Erasmus University Rotterdam

Erasmus School of Economics

MSc Economics and Business: Data Science and Marketing Analytics

# Mind the gap

## A quantitative analysis to bridge the attitude-behaviour gap among adolescents in terms of plastic consumption

*Author:*

Danks, C.

524733

*Supervisor:*

Jong, M.G. de

*Second assessor:*

Groenen, P.J.F.

A thesis submitted in fulfilment of the requirements for the degree of

MASTER OF ECONOMICS AND BUSINESS

Submission Date: 22nd of July 2020

# Executive Summary

**Purpose:** Adolescents and young adults represent of the generation Z are increasingly aware of the destructive consequences of climate change, more specifically plastics, on the environment. Despite, their enhanced pro-environmental attitudes towards sustainable and durable consumption schemes, they often fail to appropriately convert those attitudes into behaviour patterns at the checkout counter. The resulting attitude-behaviour gap constitutes the fundamental starting point of the present research paper. Indeed, this analysis aims to explore the attitude-behaviour gap phenomenon by investigating the significant factors affecting the attitude-behaviour inconsistency. Moreover, by exploring the socio-demographic characteristics associated with the attitude-behaviour gap, a demographic profile of environmentally conscious consumers misaligning their behaviour can be painted. The fundamental assumption of a mismatch between attitudes and behaviour in terms of plastic consumption and waste production is the starting point of the present analysis. The findings of the paper should enable a better understanding in explaining and predicting the attitude-behaviour inconsistency in the aspiration to bridge the gap.

**Research Methodology:** A behavioural study was conducted on external data collected through a worldwide preferendum survey. The survey data quantifies individual-level intentions, attitudes, behaviour, habits, norms, and personality traits regarding sustainability. Since the present analysis follows an inductive research approach, the study starts with exploratory research, namely principal component analysis, to depict underlying factor patterns influencing the attitude-behaviour gap among the available data. Subsequently, the retrieved factors are implemented as covariates in linear regressions to assess the statistically most significant factors on the attitude-behaviour inconsistency.

**Results:** The analysis results show that eight dimensions, each representing a factor impacting the gap, were found through principal component analysis. The retrieved factor components include; a respondent's income level, environmental education and communication, altruism, willingness-to-pay a price premium, social influence, perceived climate change risk, and socially desirable responding. The regression analysis shows that the enumerated factors all have a statistically significant effect on the attitude-behaviour gap. Subsequently, the results of the association tests between socio-demographics and the disparity between attitudes and behaviour show that there are substantial differences in terms of demographic characteristics between individuals displaying a misalignment and those who do not. Consequently, the findings enable to define the most impactful components on the attitude-behaviour gap and paint precise demographic profile on individuals most likely to misalign their attitudes and behaviour.

**Research Limitations:** The first limitation of the present paper concerns the nature of the preferendum survey and the somehow restricted research settings. The niche target audience, the overrepresentation of individual countries, and the restricting scope of the attitude-behaviour gap exploration in terms of plastics consumption jeopardise the generalisation and extra-polarisation of the research findings on a more diversified population sample. Moreover, the online implementation of the preferendum questionnaire

reduced the controllability of the research environment. In the same vein, the accessibility of the survey through a mediation platform and the prior need to create an account are partially accountable for the unexpected low participation rate. Finally, the two most considerable biases introduced and to some extent, corrected for are; the socially desirable responding bias, and the non-reverse-coding of negatively phrased survey item bias.

**Managerial and Academic Relevance:** The academic relevance of the present study lies in the aim to bridge the remaining lack in the quantitative mapping of the attitude-behaviour gap through the medium of data science and machine learning techniques. Moreover, the target audience of the survey, namely Gen Z, remains a poorly investigated demographic segment in terms of attitude-behaviour inconsistencies. Hence the particular interest in exploring this niche target population sample. Besides the academic relevance, the research is of high managerial significance. Indeed, the findings can assist in future attitude-targeted marketing strategies to adequately promote alternative products to reduce the total plastic waste production. Next to marketers, policymakers could implement the research conclusions to formulate recommendations and interventions to stimulate sustainable consumption behaviours, especially among adolescents and young adults, that represent the consumers of the tomorrow.

**Keywords**: attitude-behaviour gap, sustainable consumption behaviour, principal regression analysis, worldwide youth preferendum survey

# Table of Contents

# List of Abbreviations

| | |
|---|---|
| **ABG** | Attitude-Behaviour Gap |
| **AUCROC** | Area Under the ROC curve |
| **BIDR** | Balanced Inventory of Desirable Responding |
| **CC** | Climate Change |
| **EFA** | Exploratory Factor Analysis |
| **GNI** | Gross National Income |
| **GOF** | Goodness-of-fit |
| **HI** | High Income |
| **HL** | Hosmer-Lemeshow |
| **KMO** | Kaiser-Meyer-Olkin |
| **LI** | Low Income |
| **LMI** | Low-Middle Income |
| **LR** | Likelihood ratio |
| **ML** | Maximum Likelihood |
| **PA** | Parallel Analysis |
| **PC** | Principal Component |
| **PCA** | Principal Component Analysis |
| **ROC** | Receiving Operating Characteristic |
| **SDR** | Socially Desirable Responding |
| **TPB** | Theory of Planned Behaviour |
| **TRA** | Theory of Reasoned Action |
| **UMI** | Upper- Middle Income |
| **UN** | United Nations |
| **WTP** | Willingness-To-Pay |

# List of Figures



# List of Tables

# Chapter 1 – Introduction

The first chapter introduces the reader to the attitude-behaviour gap resulting from a misalignment between consumer's attitudes and behaviour. In the first subsections, the key concepts and the overall problem are defined and discussed. Moreover, the arising research purpose, including the research question, are displayed. Finally, the academic and managerial relevance of the study is debated.

## 1.1. Problem Definition

Since its commercialisation in the 1940s, the plastic industry has experienced a 20-fold surge. In 2018, worldwide plastic production reached 360 million tons, generating over 1.6 million jobs in Europe and a turnover of 360 billion euros (PlasticsEurope, 2019). Plastics has been one of the most revolutionary findings since the 20th century and is shaping everyday life since then (Heidbreder et al., 2019; Parker, 2019).

The commercial success of plastics is extensively due to its polyvalent properties enabling countless applications. Among many other attributes, one can list its low production cost, high resistance, durability, lightness, and malleability. This game-changing material is omnipresent in an infinite range of products and industries including, packaging, construction, vehicles, electrical and electronic devices, agriculture, clothes, and footwear (Boucher & Billard, 2019). Many life-changing and life-saving innovations have emerged through plastics production. For illustration, in the medicine and health industry, helmets, incubators, and drinking water sanitation facilities were made accessible. Moreover, in the automobile industry, plastics led to the democratisation of automobile vehicles becoming increasingly energy-efficient, saving fuel and reducing pollution (Parker, 2019).

Nevertheless, besides the various economic benefits of plastics, the downside effects of exponential plastic consumption are increasingly visible (Boucher & Billard, 2019). The environmental crisis is mainly defined by marine littering through plastic debris, leading to dramatic side effects on marine wildlife, the oceans, and human's health (Jambeck et al., 2015). Yearly, approximately ten million tons of plastic waste leak into the marine ecosystem (Boucher & Friot, 2017). According to the International Coastal Cleanup report (2019), the top ten littered items consist of single-used food-related packagings. The widespread use of plastics in the food industry because of the cheapness and multifaceted nature of the material entailed a throw-away culture. Even though single-used plastic items usually have a very limited time span ranging from few minutes to hours, once it enters the ecosystem, it persists for hundreds of years without being decomposed (Parker, 2019; Adane et al., 2011).

Although considerable efforts and actions have been undertaken to limit plastic footprint, global plastic leakage remains a persistent and growing issue. Considering that no "global peak waste" is expected before 2100, the need for appropriate solutions to reduce plastic consumption and waste is paramount (Jambeck et al., 2015). Accordingly, the notion of durable and responsible consumption patterns has become increasingly relevant, especially in developing economies where overconsumption of plastics has destructive consequences on the air quality and the environment per se (Nguyen et al., 2019). Besides, population and country effects, habits, norms, situational, and societal factors play a crucial role in plastic consumption

behaviour (Heidbreder et al., 2019). Consequently, to shape long-term actions, the understanding of consumers preferences and perceptions is a necessary condition.

The SHIFT framework defined by White and her colleagues (2019), illustrates the fundamental factors impacting sustainable consumption behaviour. The elements accounting for the model's acronym are the following; social influence, habit formation, the individual self, feelings and cognition, and tangibility. Moreover, the research lays the ground for the analysis of the attitude-behaviour gap. Indeed, consumers' willingness to reduce plastic waste and to shift towards more sustainable consumption schemes is the first step in the right direction. Nevertheless, if concrete and tangible actions do not follow the intentions for 'greener' consumption behaviour, the climate change issue remains unchanged (White et al., 2019).

To date, the prevalent response to environmental challenges is a 'green-growth' paradigm. However, the oxymorons; pro-environmental, green, and ethical consumerism, designating this transition speak for themselves. Those wordings are a juxtaposition of two contradictory terms, namely the notion of 'green' and ethical use of natural resources and the concept of homo oeconomicus standing for utility maximisation comprising resource devastation. All in all, pro-environmental consumerism describes a sustainable-oriented consumption scheme minimising negative externalities on the environment while preserving the current consumption levels (Nguyen et al., 2019; Kim et al., 2012).

## 1.2. Research Purpose

The objective of the present study is to conduct an exploratory analysis on the attitude-behaviour gap in the domain of sustainable consumption behaviour, more specifically in the human solutions to the plastic waste issue including the utilisation of recycled and, or, the re-utilisation of plastics. The paper intents to depict the factors having the most significant impact on the attitude-behaviour misalignment and hence are critical drivers for consumers to switch to more environmental-friendly behaviour and consumption schemes bridging the gap.

The conclusion on the most substantial factors in explaining and predicting the gap can subsequently be tackled to formulate future recommendations and targeted interventions to promote sustainable behaviour and reduce plastic waste effectively. The focus of the study on young adults is of particular importance since those are the consumers of the future, susceptible to make a difference for upcoming generations (Vermeir & Verbeke, 2008; Heaney,2007).

