower than in the studies by Graham et al. (2011) and Nilson and Erlandson (2015).

Table 25

Comparison CFA of current research with Nilson and Erlandson (2015) and Graham et al. (2011)

Item	Starting point	5-factor model and data of Nilsson and Erlandson (2015)	2-factor model of Nilsson and Erlandson (2015)	2-factor model of current research	5-factor model of current research
Model		5-factor model	2-factor model	2-factor model	5-factor model

Data		Highly educated people in Denmark.	Highly educated people in Denmark.	Only highly educated people	Only highly educated people
χ^2 and degrees of freedom		χ^2 (396) = 1517.8	χ^2 (400) = 1540.8	χ^2 (404) = 5447	χ^2 (395) = 4992
χ^2 / df	>= 5 is a reasonable fit according to Marsh and Hocevar (1985).	3.8	3.9	13.9	12.6
Comparative Fit Index (CFI)	CFI above .90 according to Bentler (1990) and Carlson and Mulaik (1993)	.679	.674	.505	.549
Root Mean Square Error of Approximation (RMSEA)	Value of .06 to .10 is required for model fit according to Hu & Bentler (1999) and Ridgon (1996).	.072 (95% CI [.069, .076])	.073 (95% CI [.069, .077]).	.099 (90% CI [.096, .101])	.095 (90% CI [.093, .097])

F1 Model fit improvements

Based on the results of this CFA, the researcher decides to implement several improvements in the model to improve the model fit and reduce the sources of misfit. Based on Table 25 the researcher chooses to start with a model containing only the relevance items from the questionnaire, because these models have a significantly higher model fit. There are two options: the two-factor model or the five-factor model. The researcher chooses the two-factor model based on two arguments: 1) after solving the sources of misfit in the 5-factor model, too few items remain per concept, 2) and the 2-factor model uses the individualizing and grouping concepts as used in the research.

shows the two-factor model with only the first items for resolving the sources of misfit. The researcher has fixed the following sources of misfit: The researcher removes MF_CARE_Q3, because it leads to a high discriminant validity based on the criteria of Fornell and Larcker (1981). MF_CARE_Q3 also leads to a low convergent validity. Based on Hair et al. (2019), it is possible to remove items that have a lower factor loading. The researcher removes the items MF_FAIR_Q2, MF_FAIR_Q3, MF_AUTH_Q1, MF_AUTH_Q2, MF_LOY_Q1, MF_LOY_Q2 and MF_PUR_Q3 because they have too low factor loading (<.55). shows the two-factor model after solving the sources of misfit. Table 26 shows the model fit measures of the 2-factor model with only the first three items after solving the sources of misfit.