The starting point of the study is the fundamental assumption that a positive attitude towards sustainability does not necessarily translate into concrete actions to fight climate change. The attitude-behaviour gap phenomenon puts into question many theories on rational consumer behaviour, among others, the well-known Theory of Planned Behaviour. Hence, there is a need for a more generalised and adapted framework. The present paper investigates the following research questions;

Research question 1: *"How can psychological and social factors on sustainable attitudes and behaviour measured through a preferendum survey be used to explain an attitude-behaviour gap?"*

Research question 2: *"Which factors have a statistically significant effect on the attitude-behaviour gap?"*

## 1.3. Academic and Managerial Relevance

The academic relevance of the present study is meaningful for diverse reasons. First, despite prior research on the attitude-behaviour gap in the field of sustainable consumption, there is a remaining lack in its quantitative mapping. Consequently, there is a substantial need for quantitative research on the inconsistency through the medium of data science and machine learning techniques. Indeed, pre-existing studies mainly focus on the qualitative analysis based on individual and personal interviews which is a standard method in psychological and social research to uncover individuals' attitudes, intentions and behaviour (Boulstridge & Carrigan, 2000). Second, prior quantitative analyses on the main factors explaining the attitude-behaviour gap focus on a single country, limiting the scope of the study since no comparison among other countries are possible (Vermeir & Verbeke, 2008; Adane & Diriba, 2011; Campbell et al., 2010). Conducting a quantitative analysis of the attitude-behaviour gap on a worldwide scale enables to provide generalised conclusions on the main drivers influencing its presence.

Moreover, the present analysis of the attitude-behaviour gap based on external survey data focuses on a young audience target aged between 13 and 24 years. The afore defined generations are commonly referred to as Generation Z (5 to 22 years) and the younger section of the Generation Y (18-24 years). These generations are represented by the most abundant and youthful share of the population. These young consumers might be able to make a difference for the upcoming generations by altering their consumption behaviour (Heaney, 2007). Despite their enhanced positive attitude towards sustainable consumption patterns, they seldomly convert these pro-environmental attitudes in sustainable behaviour, accordingly (Hume, 2019; Gaudelli, 2009). Consequently, the investigation of the driving factors contributing to the attitude-behaviour gap among adolescents and young adults is of particular interest. Furthermore, to date, the presence of an attitude-behaviour gap among the latter target group has not extensively been explored (Kolkailah et al., 2012).

Besides its academic relevance, the research is of high managerial significance. The identification of the driving factors influencing the attitude-behaviour gap is a promising starting point for reducing and even bridging the misalignment between pro-environmental attitudes and behaviour, which is crucial for marketers of ethical product alternatives. Hence, results can assist in future attitude-targeted marketing strategies to adequately promote alternative products to reduce the total plastic waste production. Next to marketers, policymakers could implement the findings to formulate recommendations and interventions to stimulate sustainable consumption behaviours, especially among adolescents and young adults, that represent the consumers of the tomorrow (White et al., 2019).

# Chapter 2 – Data collection

As outlined in chapter 1, the present analysis is inductive research based on an external survey source. Hence, the collected data is the foundation for deriving an alternative theoretical landscape to explain and predict the attitude-behaviour gap in terms of plastic consumption among adolescents. Consequently, the survey data is presented prior to the conceptual framework enclosing the developed hypothesis that will be tested throughout the present paper. The current chapter traces back the source of the manipulated survey data. Moreover, the performed pre-processing steps on the raw data are displayed as well as the measurement framework. Finally, the dependent variable, namely the attitude-behaviour gap, is modelled, which will be the dependent variable for all subsequent models of the paper.

## 2.1. Data Source

The confidential data to investigate the attitude-behaviour gap was collected through a worldwide preferendum survey carried out between September and November 2019. The preferendum is a result of a close collaboration between the 'State of Youth' and the 'KidsRights' Foundation, which is an academic partner of the Erasmus School of Economics (Rotterdam, Netherlands). The survey items were developed by the United Nation (UN) youth delegates. As the name of the survey, 'global state youth preferendum' implies, the target group of the analysis are adolescents and young adults between 13 and 24 years around the globe.

The survey was implemented via the Qualtrics online survey tool and spread through a connection with Facebook (KidsRights, 2019). The click ratio for the preferendum was unexpectedly low; there are various reasons for this. First, the website link to the questionnaire was available through the Facebook platform, raising suspicion and apprehension since users are mistrustful when it comes to Facebook's data privacy protection (Feng and Xie, 2014). Secondly, the survey was exclusively visible for users having set the default language of the platform to English. Finally, users willing to take part in the study had to create an account on the KidsRights' webpage.

## 2.2. Data Preparation

The raw data set includes 10.900 responses over 27 questions, some bundling various sub-questions. The first step after the data collection is the pre-processing of the data for the upcoming analysis. This step embodies variable elimination through prior knowledge and common sense. Hence, the original data set with 70 variables is reduced to a dimensionally lower subset of 45 variables. Annex A (section A.2.), provides an exhaustive listing of all removed variables and the reason for it. Since the questionnaire is partitioned into a mandatory and optional part, the valid responses are reduced to 3.669 observations when only considering complete cases.

## 2.3. Measurements

Most survey items are self-reporting questions measured on a 7-point or 10-point Likert scale. Table A.1 in Annex A provides an overview of the measurement method for each considered variable. Likert scales are particularly appreciated in the field of social sciences and psychometrics. Even though the latter measurement technique is very convenient at measuring respondent's views, attitudes and behaviour, the optimal method to treat the subsequent data remains a hot topic. There are two points of views among researchers; the ones considering Likert scale data as categorical data, more precisely ordinal variables. And the ones assuming it to be metric variables, namely interval-level variables (Jamieson, 2004).

When considered as ordinal data, only non-parametric statistic techniques can be applied to the data. Indeed, the distribution-free tests can be implemented without further concern since they do not assume any specific distribution. Statistics relying on some arithmetical manipulations (e.g. mean, standard deviation, chi-square) cannot be computed since the numbers on the scale correspond to statements and not numerics. On the other hand, when viewed as interval data, the interval distance on the Likert scale is equal and standardised. Hence, the numbers on the Likert scale can be translated into the corresponding integer value. Consequently, the data set verifies some underlying assumptions (i.e. homogeneity of variances, normality, linearity, independence), allowing the implementation of parametric tests (Jamieson, 2004).

According to Blaikie (2003), in practice, it has become widely accepted among researchers to treat Likert-scale data as interval-level variables and thus allowing the use of parametric statistics. One reason for the practise to differ from theory is that some statistical methods used to explore the relationships among the data (e.g. regression analysis, general linear model, factor analysis, item response theory) are based on parametric statistics (Harpe, 2015).

## 2.4. Dependent Variable: Attitude-Behaviour Gap

The purpose of the analysis is to evaluate the significant factors of the attitude-behaviour gap through regression analysis. Hence, the attitude-behaviour gap is the dependent variable of the model one tries to explain and predict. However, since the considered survey does not capture a direct measurement of the misalignment, it is essential to operationalising its central concept. The mapping of the gap is based on two constructs; respondent's attitude towards climate change and their behaviour.

Respondent's attitude towards climate change is quantified by two questions (Att1 and Att2). Both are measured on a 7-point Likert scale (1='not at all'; 7=' very much'), asking respondents to define how strongly they agree with an affirmation, keeping in mind that a neutral attitude is represented by the value of four (Jamieson., 2004). The behaviour towards plastic consumption is measured by a single item (Bhv) that is measured as a dichotomous variable. In case respondents did show sustainable action to reduce plastic waste over the past 12 months, the behaviour variable takes the value of one if-else, the variable is equal to zero.

Attitude 1 (Att1): Do you think that climate change is a serious problem?

Attitude 2 (Att2): Do you think that extra actions are needed to tackle climate change?

Behaviour (Bhv): In the past 12 months, did you and/or your family re-use plastic shopping bags and/or recycle plastic bottles?

Following the definition of the attitude-behaviour gap, an inconsistency occurs when respondents state that climate change is a serious issue and extra action should be taken but do not reduce their plastic footprint by recycling or re-using plastics. Conceptually, a pro-environmental attitude towards climate change is defined by scoring above four simultaneously on both items measuring the attitude (Att1 and Att2). Similarly, anti-environmental behaviour occurs for values on the behaviour variable (Bhv) equal to zero. Finally, the attitude-behaviour gap (hereafter ABG), is constructed, as shown in Equation 2.1. The resulting dependent variable is a dichotomous variable equal to one when there is an inconsistency between attitudes and behaviour, and equal to zero if-else. Out of the 3.669 complete survey responses, 456 respondents (12.43 %) show an attitude-behaviour gap. The variables explaining this phenomenon will be investigated in the upcoming chapter.

$$ABG = \begin{cases} Att1 > 4 \; and \\ Att2 > 4 \; and \\ Bhv = 0 \end{cases} \tag{2.1}$$

# Chapter 3 – Theoretical Framework

This chapter discusses prior literature about the fundamental concepts necessary to investigate and define the attitude-behaviour gap and the factors influencing this misalignment. First, the notion of the attitude-behaviour inconsistency is discussed. Next, the factors influencing the latter mismatch are outlined enclosing hypotheses based on prior literature and the available survey items in the preferendum. These developed assumptions lay the theoretical ground for the conceptual model of the present paper.

## 3.1. The Attitude-Behaviour Gap

The growth and popularisation of 'green' consumerism have enhanced consumer's awareness; nevertheless, sustainable actions remain weak at the checkout counter. Indeed, the "30:3 syndrome" confirms this phenomenon; out of 30% consumers showing a positive intention towards buying ethical products, only 3% consistently purchase these alternative products and hence translate their attitudes into behavioural choices (Cowe & Williams, 2000). The defeat to turn intentions into behaviour defines the attitude-behaviour gap (Carrington et al., 2014). In the pursuit of explaining the nature of this gap, many scientists have developed theoretical frameworks based on the Theory of Planned Behaviour (TPB) (Ajzen, 1991) and the Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1980). The central concept is the following; "beliefs determine attitudes, attitudes lead to intentions and intentions form behaviour" (Carrington et al., 2010). Hence, the stronger the interest in ethical consumerism, the more likely consumers will align attitude and behaviour (Vermeir & Verbeke, 2008). Although intentions are a sine qua non condition in the formation of behaviour, not all intentions translate in behavioural actions. Ajenz's (1991) decision-making model indicates that the decision process is affected by individual, social, and situational factors.