Figure 20

2-factor model with only the first three items

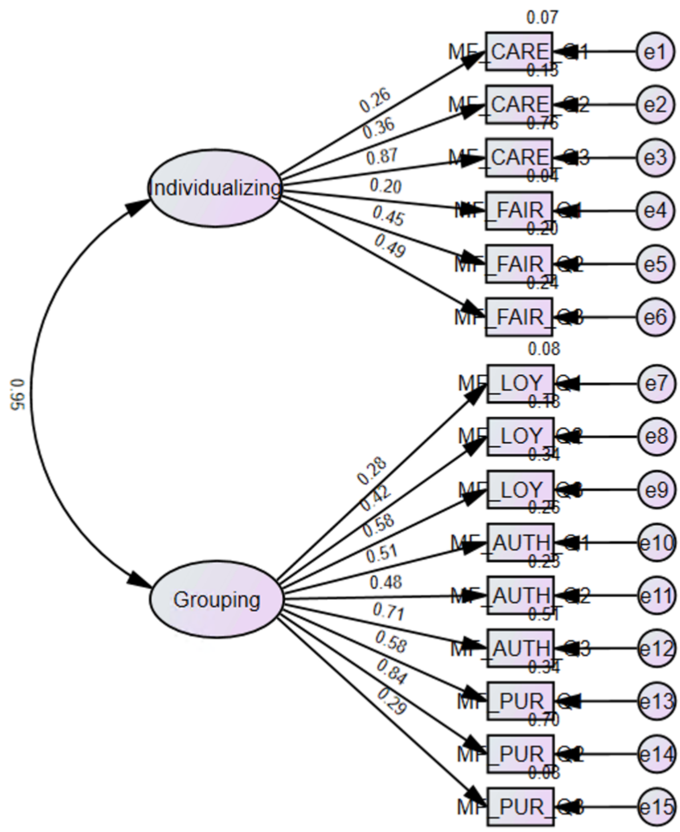


Table 26

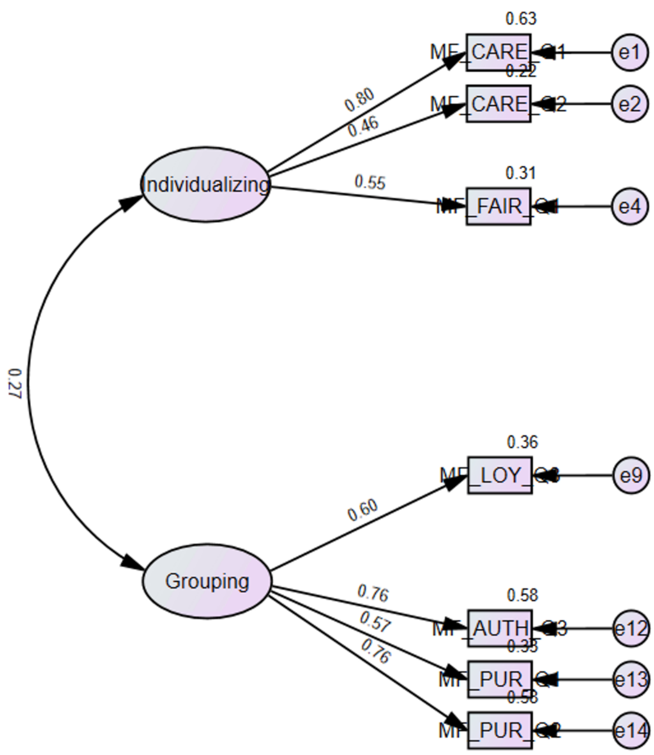
Results Confirmative Factor Analysis (All data)

Item	Baseline according literature	Optimized 2-factor model with only first three items with all data (Dataset5)
Model		2 factor model
χ^2 and degrees of freedom	N.a.	χ^2 (13) = 86
χ^2 / df	Value of .06 to .10 is required for model fit according to Hu & Bentler (1999) and Ridgon (1996).	6.6
Goodness of Fit Index	Value of .06 to .10 is required for model fit according to Hu & Bentler (1999) and Ridgon (1996).	.978

Comparative Fit Index (CFI)	Value of .06 to .10 is required for model fit according to Hu & Bentler (1999) and Ridgon (1996).	.956
Root Mean Square Error of Approximation (RMSEA)	Value of .06 to .10 is required for model fit according to Hu & Bentler (1999) and Ridgon (1996).	.071 (90% CI [.057., .086])
Factor loadings		See Figure 21.

Figure 21

2-factor model with only the first three items (after solving sources of misfit)



Appendix G: Normality analysis – real run

The researcher performed a normality analysis for the constructs as found in the confirmatory factor analysis. The first construct is the individualizing moral foundations, consisting of the following items: MF_CARE_Q1, MF_CARE_Q2, and MF_FAIR_Q1. The second construct is the grouping moral foundation, consisting of the items: MF_AUTH_Q1, MF_AUTH_Q3, MF_LOY_Q3, MF_PUR_Q1, MF_PUR_Q2.

The figure below provides an overview of the descriptive statistics, where IND is the Individualizing moral foundation and GRO is the grouping moral foundation. Hair et al. (2019) provides formulas to calculate the $z_{skewness}$ and the $z_{kurtosis}$. These are listed in Table 27. For a normality distribution the $z_{skewness}$ should be <= 2.58 and the $z_{kurtosis}$ <= 1.96. Only the $z_{kurtosis}$ falls outside the boundaries, the other items are within the boundaries for both constructs. Visually, the histograms also indicate that there is a normal distribution for both histograms.

Figure 22

Descriptive statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
IND	1118	1.00	6.00	3.6983	1.02655	-.118	.073	-.265	.146
GRO	1118	1.00	6.00	4.0461	1.09778	-.571	.073	.077	.146
Valid N (listwise)	1118								

Table 27

zkurtosis and zskewness

Item	Individualizing moral foundation	Grouping moral foundation
$z_{skewness}$	-1.611	-7.794
$z_{kurtosis}$	-1.809	-0.523