## 3.2. Influencing Factors on the Attitude-Behaviour Gap

The present section enumerates the factors that impact the attitude-behaviour gap. The hereafter outlined factors are principally derived from the SHIFT framework developed by White and her fellow partners (2019). The structure of the SHIFT model is based on a review of 320 scientific papers aiming to provide a comprehensive framework on the fundamental factors influencing sustainable consumption behaviour (White et al., 2019). Since the present analysis is inductive based on the preferendum survey, as described in chapter 2, only a subset of the factors defined in the SHIFT framework are retained and explained. Indeed, only SHIFT elements that were operationalised in the form of a survey item are tested.

### 3.2.1. Environmental Education and Communication

Prior research on the impact of environmental knowledge on 'green' behaviour shows contradictory results. One the one hand, a higher level of environmental education and knowledge results in increased engagement in more sustainable behaviour. Indeed, people that are aware of the climate change issue and the consequences of their behaviour are more inclined to engage in pro-environmental behaviour schemes (Joshi & Rahman, 2015). However, the fraction of pro-environmental behaviour that is directly influenced by an increase in knowledge and education is limited (Kollmus & Agyeman, 2002). Climate change education among the younger generation has become an increasingly important concern in the hope to foster long-term changes towards more sustainable behaviour. Educating future consumers on the

consequences of their actions on the environmental footprint might be a turning point in changing deep-rooted habits.

Nevertheless, erroneous conceptions, misunderstandings, and misinformation concerning climate change remain a substantial challenge (Hung, 2014). Besides climate change education, reliable communication is equally important. Since the environmental crisis is not directly perceived and visible, the phenomenon can rapidly be underestimated. Hence, the crucial role of trustworthy and credible information and communication sources to bring the seriousness and the urgency to act appropriately closer to the citizens (Hans & Cox, 2015).

Hypothesis 1: Environmental knowledge and education positively influence the ABG.

Hypothesis 2: Trustworthy and credible communication sources on climate change shape the general perception and hence have a positive impact on the ABG.

### 3.2.2. Willingness to Pay Price Premium and Household Income

A crucial barrier when it comes to implementing environmental-friendly product alternatives in everyday consumption habits is the price. A price premium, an extra cost added to the initial price, defines green product alternatives. The price premium considerably reduces the accessibility and democratisation of green product alternatives. Hence, it can be perceived as a metric for consumer's demand for alternative products (Tse, 2001). Accordingly, the decision on switching towards eco-responsible product alternatives depends on situational factors, as a consumer's financial situation and its willingness-to-pay price premium (WTP). Consumers with a high WTP, do not consider extra prices as a constraint to eco-responsible consumption and are willing to sacrifice some expenditures for sustainable behaviour yielding long-term environmental welfare. In the same vein, household's with higher resources are less price-sensitive and thus more likely to switch towards more expensive pro-environmental product alternatives (Moser, 2015). Recall, that the WTP a price premium and a household's income level only have an impact on product alternatives (buy re-usable shopping bags and drinking bottles). The latter factors don't necessarily impact sustainable behaviour that does not rely on physical consumption (e.g. reduce water consumption, reduce and sort waste).

Hypothesis 3: A household's income level positively influences the ABG.

Hypothesis 4: Consumer's willingness to pay a price premium to tackle climate change has a positive effect on green consumption behaviour.

### 3.2.3. Perception of Risk

Climate change is often perceived as an impalpable, vague, and abstract concept. Moreover, the consequences of the environmental crisis are considered as a distant phenomenon only impacting other countries and future generation (White et al., 2019). However, climate change is a multidimensional challenge noticeable through various environmental changes including, global temperature and sea-level rise, warming oceans, shrinking ice sheets, a surge in extreme weather events, and air pollution (Evan, 2019). As an illustration, over the last decade, a record number of extreme weather events has been recorded, among those heat waves, droughts, floods, and hurricanes. Besides a rise in the frequency of natural catastrophes, its severity and longevity have also increased (Monirul Qader Mirza, 2003).

Prior research has demonstrated that as long as climate change does not impact every-day life and people can adapt to it, the environmental crisis is not perceived as an urgent threat and remains a far-off problem in the public's opinion. The perceived risk related to climate change is a central determinant for people to realise the severity of the issue and thus increase the likelihood of pro-environmental behaviour (Slovic, 2000; Lowe et al., 2006).

Hypothesis 5: The higher the perceived risk of climate change, the higher the likelihood to engage in pro-environmental behaviour.

### 3.2.4. Social Influence

According to Cialdini et al. (1990), descriptive norms refer to how people commonly behave and can be interpreted as unwritten standards or guidelines. More precisely, these norms rely on the perception of peers' behaviour and the willingness to comply with the observed behaviour schemes. Following the social learning theory, human behaviour is derived from the observation of other's people behaviour and by imitating their actions. The influence of descriptive norms on consumer's behaviour is generally underestimated since the imitation of an individual's behaviour with whom one has communicated, occurs naturally and unconsciously (Cialdini et al., 1990). Prior research on the relationship between the descriptive norms and the attitude towards sustainability is mitigated. Hence, the need to test, whether there is a correlation between the descriptive standards and the attitude-behaviour gap (Fang et al., 2017).

Hypothesis 6: Descriptive norms have a significant positive effect on the ABG.

### 3.2.5. Altruism and Personality traits

Sense of responsibility and consumer's personal norms have significant repercussions on the attitude-behaviour gap. The individual sense of responsibility, and the personality characteristics defined by the Big Five personality traits theory reflecting individuals' standards, substantially increase sustainable and responsible behaviour engagement. Among those traits, one can list, altruism, helpfulness, the willingness to make concessions, conscientiousness and many more individual-specific values. Indeed, consumers defending those attributes are individuals prioritising the community over the self, which can also be defined as collectivism (Kollmus & Agyeman, 2002; Moser, 2015).

Hypothesis 7: Personal norms and responsibility have a positive effect on bridging the ABG.

### 3.2.6. Socio-Demographic Characteristics

Prior research (White et al., 2019) has demonstrated that consumers' socio-demographic characteristics have an impact on environmental-friendly attitudes and behaviour. Trying to define environmental conscious consumers in terms of demographics has resulted in the following conclusions. (1) Women demonstrate a stronger trend towards sustainable behaviour and are more open to changing their everyday habits, (2) the higher the education level, the higher the accumulation of knowledge about the climate change issue. However, the relationship between the highest achieved level of education and environmental consciousness is not necessarily linear. (3) Moreover, a consumer's age plays a role in their willingness to behave in a pro-environmental manner. (4) Finally, from a statistical point of view, consumers who originated from less developed countries are less environmentally conscious. However, when considering

the ecological footprints, richer countries exhibit far more extensive negative externalities on the planet (Kollmus & Agyeman, 2002; White et al., 2019).

Hypothesis 8: Socio-demographic traits have a significant impact on the ABG.

H8.1: Women are more environmentally engaged than men. Hence female gender has a positive influence on the ABG.

H8.2: A higher completed education level is positively correlated with the ABG.

H8.3: Young consumers are more likely to engage in pro-environmental behaviour.

H8.4: The income and development level of consumers' country of residence has an impact on the ABG.

### 3.2.7. Socially desirable responding in self-reporting survey research

In social sciences, a common approach to assess consumer's attitudes towards ethical consumerism and their ensuing behaviour, are self-reporting survey instruments (Carrington et al., 2010). Through rating scales (i.e. Likert scales), consumers are asked to evaluate to what extent they care about an environmental issue and how they would or already have adapted their behaviour accordingly (Auger & Devinney, 2007). Hence, researchers rely on participants answers to investigate thoughts, opinions, feelings, intentions, and behaviours. Unfortunately, it is confirmed that the nature of rating scale questions; question-wording, -format, and -context, have a substantial impact on respondent's answers and thus are an imperfect and a non-exclusive data source (Schwarz, 1999). Moreover, social desirability and self-presentation are additional factors leading to survey bias. Respondents may overstate their concern on ethical issues since consumers tend to adapt their responses to fit social norms (Auger & Devinney, 2007; Carrington et al., 2010).

Hypothesis 9: The bias introduced by social desirability responding might amplify the ABG.

## 3.3. Conceptual Framework

Figure 3.1 illustrates the conceptual model explaining the attitude-behaviour gap concerning sustainable plastic consumption behaviour based on the developed hypotheses throughout chapter 3. Recall, the assumptions are based on prior research and their operationalisation in terms of survey items. The correlation between the factors influencing the attitude-behaviour gap, and the misalignment itself are tested in the subsequent analysis.

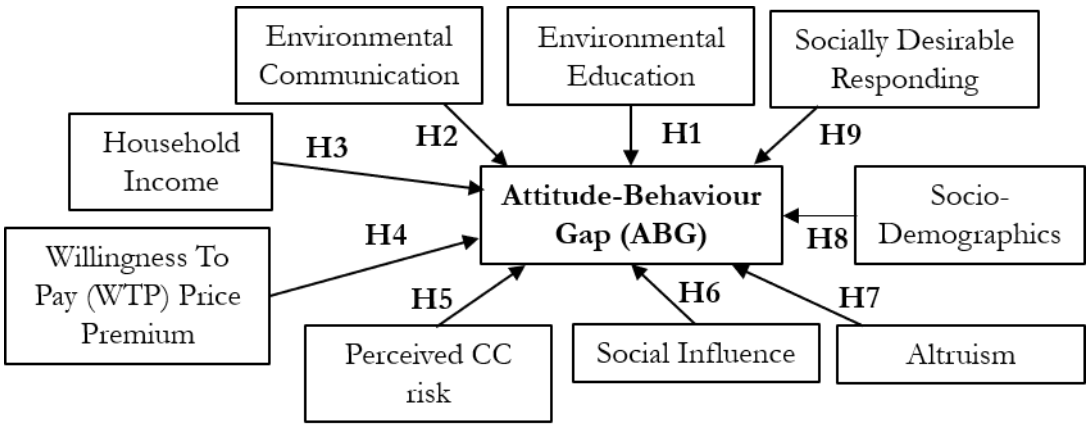


**Figure 3.1:** Conceptual framework of the Attitude-Behaviour Gap

# Chapter 4 – Methodology

The present chapter discusses the methodological framework of the analysis. First, the methodological approach and the research settings are presented, followed by the principal component theory. Finally, the regression analysis based on the principal component results is defined and evaluated.