Figure 23

Histogram Individualizing construct

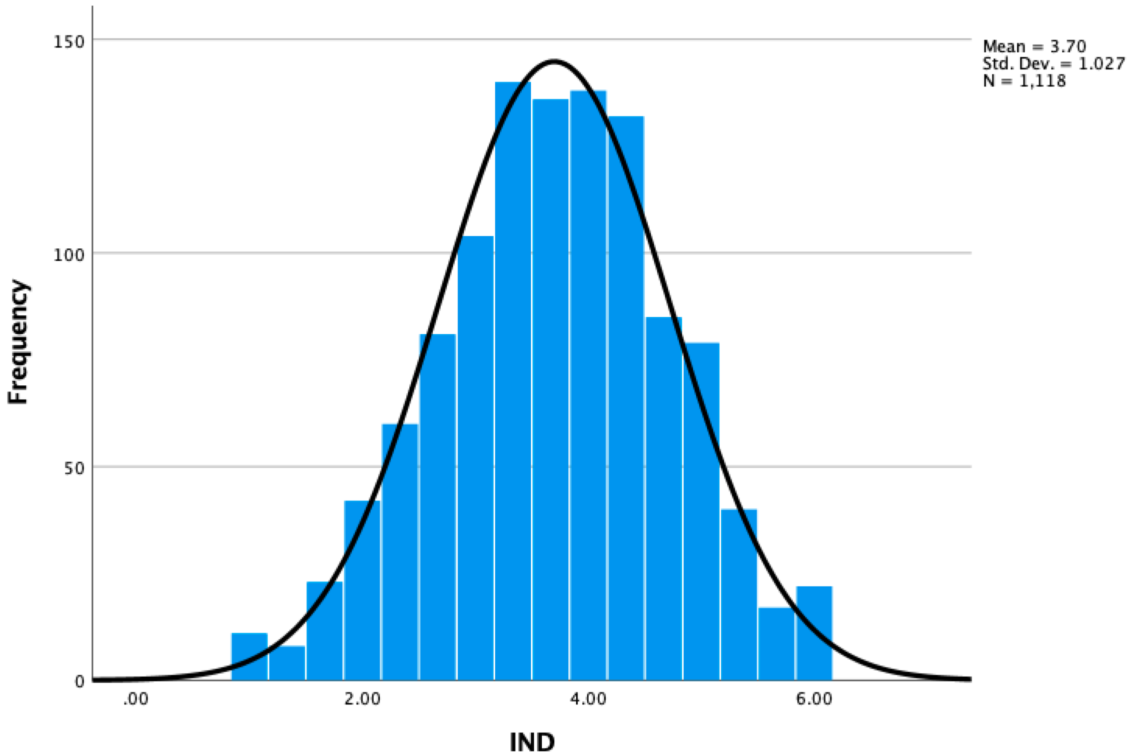
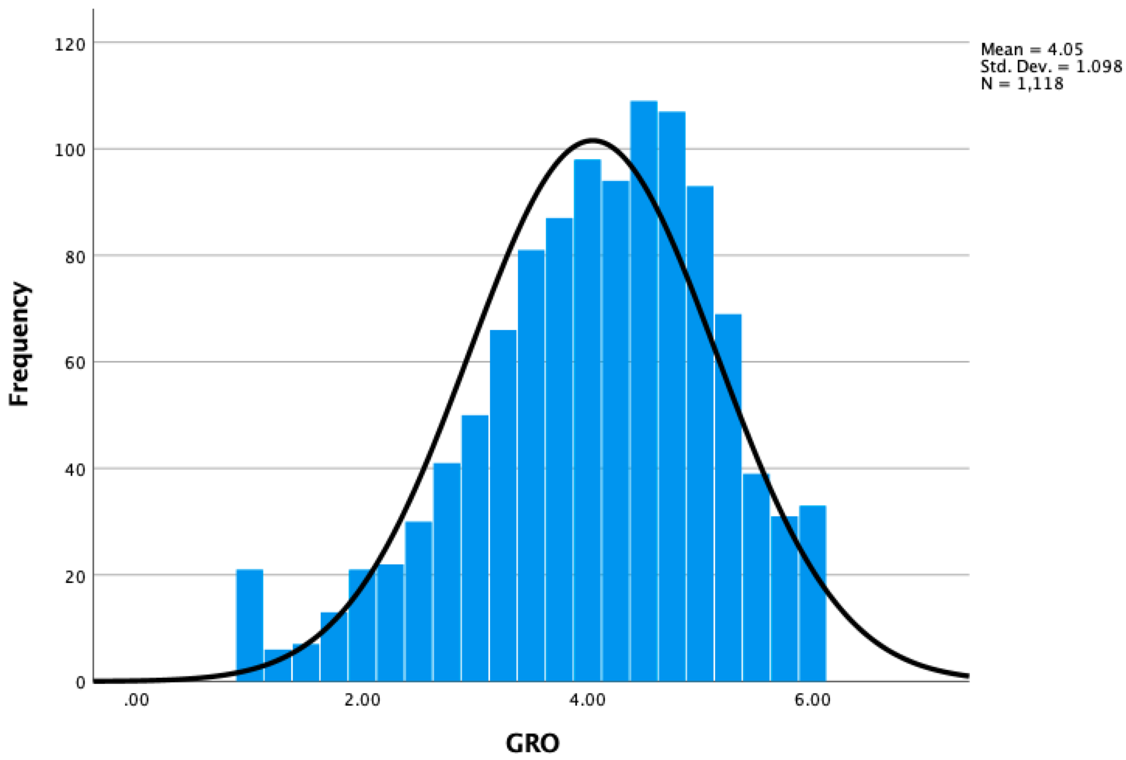


Figure 24

Histogram Grouping construct



Appendix H: Regression analysis

Hair et al. (2019) provide a step-by-step plan consisting of six steps to perform a multiple regression analysis. This appendix provides a description of the steps performed in the multiple regression analysis and describes the choices made.

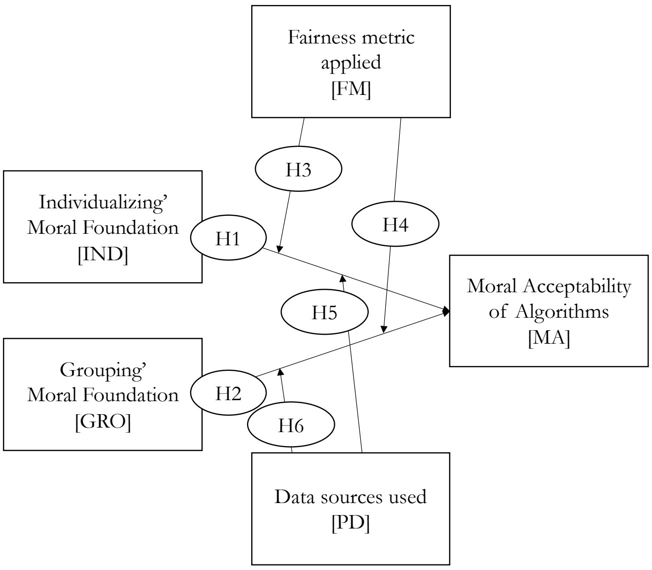
Stage 1: Objectives of Multiple Regression

According to Hair et al. (2019), multiple regression analysis is used for two purposes, namely prediction and explanation. In the explanation, the model is used to verify a theoretical model in the data. In prediction, the model is used to predict a dependent variable with an independent variable. For this application, it is about explaining a relationship between the independent variables and the dependent variables.