## 4.1. Methodological Approach

The paper investigates the variables significantly influencing the attitude-behaviour gap. More specifically, one wants to understand the main factors for consumers to translate their green attitudes into sustainable behaviour. To date, in social sciences, pro-environmental attitudes and behaviour constitute a complex relationship and have been subject to prior research for decades. According to Field (2013), exploratory research methods are particularly well-suited to investigate underlying patterns on consumers showing an inconsistency between their attitudes and behaviour. Among the most popular exploratory techniques, one counts the Principal Component Analysis (PCA) and the Exploratory Factor Analysis (EFA).

Even though, in general PCA and EFA are assumed to be different varieties of the same technique, they are conceptually distinct, leading to their common misuse and misinterpretation, especially in social and psychological research. On the one hand, EFA estimates unknown latent constructs within the data that cannot directly be measured. By investigating hidden structures in the data, EFA can also be interpreted as a technique to 'generate a new theory'. On the other hand, PCA does not assume any assumptions about underlying latent models and is primarily a dimensionality reduction technique. All in all, EFA is a structural explanation technique, whereas PCA is a dimensionality reduction tool (Matsunga, 2010).

In the present study, the choice fell on PCA for various reasons. First, even though the use of EFA might be more accurate, it is also more complex. Since the results generated through PCA are very similar to the ones extracted from EFA, the tendency towards PCA is undeniable (Field, 2013). Second, the implementation of PCA results in posterior regression analysis eliminates encountered issues when considering the full set of variables. Indeed, PCA accounts for uncorrelated components, reducing high multicollinearity in the original set of variables. The reduction of the initial set of variables to a more parsimonious and relevant subset facilitates the variables' interpretation (Byrne, 2005). The extracted dimensions through PCA are implemented in a regression model to assess the magnitude of the contribution of every component to the response variable, namely the attitude-behaviour gap. At the same time, the hypotheses developed in chapter 3 are tested and evaluated.

## 4.2. Principal Component Analysis

PCA is an unsupervised tool implemented for exploratory purposes on numerical data. The dimension $(nxp)$ of the original data matrix, $x$, is reduced by fewer variables $(\ll p)$, capturing most of the variability in $x$ while guaranteeing for minimal information loss. By linear combinations of the original variables, PCA generates new independent and uncorrelated variables, also called principal components (PCs). Besides its ability to reduce the original data dimension to a smaller set of representative variables, PCA is also a handy visualisation tool (James et al., 2013).

However, PCA can only be carried out if the underlying data structure verifies two assumptions; (1) the adequate sampling of the data and (2) the intensity of the correlation between variables. The Kaiser-Meyer-Olkin (KMO) test assesses the sampling adequacy. The latter metric indicates the degree of variance among all variables that might be common variance. The KMO statistic ranges between zero and one, values above 0.6 confirm that the sampling is adequate. Indeed, values below 0.6 or close to zero indicate widespread correlations, which is a violation of the PCA hypothesis. The Bartlett's test of sphericity verifies the assumption that the correlation matrix of $x$ is an identity matrix. A significant Bartlett's test ($p < 0.05$) confirms that the correlation matrix yields substantial correlations among variables (for some at least), which is a fundamental condition for the PCA implementation (Pallant, 2013; IBM, 2019).

After verification of the here before listed assumptions for the use of PCA, the first principal component can be derived. Let's start with the geometrical interpretation of PCA to understand how one can represent multidimensional data in a coordinate system. Subsequently, through the spatial representation of PCA, the statistical dimensionality reduction will be illustrated.

Geometrically, the matrix $x$ can be represented in a coordinate system where each variable $p$ defines an axis, and each object $n$ is a point in the space. The more variables in the data set, the higher the dimensions of the space, and the more complex the representation becomes. First, the variables must be standardised to ensure the comparability of variables measured on different scales. PCA commonly accepts z-score normalisation, where all variables have zero-mean (by subtracting the mean from each variable, $\mu$) and unit-variance (divide by the variable's standard deviation, $\sigma$) as shown in Equation 4.1. In the case of non-normalisation, the PCA result is very dependent on variables with high variances. Secondly, the variables are mean-centred, meaning that they are moved towards the centre of the space (James et al., 2013).

$$z = \frac{x - \mu}{\sigma} \tag{4.1}$$

The first principal component, $z_1$, is derived by fitting a line through the centre of the data cloud. This line should be as close as possible to the cloud by minimising the sum of the squared distance between the first PC and each data point. The direction of the first PC is defined by the direction of the maximum variability in the data. The higher the correlations between that line and the data points, the higher the ability of the first PC to explain the original data. The loadings also called eigenvalues of $z_1$, are obtained by perpendicular projections of each observation onto that line. The distance from the origin to the projected points on the

fitted line designates the loading. For each component, there are as many loadings as there are observations in the data forming the component's eigenvector (Jolliffe, 1986; James et al., 2013).

Even though the 1st dimension captures the largest variability in the data, it is not sufficient to describe it entirely. Hence, the second component, $z_2$, can be added to the coordinate system. The 2nd PC is an additional line drawn through the data cloud, crossing the origin of the coordinate system and orthogonal to $z_1$. Similarly, each data point is orthogonally projected on the second vector to obtain its eigenvalues. This process is repeated for each principal component, $(z_1, z_2, \dots, z_m)$, until the maximal number of components, $min(n-1, p)$, has been generated (Jolliffe & Cadima, 2016).

Statistically, PCA relies on correlations and dependencies among the original variables. Hence, the first step, after z-score normalisation, is to compute the correlation matrix within the variables. The output is stored in the $(pxp)$ covariance matrix $\mathbf{\Sigma}$, indicating the magnitude of the co-dependence of the variables to each other. The $(i, j)$th element of $\mathbf{\Sigma}$ is the variance of the $j$th variable of the original matrix $\boldsymbol{x}$ when $i = j$. In case $i \neq j$, the element of $\mathbf{\Sigma}$ is the covariance between variables $i$ and $j$, respectively (Jolliffe, 1986).

Mathematically, the linear combination of the z-scored variables $(x_1, x_2, \dots, x_p)$ accounting for the largest variability in $\boldsymbol{x}$ is denoted by the normalised linear combination, $\phi_1' x$, where $\phi_{1j}' = (\phi_{11}, \phi_{12}, \dots, \phi_{1p})$ represents the first eigenvector. The coefficients attached to the eigenvector, $\phi_{11}, \phi_{12}, \dots, \phi_{1p}$, are the eigenvalues, also known as the loadings. The function is said to be normalised, constraining the eigenvalues such that their sum of squares must be equal to 1, $\sum_{j=1}^{p} \phi_{1j}^2 = 1$. The derived function in Equation 4.2 corresponds to the 1st principal component (James et al., 2013).

$$z_1 = \phi_1' x = \phi_{11} x_1 + \phi_{12} x_2 + \cdots + \phi_{1p} x_p = \sum_{j=1}^{p} \phi_{1j} x_j \tag{4.2}$$

Concretely, Equation 4.2 is adapted to obtain the 1st component of the data sample made up of $n$ observations for $i = 1, \dots, n$ and $p$ variables for $j = 1, \dots, p$ which can subsequently be formulated in terms of a maximisation problem as shown in Equation 4.4.

$$z_{1i} = \phi_{11} x_{i1} + \phi_{12} x_{i2} + \cdots + \phi_{1p} x_{ip}\,. \tag{4.3}$$

$$max(\phi_{11}, \dots, \phi_{p1}) \left\{ \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1}^{p} \phi_{1j} x_j \right)^2 \right\} \text{ subject to } \frac{1}{n} \sum_{j=1}^{p} z_{1i}^2 \tag{4.4}$$

Similarly, the 2nd linear function $\phi_2' x$, is orthogonal to the 1st component, maximising the remaining variance in the data. This process is repeated until the last $m$th dimension, $\phi_m' x$, is found that is uncorrelated with $\phi_{m-1}' x$. Finally, the $m$th principal component is defined by $z_m = \phi_m' x$, where $\phi_m$ is an eigenvector of $\mathbf{\Sigma}$ corresponding to the $m$th largest eigenvalue $\lambda_m$. Recall that the normalisation constraint of the eigenvalues is still applicable, thus the variance $var[z_m] = \lambda_m$. By ranking the eigenvectors in descending order with respect to their attached eigenvalues, the $m$ principal components are presented in descending significance order (Jolliffe, 1986; James et al., 2013).

Although, PCA is a convenient technique to reduce a broad set of variables to fewer artificially created components capturing most of the variability, the interpretation of those components can be tricky. Orthogonal rotations on the eigenvectors increase the interpretability of the retrieved components. Hence, their analysis is simplified while maintaining the assumption of orthogonal (uncorrelated) principal components. The Varimax-Kaiser rotation is a powerful rotation technique in PCA. The method aims to reallocate the principal components' eigenvalues such that each dimension captures a small number of high loadings and a high number of low (or zero) loadings. After the Varimax rotation, each component is defined by fewer variables showing high eigenvalue scores which substantially facilitates their interpretation. Through the reallocation, the initial variance captured by each dimension is dismantled; the first rotated component is not necessarily the dimension accounting for the maximum variability anymore. Hence, proportions of captured variance for each rotated component must be recomputed. (Stevens, 2002).

The derived PCs rank in decreasing order in terms of the proportion of captured variability. However, in general, the first few dimensions are most representative of the original data. They account for the most significant percentage of variance explained. Hence the last components are removed since they reflect potential noise. Consequently, the optimal number of dimensions to retain has to be determined. The most popular methods to define the optimal amount of components are; (1) Screeplot, (2) Kaiser's rule, and (3) Horn's parallel analysis (Hadi et al., 2016).

(1) The scree plot is a graphical representation of the components' eigenvalues, which are plotted in decreasing order, illustrating the cumulative variance explained by each dimension. The number of significant dimensions to preserve is defined by the location where there is a substantial change in the plot's slope, and the eigenvalues level off. This method is known as the elbow test and is heavily criticised because of its subjective nature, questioning its reliability (Zwick and Velicer, 1982).
(2) The Kaiser's rule stipulates that every component with eigenvalues below one should be discarded. Indeed, components with eigenvalues below the defined threshold explain the same or even less variability than a single variable on its own. The drawback of this trivial criterion is the overestimation of the number of components to be extracted (Kaiser, 1991).
(3) Horn's parallel analysis (PA) is considered as an effective and non-trivial method to define the optimal number of dimensions. The first step of PA consists of generating normal distributed random data through the Monte Carlo sampling algorithm. The reshuffled data is parallel to the initial data in terms of observations and the number of variables. In a second step, PCA is conducted on the artificial data obtained in step one. Both steps are usually repeated between 500 and 1000 times. Hereafter, the obtained eigenvalues are averaged and compared to the eigenvalues retrieved from the original data. Finally, a dimension is significant, worthy of being preserved, if the eigenvalues of the initial component are larger than the averaged eigenvalues on the parallel component (Peres-Neto et al., 2005). In case the eigenvalues of the original dimension are equal or smaller than the averaged eigenvalues derived from PA, the component is equally significant than a random dimension and thus can be dismissed (Peres-Neto et al., 2005; Matsunaga, 2010).

Recall, the significant components are interpreted in terms of the absolute value of their corresponding loadings. Stevens (2002) suggests that variables with associated component loadings above 0.4 in absolute terms, share at least 15% of their variance with the dimension and are consequently proven to be statistically significant. Accordingly, the variables within a component associated with the largest loadings $(> 0.40)$ are clustered within that dimension and analysed to find any similarity and common ground (Stevens, 2002).

The dimension names are derived from that similarity found within those clustered variables. Furthermore, dimensions defined by a few lower loadings $(\simeq 0.4 \ or < 0.4)$ might not be interpreted, since those variables seem to be barely significant and might not be impactful enough to define a component. Finally, to avoid dimensions to be variable-specific, only factors enclosing more than two significant variables are taken into consideration (Stevens, 2002).

After having defined the optimal number of dimensions and the variables present within each component, comes the point to assess the internal consistency and reliability of the extracted factors. The internal consistency in PCA is associated with the components' estimated reliability. In social sciences, the most popular metric to assess internal consistency is Cronbach's alpha coefficient (1951) (Cortina, 1993). More specifically, it refers to the degree of intercorrelation, the commonality, in the items within a dimension. Large coefficient values imply that the variance within a component is hardly associated with item-specific variance. Hence a large proportion of the variability is characterised by an underlying common factor. The higher the alpha coefficient, the lower the uniqueness, in other words, the item-specific variance, of the items within that dimension. In practice, a value above 0.45 ensures acceptable internal consistency (Cortina, 1993; Nunnally, 1978; Taber, 2018).

## 4.3. Logistic Regression

Regression analysis is a statistical technique able to mathematically model the relationship between a dependent variable and two or more independent variables. There are many variants of regression analysis, dependent on the type of data of the dependent variable. The binary logistic regression is an extension of the linear regression when the dependent variable is dichotomous (Aguilera et al., 2006).

Traditional linear regression expresses the response variable as conditional mean, $E(Y|x) = \beta_0 + \beta_1 x$, where $E(Y|x)$ represents the 'expected value of $Y$ given the value $x'$ (Hosmer et al., 2013). Since $x$ ranges from $-\infty$ to $+\infty$, the value of $E(Y|x)$ also varies within the same range. However, when the dependent variable is dichotomous, $E(Y|x)$ must range between zero and one $(0 \leq E(Y|x) \leq 1)$. The S-shaped logistic function (i.e. sigmoid function), which will be explained in more details at a later stage in the present chapter, constraints the output of the linear function to range between zero and one (Hosmer et al., 2013).

Contrary to alternative regression models, logistic regression does not assume linearity, normality, and homoscedasticity. Nevertheless, binary logistic regression is not assumption-free. Some assumptions must be verified beforehand, namely; (1) binary structure of the dependent variable, (2) observation

independence, (3) little to no multicollinearity among the independent variables, (4) linearity between the independent variables and log odds, and (5) a large enough sample size (Schreiber-Gregory, 2018).

Let's start with the basic theory of binary logistic regression. Let $Y_i(i = 1, \dots, n)$ be the binary dependent variable associated with the observations in $\boldsymbol{X}$, where $\boldsymbol{X} = (x_{ij})_{nxp}$, the matrix embodying a set of continuous variables, $X_1, X_2, \dots, X_p$, observed for n observations. The dependent variable follows the distribution of the Bernoulli probability function, $Y_i \sim Bernoulli(Y_i|\pi_i)$. Hence the response variable, $Y_i$, is equal to one with a probability of success $\pi_i$ and takes to value of zero with a probability of $(1 - \pi_i)$.

The logistic regression model, given by Equation 4.5, constraints the output $E(Y|x)$ to range between zero and one for all values of $X_p$. Thus, the logistic function will produce an S-shaped curve, limiting the range of probabilities to lie between zero and one (Aguilera et al., 2006). The vector of parameters gives the unknown regression coefficients, $\beta = (\beta_0, \beta_1, \dots, \beta_p)$, which values must be estimated. According to Equation 4.5, the relationship between the probability of $Y$ and all independent variables $(X_1, X_2, \dots, X_p)$ is non-linear, which represents a violation of an assumption of the binary logistic model. The logistic function, also known as the sigmoid function, presented in Equation 4.6, allows for meeting the assumption of linearity between the logit transformed dependent variable and the independent variables by approximating $E(Y|x)$ as a sigmoid function (Peng et al., 2002).

$$E(Y = 1|X_1 = x_{i1}, \dots, X_p = x_{ip}) = \pi_i = \frac{exp\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\right)}{1 + exp\left(\beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j\right)} \tag{4.5}$$

The left part of the equation, $\frac{\pi_i}{1-\pi_i}$, is defined as the log-odds or logit of the output $Y$ that is linear in $X_p$. The relationship between the response variable and each predictor variable must be understood in terms of the odds ratio. A one-unit increase in $X_p$, while keeping all remaining predictors constant, modifies the log odds by $\beta_p$, which is equivalent to multiplying the odds by $e^{\beta_p}$ (James, 2013; Aguilera et al., 2006). Even though one cannot directly evaluate the average change in $Y$ associated with a one-unit increase in one predictor variable, the coefficient indicates the direction of the relationship between the $logit(Y)$ and the regressor. For coefficients greater than zero, the relationship between a predictors value and logits of $Y$ is positive. Larger (smaller) $X$ values contribute to larger (smaller) logits of $Y$. When the coefficients are negative, the relationship is noticeably inversed (Peng et al., 2013).

$$logit(Y) = log\left[\frac{\pi_i}{1 - \pi_i}\right] = \sum_{j=1}^{p} x_{ij}\beta_j \tag{4.6}$$

For a proper interpretation of the model parameters, it is crucial to estimate the corresponding coefficients accurately. The Maximum Likelihood (ML) function, which is similar to the least-squares function used in traditional linear models, constitutes a conventional approach to estimate the regressor coefficients in case of logistic regression. The likelihood function aims to find a set of parameters, the ML estimates, that

maximises the probability of the observed dependent variable. The ML function to be optimised is derived from the Bernoulli distribution of the dependent variable, as shown in Equation 4.7 (Hosmer et al., 2013).

$$L(\beta|y) = \prod_{i=1}^{n} \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i} (1-\pi_i)^{1-y_i} \tag{4.7}$$

To simplify the equation even further, one can apply the natural logarithm to it. Assuming there is parameter maximising the likelihood function, this same parameter will also maximise the log-likelihood function. This is a direct consequence of monotone increasing transformations, which applies to the logarithm transformation. The final log-likelihood function can be reformulated as follows (Hosmer et al., 2013);

$$LL(\beta) = ln[L(\beta)] = \sum_{i=1}^{n} \{y_i \, log(\pi_i) + (1-y_i) log(1-\pi_i)\}. \tag{4.8}$$

To approximate the ML estimates, $\hat{\beta} = (\widehat{\beta_0}, \widehat{\beta_1}, \dots, \widehat{\beta_p})$, Equation 4.8 must be differentiated for each parameter. Subsequently, the partial derivatives are equalised to zero. However, since the derivatives are not linear in the parameters to be estimated, they cannot be solved in a precise way. Consequently, specific iterative algorithms (i.e. Newton-Raphson, iteratively re-weighted least squares, Fisher scoring) are implemented. However, deepening the theory of logistic regression in terms of the latter iterative algorithms is beyond the scope of the present research framework. Besides, the software used for the analysis of the logistic regression internalises an iterative method in the logistic regression function (Hosmer et al., 2013). Once the ML coefficients have been estimated, the probability of the binary response variable can be computed, corresponding to the result obtained in Equation 4.9 (James et al., 2013).

$$\hat{p}(Y = 1 | X_1 = x_{i1}, \dots, X_p = x_{ip}) = \widehat{\pi_1} = \frac{\exp\left(\widehat{\beta_0} + \sum_{j=1}^{p} x_{ij}\widehat{\beta_j}\right)}{1 + \exp\left(\widehat{\beta_0} + \sum_{j=1}^{p} x_{ij}\widehat{\beta_j}\right)} \tag{4.9}$$

### 4.3.1. Unbalanced response variable

In social research, binary response variables are commonly used to measure the presence/absence of a specific phenomenon or the change/no-change in consumer behaviour. Generally, in the domain of ecology and sustainability, the data implemented in logistic regression analysis reveals to be unbalanced. An unbalanced set is defined by an immense disproportion of success/failure or presence/absence in the response variable. These rare events have been proven to be challenging to explain and predict. Fitting a logistic regression on the minority class will usually underestimate the occurrence probability of a rare event. Hence, favouring the majority class can have severe implications in terms of model performance and the resulting conclusions. The issue related to unbalanced data can be overcome by adopting some resampling technique to create a 50-50 distributed response variable. Note that the resampling algorithm is exclusively applied to the training set to encourage the model to predict the minority class correctly. Random oversampling is the most known and easy-to-implement technique. The purpose of the latter algorithm is to duplicate minority class instances of the training data randomly. However, attention must be paid since this method can lead to overfitting (King & Zeng, 2001; Salas-Eljatib et al., 2018).

### 4.3.2. Model Evaluation and Diagnostics

This last subsection regroups the available methods to evaluate the model performance of the binary logistic regression. The hereafter enumerated statistical tests allow to assess the model fit on the data; the most significant predictor variables in the model and the accuracy of the model predictions. The displayed techniques are split into three groups; (1) the goodness-of-fit (GOF) statistics determining the model fit, (2) the statistical tests on individual predictors to investigate the most significant regressors, and (3) the predictive power metric, judging the accuracy of the model predictions (Hosmer et al., 2013).