An important starting point is that there is a theoretical basis that proves that there is a certain relationship. This theoretical basis is described in the theoretical framework. The constructs used by the researcher in this multiple regression analysis are shown in the figure below.

Figure 25

Conceptual model



Stage 2: Research design of multiple regression

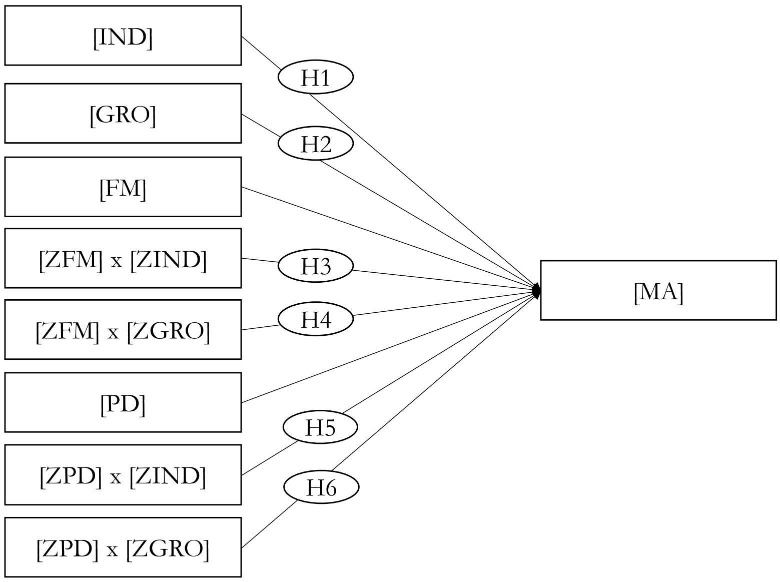
During the research design, several choices must be made about the sample size. The choice of sample size depends on predictable significance. The researcher assumes a significance of .05. According to Hair et al. (2019), with 5 independent variables and a sample size of at least 1000, an R2 of at least 1 can be predicted. In short, the sample that the researcher has available for this research is sufficient.

In addition, the conceptual model contains two moderators, and a moderator analysis must therefore be performed. The Hayes plugin is available in SPSS for moderator analyses; however, the conceptual model used in this study is not available in this Hayes plugin.

The researcher translates the conceptual model as shown in Figure 26 into the statistical model as shown in Figure . The variables ZFM, ZIND, ZGRO, ZPD are the z-scores of FM, IND, GRO and PD, respectively. The variable Moral Acceptability is the score a respondent has given on 1 of the vignettes as described in the questionnaire (see Appendix A).

Figure 26

Statististical model



Stage 3: Assumptions in Multiple Regression Analysis

Prior to a multiple regression analysis, according to Hair et al. (2019), several basic conditions must be tested, namely normality, linearity, homoscedasticity, and independence of the error terms.

Regarding normality Hayes et al. (2012) indicate that the assumption of normality is the least important assumption for multiple regression because normality is often not met because variables are dichotomous. This is the case for fairness metrics [FM] and public data sources [PD] used. The grouping and individualizing variables are normally distributed (see Appendix H). In addition, the moral acceptability [MA] is also not normally distributed; this is because many respondents score 100 or 0 on 1 of the cases. This is not problematic because this assumption is rarely met, as indicated by Hayes et al. (2012).

In relation to linearity, the researcher made several scatter dots in SPSS to test whether there is linearity between the independent and dependent variables. Based on a visual inspection, the researcher concludes that there is linearity in these cases. See Figure 27 and Figure 28 for a screenshot from SPSS.

Figure 27

Visual inspection linearity grouping [GRO] versus moral acceptability [MA]