**(1) Goodness-of-fit statistics**

a) Cox & Snell $R^2$

Cox & Snell $R^2$ measures the proportion of variability in the independent variable associated with the regressors. The metric takes into account the deviance, which is the difference in log-likelihoods between the fitted and the intercept-only model, only consisting of the model intercept. The $R^2$ can be interpreted as a ratio indicating how close the fitted logistic model is to the perfect model or the worst intercept-only model as defined in Equation 4.10. When adding regressors to the intercept-only model does not improve the model performance $(D = D_0)$, $R^2$ is equal to 0. On the contrary, if the model is perfectly fitted $(D = 0)$, $R^2$ is equal to 1 (Peng et al., 2002; Hosmer et al., 2013).

$$R^2 = 1 - \frac{-2log\ lik(\hat{\beta}) + 2log\ lik(null\ model)}{-2log\ lik(\widehat{\beta_0})} = 1 - \frac{D}{D_0} \qquad (4.10)$$

a) Hosmer-Lemeshow (HL) test

For the HL test, the observations in the data are segmented groups based on their model-predicted probability similarity $(\widehat{\pi_i})$. The observations are approximately allocated within each group to form equally sized groups. In general, the data is split based on the percentiles of the predicted probabilities, forming ten groups $(g = 10)$. The HL investigates the similarity between the observed and expected frequencies in the defined groups based on the chi-square statistic, $\chi^2$. The null hypothesis suggests that the expected probabilities are equal to the observed frequencies. In contrast, the alternative hypothesis states that they are not the same and that the model poorly fits the data.

**(2) Statistical tests on individual predictors**

a) Wald test

The Wald chi-square statistic, $\chi^2$ assess the statistical significance of each regressor coefficient. The metric aims to test the hypothesis that a regressor coefficient is significantly different from zero for each regressor of the model. In case the assumption is verified, the tested regressor substantially contributes to an increase in the model fit. In the alternative case, where the hypothesis is rejected, the tested variable can be discarded since it does not help in predicting the dependent variable (Hauck and Donner, 1977).

(a) Variable Importance

The importance of individual regressors can be quantified by the standardised regression coefficients indicating the normalised change in the response variable related to one standardised-unit increase in the corresponding regressor, keeping all remaining regressors unchanged. However, in the case of correlated regressors, this method does not account for the relative variable importance. Indeed, in the case of multicollinearity, the variable importance of a regressor must be considered in combination with other correlated regressors. Hence, the variable importance analysis might conclude that no regressors in the model are important, which does not reflect reality. Consequently, considering the standardised coefficients to determine regressors' variable importance is only consistent in the absence of multicollinearity among regressors. More sophisticated techniques have recently been developed to assess the relative importance of correlated regressors. Among those, one can count the dominance analysis, considering multicollinearity among the variables (Azen and Traxel, 2009).

**(3) Predictive Power Evaluation**

a) ROC curve

The Receiving Operating Characteristic (ROC) is a measure to evaluate the prediction performance of a model. In the case of a binary response variable, the ROC evaluates a model's classification performance. Moreover, the ROC enables to determine the optimal threshold value for predicting whether a new observation should be classified as a success or failure. In practice, the ROC curve illustrates the sensitivity, the proportion of correctly classified successes (true positive error rate), and specificity, the percentage of correctly classified failures (true negative error rate) for every decision rule threshold ranging from zero to one. The ROC plot graphically defines the trade-off between the true positive rate; one tries to maximise while minimising the false positive rate. The ideal example would show a ROC curve with a 90-degree angle, showing a cut-off for which specificity and sensitivity are equal to 100%, which means that the model correctly classifies every new observation without misclassifications. Besides, the area under the ROC curve (AUROC) indicates how accurately the logistic regression classifies every new observation for all possible threshold values. The AUCROC metric ranges from 0.5 to 1, where a value above 0.7 indicates that the logistic regression has reasonable predictive power (Hosmer et al.,2013).

b) Train-test process

The full data set is partitioned into two subsets, the training (learning) set and the test set to evaluate the overall model performance. The model is built on the training set, whereas its performance is tested on the set-aside test set by considering the test error rate. The most commonly used splitting threshold of two thirds and three quarters refers to a rule of thumb. Furthermore, the train-test split enables to evaluate whether the model is overfitting. A significant difference in the training and test accuracy rate indicates that the model perfectly fits the training data but is performing poorly on new unseen data (James et al., 2013; Harrell et al., 1996).

## 4.4. Logistic Regression on Principal Components

The model performance of the binary logistic regression seems to be especially unsteady and limited when facing a high number of multicollinear independent variables. Hence, the need for some dimensionality reduction to mitigate against multicollinearity. Remember, PCA is a technique that summarises the initial set of variables into fewer uncorrelated dimensions, the principal components while maximising the captured variability in the original data. Consequently, PCA is an optimal technique to ensure the introduction of the collinearity-free variable subset in the binary logistic regression model (Jolliffe, 1986). Besides solving the issue of multicollinearity, PCA additionally controls the risk of overfitting, since all variables not statistically contributing to the variance in the data, are considered as noise and are discarded from the model (James et al., 2013). In the same vein, scientific parsimony, advocating simple and understandable theory conceptualisations is complied with (Stevens, 2002).

Substituting the orthogonal components found in PCA as covariates in the logistic regression can be summarised in two major steps. First, perform PCA as preliminary data processing step to reduce the dimensionality of the original data. The newly created orthogonal dimensions through PCA overcome the critical problem of multicollinearity among the regressor variables and enable to conclude to the relative importance of each covariate in the explanation and prediction of the binary dependent variable. Section 4.2 is dedicated to the PCA theory, including how to extract the optimal number of components. Subsequently, the first few dimensions accounting for the largest variance are stored and will be incorporated as covariates in logistic regression. The factors accounting for a small proportion of variance are discarded based on the assumption that these components have no significant impact on the predictive power of the subsequent regression analysis (Jolliffe, 1982). Second, proceed to a binary logistic regression, including the selected principal component dimensions as covariates. By mean of the vector of estimated regression coefficients, the outcome of the logistic analysis, one can draw some conclusions on the importance of each component.

Figure 4.1 summarises the concept of implementing component factors obtained through PCA in a regression analysis. The aforementioned 2-step decomposition of the study, embodying the PCA as a pre-processing step and regression analysis as the final step, is reinforced by the visual representation of the workflow.

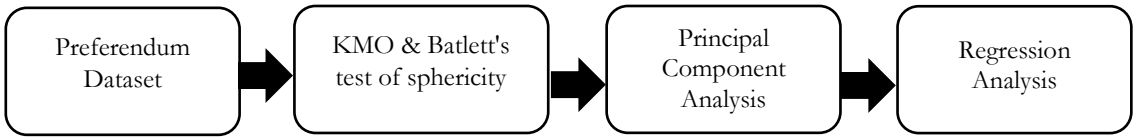


**Figure 4.1:** A workflow diagram of all analysis steps

# Chapter 5 – Results

This section outlines the findings of the analysis conducted on the preferendum survey, using the statistical software R (version 3.6.1; R Core Team, 2019). First, an overview of the sample characteristics is given, which is then followed by descriptive statistics. After that, the PCA results are presented. Finally, by substituting the principal component dimensions as covariates in regression analysis, the significance of the hypothesised assumptions, defined in section 3, are tested and evaluated.

## 5.1. Sample Characteristics

The retrieved population sample comprises adolescents of an average age of 18 years with a slight majority of female respondents (55%). It is notable that even though 56% of respondents are from lower-middle-income countries, 61% of respondents being 18 years or older, hold a university degree. The sample characteristics of educated and young adults are especially meaningful in the analysis of sustainable behaviour. Indeed, young adults are the consumers of tomorrow and are thus able to make a difference for the upcoming generations by altering their consumption behaviour. Furthermore, at their early stage of life, they are developing habits which are susceptible to be translated into future behaviour schemes (Heaney, 2007). Additionally, the underlying educated sample size is very likely to have prior knowledge about sustainability and climate change, making responses on questions related to the environmental crisis increasingly reliable (Vermeir & Verbeke, 2008).

**Table 5.1**: Socio-demographic composition of the survey sample

| **Socio-demographic variables** | **Total Sample (N=3.669)** | |
|---|---|---|
| | n | (%) |
| **Age**[a] | | |
| [13,16] | 1468 | 40.00 |
| [17,19] | 1270 | 34.61 |
| [20,23] | 916 | 24.97 |
| >24 | 15 | 0.40 |
| **Country of residence**[b] | | |
| Low income | 56 | 1.53 |
| Lower-middle income | 2056 | 56.04 |
| Upper-middle income | 690 | 18.80 |
| High income | 867 | 23.63 |
| **Gender** | | |
| Female | 2024 | 55.16 |
| Male | 1600 | 43.61 |
| Other | 45 | 1.23 |
| **Highest educational achievement** | | |
| No formal education | 8 | 0.22 |
| Education up to age 12 | 98 | 2.67 |
| Education up to age 14 | 395 | 10.77 |
| Education up to age 16 | 771 | 21.01 |
| Education up to age 18 | 798 | 21.75 |
| Education after age 18 (non-university) | 360 | 9.81 |
| University | 1239 | 33.77 |

[a] Age has been split into 4-factor levels to admit all socio-demographics as factor variables
[b] Country of residence is classified by income level following to the World Bank benchmark (see Annex A.3)

## 5.2. Descriptive Statistics

The Chi-square test assesses a possible association between different demographic variables with respect to two groups, namely respondents displaying a positive attitude-behaviour gap and the ones showing no misalignment. The null hypothesis of the Chi-square statistic states that there is no association between the tested demographic variable(s) and the response variable; whether there is an attitude-behaviour mismatch or not. The alternative hypothesis affirms that the ABG is dependent on the demographic variable(s). The results of the Chi-square statistic are shown in Table 5.2.

By means of the significance level of the test, one can conclude that for all tested variables, except for the highest educational achievement variable, there is a substantial difference in the ABG and any remaining demographic variable. To assess the effect size of individual factor levels on the attitude-behaviour gap, one can run a binary logistic regression for every demographic factor variable. Hence, the logistic regression defines a reference category among the tested factor levels. After that, the coefficient of the remaining factor levels describes the difference in terms of standard units between the reference level's coefficient and the coefficient of the tested categories. The associated $p$-value defines whether the coefficient difference is statistically significant or not (Chapman & Feit, 2015).

**Table 5.2**: Chi-square test between demographic factor variables among respondents showing a positive attitude-behaviour gap and those who do not

| **Socio-demographic variables** | **No ABG (n=3.213)** | | **ABG (n=456)** | | $\chi^2$ **$p$-value Sign. Level** [c] |
|---|---|---|---|---|---|
| | n | (%) | n | (%) | |
| **Age**[a] | | | | | 82.941 |
| [13,16] | 1.372 | 42.70 | 96 | 21.05 | 0.0004 |
| [17,19] | 1.076 | 33.49 | 194 | 42.54 | ⋆⋆⋆ |
| [20,23] | 751 | 23.37 | 165 | 36.19 | |
| ≥24 | 14 | 0.44 | 1 | 0.22 | |
| **Country of residence**[b] | | | | | 149.57 |
| Low income | 34 | 1.06 | 22 | 4.82 | 0.0004 |
| Lower-middle income | 1.708 | 53.16 | 348 | 76.32 | ⋆⋆⋆ |
| Upper-middle income | 636 | 19.79 | 54 | 11.84 | |
| High income | 835 | 25.99 | 32 | 7.02 | |
| **Gender** | | | | | 66.123 |
| Female | 1.849 | 57.55 | 175 | 38.38 | 0.0004 |
| Male | 1.321 | 41.11 | 279 | 61.18 | ⋆⋆⋆ |
| Other | 43 | 1.34 | 2 | 0.44 | |
| **Highest educational achievement** | | | | | 0.559 |
| No formal education | 7 | 0.22 | 1 | 0.22 | 0.700 |
| Education up to age 12 | 85 | 2.65 | 13 | 2.85 | |
| Education up to age 14 | 367 | 11.42 | 28 | 6.14 | |
| Education up to age 16 | 722 | 22.47 | 49 | 10.75 | |
| Education up to age 18 | 699 | 21.76 | 99 | 21.71 | |
| Education after age 18 (non-university) | 303 | 9.43 | 57 | 12.50 | |
| University | 1.030 | 32.06 | 209 | 45.83 | |

[a] Age has been split into 4-factor levels to admit all socio-demographics as factor variables
[b] Country of residence is classified by income level following to the World Bank benchmark (see Annex A.3)
[c] Significance codes : $'\star\star\star' \; p \leq 0.001$ ; $'\star\star' \; p \leq 0.01$ ; $'\star' \; p \leq 0.05$ ; $'.' \; \mathrm{p} \leq 0.1$

Let's start the analysis of the magnitude of the association between the age variable and the ABG, as displayed in Table 5.3. The age level [13,16] is the reference category. The difference between the coefficient of the reference level and the coefficient of levels [17,19] and [20,23] are respectively, statistically significant. Consequently, one can deduce that compared to respondents aged between 13 and 16 years, on average respondents aged between [17,19] and [20,23], show an ABG likelihood that is respectively 0.95 and 1.14 standard units higher than the reference level. The coefficient level of respondents older than 24 years is statistically indifferent compared to the reference category. Hence respondents being 24 years and older are equally likely to show an ABG (or not) compared to respondents aged between 13 and 16 years. To conclude, young respondents, are more likely to align their attitude and behaviour in terms of sustainable consumption.

**Table 5.3**: Logistic regression result between age variable and the ABG

| Model | Standardised Coefficients | | T-statistic | p-values | Significance level[1] | Odds ratio |
|---|---|---|---|---|---|---|
| | Beta | Std. Error | | | | |
| Age: [13,16] | Reference level | | | | | |
| Intercept | -2.660 | 0.106 | -25.193 | <2e-16 | ⋆⋆⋆ | |
| Age: [17,19] | 0.947 | 0.131 | 7.211 | 5.55e-13 | ⋆⋆⋆ | 2.577 |
| Age: [20,23] | 1.144 | 0.136 | 8.404 | <2e-16 | ⋆⋆⋆ | 3.140 |
| Age: ≥24 | 0.021 | 1.040 | 0.020 | 0.984 | | |

[1] Significance codes : $'\star\star\star' \; p \leq 0.001$ ; $'\star\star' \; p \leq 0.01$ ; $'\star' \; p \leq 0.05$ ; $'.' \; p \leq 0.1$
Dependent variable: ABG

Table 5.4 displays the effect size of the association between the income and development level of a respondent's residence country and the ABG. One can conclude that the difference between the reference level coefficient and the remaining income level coefficients are statistically significant, respectively. Hence, respondents from low-income economies have an ABG risk that is 2.83 standard units higher compared to respondents originated from high-income level countries. The difference between the reference category and lower-middle and upper-middle-income countries is more moderate but still statistically significant. The likelihood of a positive ABG is 1.67, and 0.54 standard deviations larger for respondents originated from lower-middle- and upper-middle-income countries compared to the ABG probability of respondents from high-income level economies. Conclusively, compared to high-income countries, respondents from middle- to low-income countries are more likely to misalign their attitudes and behaviour.

**Table 5.4**: Logistic regression result between the economic status variable and the ABG

| Model | Standardised Coefficients | | T-statistic | p-values | Significance level[1] | Odds ratio |
|---|---|---|---|---|---|---|
| | Beta | Std. Error | | | | |
| High Income | Reference level | | | | | |
| Intercept | -3.262 | 0.180 | -18.107 | <2e-16 | ⋆⋆⋆ | |
| Low Income | 2.826 | 0.328 | 8.628 | <2e-16 | ⋆⋆⋆ | 16.885 |
| Lower-middle Income | 1.671 | 0.190 | 8.817 | <2e-16 | ⋆⋆⋆ | 5.316 |
| Upper-middle Income | 0.536 | 0.240 | 2.235 | 0.025 | ⋆ | 1.709 |

[1] Significance codes : $'\star\star\star' \; p \leq 0.001$ ; $'\star\star' \; p \leq 0.01$ ; $'\star' \; p \leq 0.05$ ; $'.' \; p \leq 0.1$
Dependent variable: ABG

The same analysis procedure holds for the association between respondent's gender and the explored ABG as outlined in Table 5.5. The difference between the female and male coefficients and the ABG is statistically significant for both genders, respectively. Hence, compared to female respondents, the ABG on male participants is 0.80 standard units larger. Note, that as represented in Table 5.5, compared to female respondents, respondents defining their gender as 'other' are equally likely to display an ABG, since the difference between both coefficients is statistically insignificant.

**Table 5.5:** Logistic regression result between the gender demographic and the ABG

| Model | Standardised Coefficients | | T-statistic | p-values | Significance level[1] | Odds ratio |
|---|---|---|---|---|---|---|
| | Beta | Std. Error | | | | |
| Gender: female | Reference level | | | | | |
| Intercept | -2.358 | 0.079 | -29.810 | <2e-16 | ⋆⋆⋆ | |
| Gender: male | 0.803 | 0.103 | -7.798 | 6.31e-15 | ⋆⋆⋆ | 2.232 |
| Gender: other | -0.710 | 0.728 | -0.976 | 0.329 | | |

[1] Significance codes : $'\star\star\star' \; p \leq 0.001$ ; $'\star\star' \; p \leq 0.01$ ; $'\star' \; p \leq 0.05$ ; $'.' \; \mathrm{p} \leq 0.1$
Dependent variable: ABG

Finally, Table 5.6 evaluates whether there is a substantial difference between the coefficient of the reference level 'no formal education' and the estimates of the remaining factor level measuring respondent's highest achieved educational level. As displayed in the last column of the table, compared to the reference level, no factor level is statistically significant. Consequently, as already concluded from the Chi-square statistic, the ABG is independent of a respondent's education level. Hence, respondents are equally likely to mismatch their attitudes and behaviour (or not), regardless of their highest educational achievement.

**Table 5.6:** Logistic regression result between the highest achieved educational variable and the ABG

| Model | Standardised Coefficients | | T-statistic | p-values | Significance level[1] |
|---|---|---|---|---|---|
| | Beta | Std. Error | | | |
| No formal education | Reference level | | | | |
| Intercept | -1.946 | 1.069 | -1.820 | 0.069 | . |
| Education up to age 12 | 0.068 | 1.110 | 0.061 | 0.951 | |
| Education up to age 14 | -0.627 | 1.087 | -0.577 | 0.564 | |
| Education up to age 16 | -0.744 | 1.079 | -0.690 | 0.490 | |
| Education up to age 18 | -0.009 | 1.074 | -0.008 | 0.994 | |
| Education after age 18 (non-university) | 0.275 | 1.079 | 0.255 | 0.799 | |
| University | 0.351 | 1.072 | 0.327 | 0.743 | |

[1] Significance codes : $'\star\star\star' \; p \leq 0.001$ ; $'\star\star' \; p \leq 0.01$ ; $'\star' \; p \leq 0.05$ ; $'.' \; \mathrm{p} \leq 0.1$
Dependent variable: ABG

## 5.3. Principal Component Analysis

Before implementing PCA, the suitability of the data must be evaluated. Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) metric are performed to assess the data's factorability. Both criteria and their respective thresholds have been discussed in more details in chapter 4. Table 5.7 displays the results determining whether performing PCA on the underlying data set is reasonable. The preferendum data shows to be an adequate sample since the KMO value is equal to 0.75 (> 0.6), which implies that existing principal components explain 75% of the variance in the data. Moreover, Bartlett's sphericity test is statistically significant at a 95% significance level. Hence, the null hypothesis stating that the correlation matrix is equal to an identity matrix is rejected. Consequently, (some) variables show to have a statistically significant interrelation between each other. Finally, since both tests are verified, one can conclude that PCA is a suitable technique to be applied to the preferendum survey data.

**Table 5.7:** KMO and Bartlett's test results

| Test | Result |
|---|---|
| **KMO Measure of Sampling Adequacy** | 0.75 |
| **Bartlett's Test of Sphericity** | |
| $\chi^2$ | 23506.85 |
| DF | 435 |
| $p$-value | 0.00 |

After having verified the data adequacy for PCA, the first step of the analysis consists of deciding on the optimal number of components to extract. As already stated in chapter 4, Horn's parallel analysis is the most reliable technique to depict the right amount of dimensions. Figure 5.2 is a graphical representation of the Horn's PA results (after 1.000 iterations). The graph combines two scree plots, one on the eigenvalues resulting from PCA on the actual data (blue line), and another one on the resampled data (red lines). The eigenvalues that are higher on the original data than on the simulated data represent significant factors that should be retained. Graphically, each point on the blue line located above the red line is a considerable component to extract. In the present case, nine dimensions on the actual data line are located above the line of resampled data. However, since the last component is very close to the red line, the PCA solution for nine and eight factors should be compared in terms of interpretability.