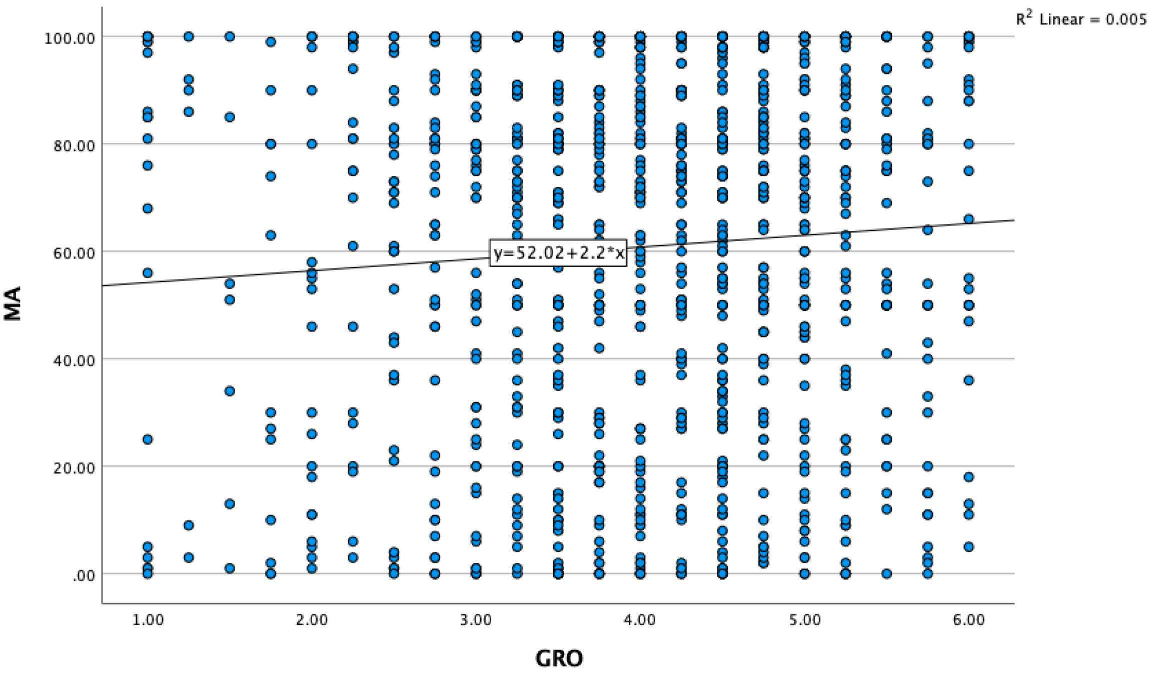
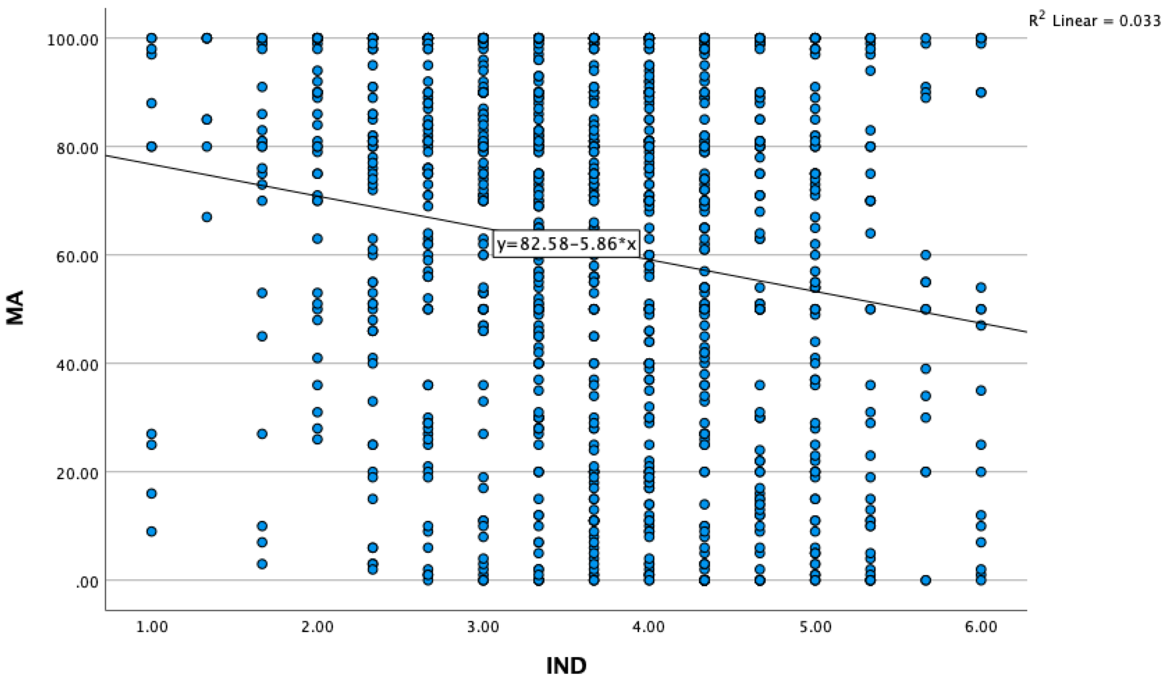


Figure 28

Visual inspection linearity individualizing [IND] versus moral acceptability [MA]



Regarding to the test for homoscedasticity, the researcher visually inspects the plot in which moral acceptability is plotted on the standardized residual for the specific variables. shows the plot of the standardized residual plot of the individualizing variable on moral acceptability. shows the plot of the standardized residual plot of the grouping variable on moral acceptability. A visual inspection shows homoscedasticity.

Figure 29

Standardized residual plot Individualing on Moral Acceptability

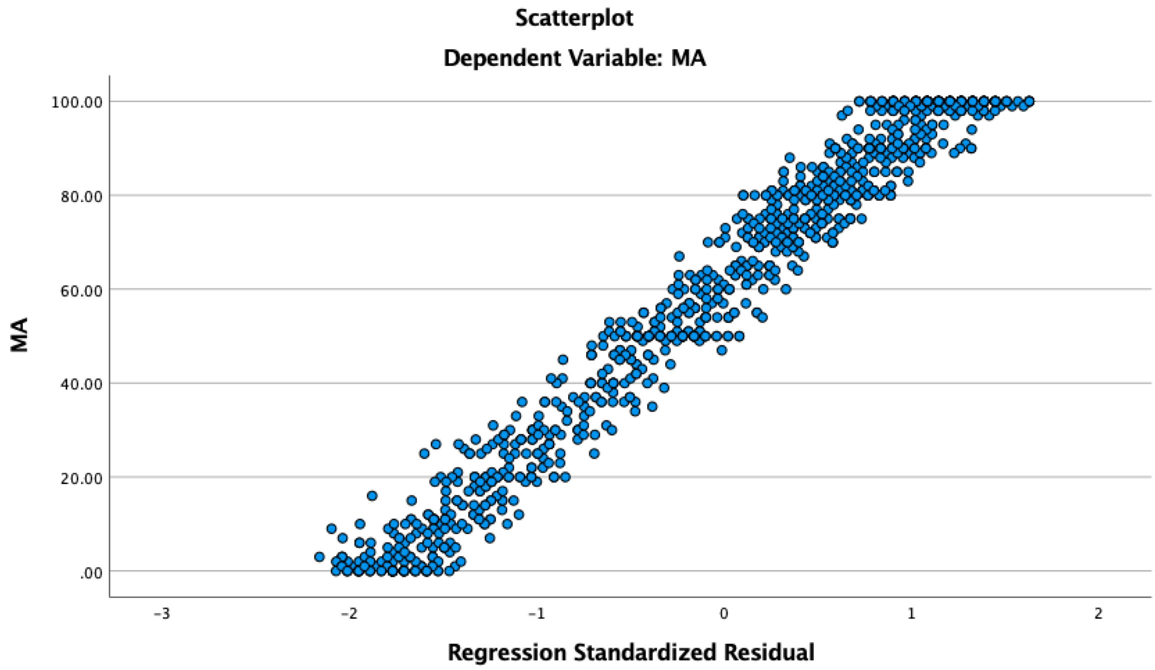
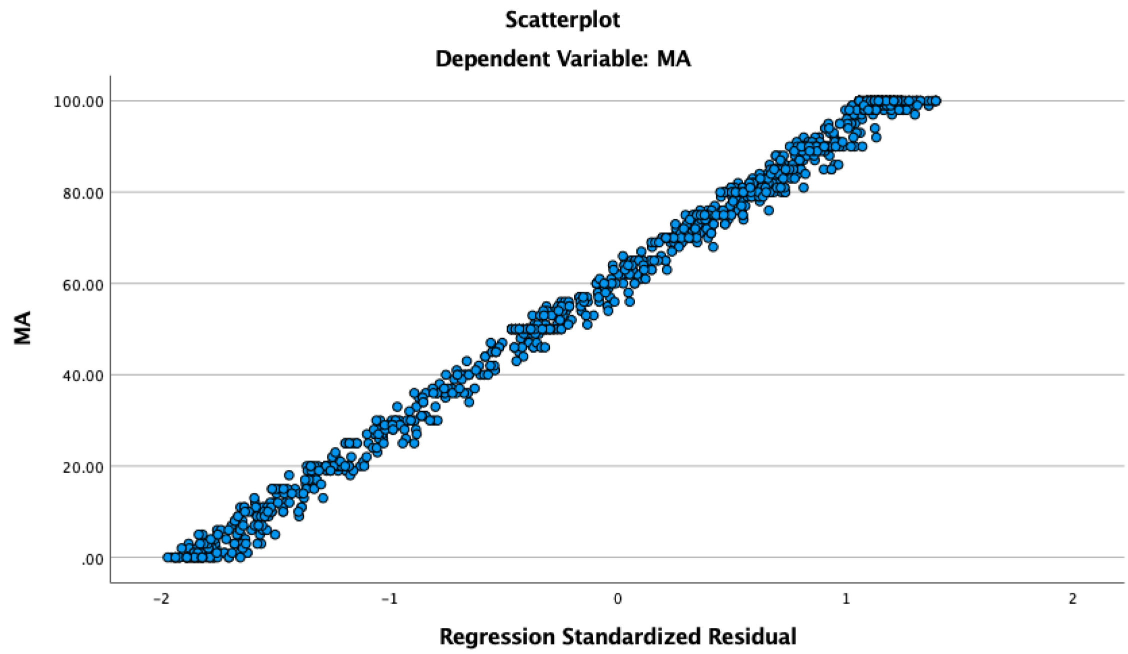


Figure 30

Standardized residual plot Grouping on Moral Acceptability



Stage 4: Estimating the regression model

The researcher performs a hierarchical multiple regression analysis. The researcher examines 9 models as shown in Table 28. In each model the researcher adds new independent variables as shown in the statistical model in Figure . In this way it is possible to gain insight into the effect of individual independent variables.

The variables INDxFM, GROxFM, INDxPD, and GROxPD are computed variables and are calculated as follows: INDxFM = z-score of IND times z-score of FM; GROxFM = z-score of GRO times z-score of FM; INDxPD = z-score of IND times z-score or PD; and GROxPD = z-score of GRO times z-score or PD.

Table 28

Hierarchical regression models

Model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Depedendent	MA	MA	MA	MA	MA	MA	MA	MA	MA
Independent	AGE	AGE IND	AGE IND GRO	AGE IND GRO FM	AGE IND GRO FM INDFM	AGE IND GRO FM INDFM GROFM	AGE IND GRO FM INDFM GROFM PD	AGE IND GRO FM INDFM GROFM PD INDPD	AGE IND GRO FM INDFM GROFM PD INDPD GROPD

The researcher uses the syntax below in SPSS to perform the hierarchical multiple regression.

```
DATASET ACTIVATE DataSet1.
REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT MA
  /METHOD=ENTER AGE
  /METHOD=ENTER AGE IND
  /METHOD=ENTER AGE IND GRO
  /METHOD=ENTER AGE IND GRO FM
  /METHOD=ENTER AGE IND GRO FM INDxFM
  /METHOD=ENTER AGE IND GRO FM INDxFM GROxFM
  /METHOD=ENTER AGE IND GRO FM INDxFM GROxFM PD
  /METHOD=ENTER AGE IND GRO FM INDxFM GROxFM PD INDxPD
  /METHOD=ENTER AGE IND GRO FM INDxFM GROxFM PD INDxPD GROxPD
  /SCATTERPLOT=(MA ,*ZRESID)
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

Stage 5: Interpreting the regression Variable

Figure 30 provides an overview of the hierarchical multiple regression analysis performed.

Figure 31

Model summary

Model Summary[j]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.002[a]	.000	-.001	32.89645	.000	.006	1	1116	.937
2	.183[b]	.033	.032	32.35574	.033	38.611	1	1115	<.001
3	.220[c]	.048	.046	32.12195	.015	17.290	1	1114	<.001
4	.361[d]	.131	.128	30.71391	.082	105.481	1	1113	<.001
5	.365[e]	.133	.129	30.68598	.002	3.027	1	1112	.082
6	.369[f]	.136	.132	30.63832	.003	4.462	1	1111	.035
7	.396[g]	.157	.152	30.28525	.021	27.056	1	1110	<.001
8	.400[h]	.160	.154	30.23968	.003	4.348	1	1109	.037
9	.402[i]	.161	.154	30.23616	.001	1.259	1	1108	.262

a. Predictors: (Constant), AGE

b. Predictors: (Constant), AGE, IND

c. Predictors: (Constant), AGE, IND, GRO

d. Predictors: (Constant), AGE, IND, GRO, FM

e. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM

f. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM, GROxFM

g. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM, GROxFM, PD

h. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM, GROxFM, PD, INDxPD

i. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM, GROxFM, PD, INDxPD, GROxPD

j. Dependent Variable: MA

Figure 32 gives an overview of the ANOVA table from SPSS.

Figure 32

ANOVA Table

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.756	1	6.756	.006	.937[b]
	Residual	1207709.12	1116	1082.177		
	Total	1207715.88	1117			
2	Regression	40428.849	2	20214.424	19.309	<.001[c]
	Residual	1167287.03	1115	1046.894		
	Total	1207715.88	1117			
3	Regression	58268.949	3	19422.983	18.824	<.001[d]
	Residual	1149446.93	1114	1031.820		
	Total	1207715.88	1117			
4	Regression	157773.668	4	39443.417	41.812	<.001[e]
	Residual	1049942.21	1113	943.344		
	Total	1207715.88	1117			
5	Regression	160623.915	5	32124.783	34.116	<.001[f]
	Residual	1047091.96	1112	941.629		
	Total	1207715.88	1117			
6	Regression	164812.641	6	27468.773	29.262	<.001[g]
	Residual	1042903.23	1111	938.707		
	Total	1207715.88	1117			
7	Regression	189627.871	7	27089.696	29.535	<.001[h]
	Residual	1018088.00	1110	917.196		
	Total	1207715.88	1117			
8	Regression	193603.573	8	24200.447	26.465	<.001[i]
	Residual	1014112.30	1109	914.439		
	Total	1207715.88	1117			
9	Regression	194754.253	9	21639.361	23.670	<.001[j]
	Residual	1012961.62	1108	914.225		
	Total	1207715.88	1117			

a. Dependent Variable: MA
b. Predictors: (Constant), AGE
c. Predictors: (Constant), AGE, IND
d. Predictors: (Constant), AGE, IND, GRO
e. Predictors: (Constant), AGE, IND, GRO, FM
f. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM
g. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM, GROxFM
h. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM, GROxFM, PD
i. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM, GROxFM, PD, INDxPD
j. Predictors: (Constant), AGE, IND, GRO, FM, INDxFM, GROxFM, PD, INDxPD, GROxPD

Figure 33 gives an overview of the coefficients of the different models.

Figure 33

Coefficients

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	60.423	6.240		9.684	<.001					
	AGE	.007	.093	.002	.079	.937	.002	.002	.002	1.000	1.000
2	(Constant)	82.538	7.094		11.634	<.001					
	AGE	.001	.091	.000	.008	.994	.002	.000	.000	1.000	1.000
	IND	-5.860	.943	-.183	-6.214	<.001	-.183	-.183	-.183	1.000	1.000
3	(Constant)	71.015	7.569		9.383	<.001					
	AGE	-7.277E-5	.091	.000	-.001	.999	.002	.000	.000	1.000	1.000
	IND	-6.837	.965	-.213	-7.083	<.001	-.183	-.208	-.207	.941	1.063
	GRO	3.753	.903	.125	4.158	<.001	.073	.124	.122	.941	1.063
4	(Constant)	62.507	7.284		8.581	<.001					
	AGE	-.003	.087	-.001	-.035	.972	.002	-.001	-.001	1.000	1.000
	IND	-7.118	.923	-.222	-7.708	<.001	-.183	-.225	-.215	.940	1.064
	GRO	3.782	.863	.126	4.382	<.001	.073	.130	.122	.941	1.063
	FM	18.880	1.838	.287	10.270	<.001	.281	.294	.287	.999	1.001
5	(Constant)	62.421	7.278		8.577	<.001					
	AGE	.000	.087	.000	-.003	.997	.002	.000	.000	1.000	1.000
	IND	-7.144	.923	-.223	-7.742	<.001	-.183	-.226	-.216	.940	1.064
	GRO	3.770	.862	.126	4.372	<.001	.073	.130	.122	.941	1.063
	FM	18.886	1.837	.287	10.283	<.001	.281	.295	.287	.999	1.001
	INDxFM	1.600	.920	.049	1.740	.082	.046	.052	.049	.999	1.001
6	(Constant)	62.562	7.267		8.609	<.001					
	AGE	-.002	.087	-.001	-.028	.978	.002	-.001	-.001	.999	1.001
	IND	-7.123	.921	-.222	-7.732	<.001	-.183	-.226	-.216	.939	1.064
	GRO	3.749	.861	.125	4.355	<.001	.073	.130	.121	.941	1.063
	FM	18.885	1.834	.287	10.298	<.001	.281	.295	.287	.999	1.001
	INDxFM	2.085	.947	.063	2.203	.028	.046	.066	.061	.940	1.063
	GROxFM	-1.997	.946	-.061	-2.112	.035	-.049	-.063	-.059	.941	1.063
7	(Constant)	67.264	7.240		9.291	<.001					
	AGE	-.009	.086	-.003	-.103	.918	.002	-.003	-.003	.999	1.001
	IND	-7.045	.911	-.220	-7.735	<.001	-.183	-.226	-.213	.939	1.065
	GRO	3.787	.851	.126	4.450	<.001	.073	.132	.123	.940	1.063
	FM	19.179	1.814	.292	10.576	<.001	.281	.303	.291	.998	1.002
	INDxFM	2.319	.937	.070	2.476	.013	.046	.074	.068	.938	1.066
	GROxFM	-2.010	.935	-.061	-2.151	.032	-.049	-.064	-.059	.941	1.063
	PD	-9.447	1.816	-.144	-5.201	<.001	-.134	-.154	-.143	.996	1.004
8	(Constant)	67.783	7.233		9.371	<.001					
	AGE	-.012	.086	-.004	-.145	.885	.002	-.004	-.004	.999	1.001
	IND	-7.225	.914	-.226	-7.909	<.001	-.183	-.231	-.218	.931	1.074
	GRO	3.867	.851	.129	4.545	<.001	.073	.135	.125	.939	1.065
	FM	19.377	1.813	.295	10.686	<.001	.281	.306	.294	.995	1.005
	INDxFM	2.301	.935	.070	2.460	.014	.046	.074	.068	.938	1.066
	GROxFM	-1.945	.934	-.059	-2.082	.038	-.049	-.062	-.057	.940	1.064
	PD	-9.453	1.813	-.144	-5.213	<.001	-.134	-.155	-.143	.996	1.004
	INDxPD	-1.895	.909	-.058	-2.085	.037	-.023	-.062	-.057	.987	1.013
9	(Constant)	67.627	7.233		9.349	<.001					
	AGE	-.012	.086	-.004	-.136	.892	.002	-.004	-.004	.999	1.001
	IND	-7.289	.915	-.228	-7.964	<.001	-.183	-.233	-.219	.927	1.078
	GRO	3.948	.854	.132	4.624	<.001	.073	.138	.127	.932	1.073
	FM	19.385	1.813	.295	10.692	<.001	.281	.306	.294	.995	1.005
	INDxFM	2.268	.936	.069	2.424	.016	.046	.073	.067	.937	1.067
	GROxFM	-1.945	.934	-.059	-2.083	.037	-.049	-.062	-.057	.940	1.064
	PD	-9.449	1.813	-.144	-5.211	<.001	-.134	-.155	-.143	.996	1.004
	INDxPD	-2.155	.938	-.066	-2.298	.022	-.023	-.069	-.063	.927	1.079
	GROxPD	1.049	.935	.032	1.122	.262	.007	.034	.031	.933	1.072

a. Dependent Variable: MA