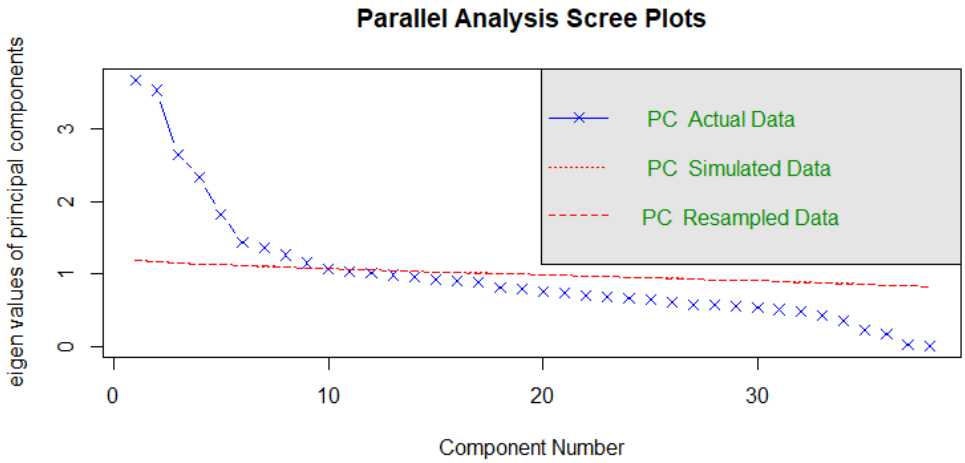


**Figure 5.1:** Horn's Parallel Analysis scree plots

Finally, one can proceed to the PCA analysis with Varimax rotation. In terms of interpretability of the factors, PCA on eight dimensions yields the best component solution. The resulting rotated component matrix, which only contains significant loadings (>0.4) for the eight components is displayed in Table 5.8. The full rotated PCA matrix including all cross-loadings is shown in Annex B. Only non-demographic variables are taken into account for PCA, excluding respondents' age, gender, educational level, and their country's economic status. Of the 35 selected non-demographic variables for PCA, four variables did not show a significant loading score in any dimension. Therefore, these variables are discarded from the model. Consequently, 31 variables are combined into components clustering at least two variables each. The eight retained dimensions explain 52% of the variability in the survey and are sorted in decreasing order.

The first dimension, named 'Household Income' is characterised by the following variables; 'Income_enough', 'Income_nowor', and 'Income_nowor_fut'. These variables describe the financial situation of a respondent's household. The second dimension is identified as the 'Environmental Education' component holding back variables related to different educational sources on climate change (e.g. parents, friends, teachers, movies). Moreover, the 'ScarceJobMen' variable, asking respondents whether women should have the same rights than men to a scares job, is included in this component. Indeed, gender equity plays a crucial role in the 21st century and can be interpreted as an indirect measure of educational level. The 'Socially Desirable Responding' component clusters five variables measuring to what extent respondents are answering the survey in a way to comply with what is socially acceptable. The latter dimension encloses the following variables; whether respondents gossip about other people's business ('DontGossip'), they avoid listening to private conversations ('AvoidListPrivConv'), and they never take things that don't belong to them ('NeverTakeThings'). Finally, the variables describing whether respondents always obey to laws ('AlwaysObeyLaws') and are not ashamed of making mistakes by covering them up ('NeverHideMistakes') are also captured by the 'Socially Desirable Responding' factor.

The fourth dimension is designated as the 'Altruism' component since the variables loading high on that dimension show the capacity of an individual to prioritise the community over the self. Among those variables, one can count, respondent's awareness of the environmental crisis and the need to take action ('Aware_CC'), their willingness to make concessions for the future ('Concess_fut'), the need to take risks ('Risktaking'), and their desire to help peers without expecting anything in return ('Helpful_noreturn'). The next component is designated as the 'Environmental Communication' dimension considering respondents' trust level in various sources (e.g. news, politicians, media, scientists, and famous people) when it comes to communication on environmental issues. The sixth dimension describes a respondent's willingness-to-pay (WTP) a price premium for alternative products and is named accordingly. Moreover, to ensure that the individual WTP is measured, respondents are additionally asked to indicate their WTP, independently of their peers' willingness to spend money to reduce climate change (i.e.' Spend_CC_ind').

The penultimate dimension is defined as the 'Social Influence' taking into account descriptive norms and individual's personality traits. On the one hand, the descriptive norm is related to the perception on socially

accepted attitudes and behaviours of young peers living in the same country ('Youg_seriousness_CC' and 'Young_spend_CC'). On the other hand, personality traits describe whether a respondent is hardworking and consistent by showing appreciation for routines ('Days_routine'). Finally, the last dimension is identified as the 'Perceived_CC_risk' factor quantifying the perceived risk of climate change on the weather.

**Table 5.8**: Rotated Principal Component results

| Construct (PC) | Survey Items | PC Loading[a] | AVE[b] | Cronbach's α |
|---|---|---|---|---|
| Household Income | Income_nowor<br>Income_enough<br>Income_nowor_fut | 0.91<br>0.86<br>0.83 | 0.12 | 0.35 |
| Environmental Education | Knwl_parents<br>Knwl_friends<br>Knwl_teacher<br>Knwl_movies<br>ScarceJobMen | 0.69<br>0.60<br>0.61<br>0.56<br>0.41 | 0.09 | 0.64 |
| Socially Desirable Responding (SDR) | DontGossip<br>AvoidListPrivConv<br>NeverTakeThings<br>AlwaysObeyLaws<br>NeverHideMistakes | 0.72<br>0.70<br>0.66<br>0.55<br>0.54 | 0.08 | 0.28 |
| Altruism | Risktaking<br>Helpful_noreturn<br>Concess_fut<br>Aware_CC | 0.74<br>0.67<br>0.66<br>0.49 | 0.06 | 0.57 |
| Environmental Communication | Trust_media<br>Trust_fam<br>Trust_poli<br>Trust_scient<br>Knwl_news | 0.76<br>0.60<br>0.58<br>0.50<br>0.49 | 0.05 | 0.59 |
| WTP | Spend_CC<br>Spend_CC_ind | 0.83<br>0.82 | 0.04 | 0.41 |
| Social Influence | Young_seriousness_CC<br>Young_spend_CC<br>Hardworking<br>Days_routine | 0.64<br>0.56<br>0.51<br>0.44 | 0.04 | 0.45 |
| Perceived_CC_risk | Extr_weath<br>Extr_weath_fut | 0.62<br>0.71 | 0.04 | 0.46 |

[a] Estimation method: maximum likelihood estimation (see Annex B for the full cross loading table)
Rotation method: Varimax with Kaiser normalisation
[b] AVE=average variance extracted

In practice, Cronbach's $\alpha$ values above 0.45 ensure acceptable internal consistency (Cortina, 1993; Nunnally, 1978; Taber, 2018). As displayed in Table 5.8, the $\alpha$ scores are acceptable for the following component constructs; environmental education and communication, altruism, social influence, and perceived CC risk. There are various explanations for low $\alpha$ values. First, since the coefficient is interpreted in terms of the number of captured items per dimension, low item numbers yield low coefficients. It is advised to have a minimum of three or four items per component to ensure a stable α score (Hair et al., 2006; Cortina, 1993). Another reason for low or negative coefficient scores is related to the need for reverse-coding of negatively phrased Likert scale items. It has been demonstrated that ignoring reverse-coding leads to a decrease in factor validity and reliability (Pilotte & Gable,1990; Marsh, 1996).

In the present case, the low alpha score on the WTP dimensions is associated with the small number of items captured by the latter component. Indeed, two items might be insufficient to establish a robust measurement scale. The two remaining low alpha values related to the household income and SDR can be argued in terms of ignored reverse-coding of several survey items. Indeed, the first component captures two survey items embodying negations which substantially change the item's meaning. Reverse-keyed items can introduce interpretation and psychometric test bias when not consistently corrected for (Suárez-Alvarez et al., 2018). Similarly, many items captured in the SDR dimension are negatively formulated. Recoding the negative-keyed Likert-scale items in the opposite direction for both factors should compensate for the low Cronbach alphas obtained without having taken into account this slight but susbtantial nuance.

## 5.4. Binary Logistic Regression Analysis

Before evaluating the binary logistic regression performed on the formerly defined components, some specifications must be provided. First, 80-20 is the considered threshold for the training and test set. Second, as outlined in chapter 2, the dependent variable, the ABG, is highly imbalanced. Only 12.43% of respondents show an ABG, whereas 87.5% align their attitude and behaviour in terms of sustainable plastic consumption. Hence, random oversampling is applied to the training data to compensate for the imbalance, resulting in a 50-50 distribution of the ABG. Lastly, Fisher's scoring algorithm used to solve the ML function and estimate the coefficients converged after four iterations.

The predictive power of the logistic regression is given by the accuracy, sensitivity and specificity rates, which are provided in Annex C. The overall model accuracy amounts to 65%. The sensitivity rate measures the positive instances that have correctly been classified as such, whereas the specificity rate considers the true negative instances. Hence, the sensitivity rate of 59% indicates that the logistic model accurately predicts the ABG in 59% of cases. Thus in the remaining 41% of considered cases, the model misclassified a positive ABG, which is still better than predictions at chance. Similarly, the 90% specificity rate correctly classifies an absence of an ABG in 90% of cases and misclassifies an ABG absence with a 10% probability. The AUCROC metric equals 0.68, indicating a reasonable predictive power of the regression on the survey. Moreover, Cox & Snell $R^2$ and the Hosmer Lemeshow (HL) test both rate the model's goodness-of-fit. The $R^2$ metric is equal to 0.23. Moreover, the HL null-hypothesis can be rejected, which implies that the model is acceptedly specified to fit the data.

Recall, the dependent variable is the ABG, and the covariates are the eight principal components obtained through PCA. Since the regression is performed on the uncorrelated components rather than on the original variables, assessing the relative variable importance through the standardised regressor estimates is a reliable technique. Table 5.3 shows that all tested components have a statistically significant effect on the ABG. Considering that some variable coefficients are positive, and others are negative, the analysis is separated accordingly. Since the reference class of the binary logistic regression is equal to an ABG absence $(ABG = 0)$, the regression predicts a respondent's likelihood to show a positive ABG $(ABG = 1)$. Hence, a negative variable coefficient indicates a negative effect on the response variable. Thus the probability of

a positive ABG is reduced, increasing the likelihood to align attitudes and behaviour. Inversely, a positive variable coefficient represents the effect of a variable to increase the ABG likelihood. Thus. To interpret the association intensity between the ABG and the negative estimates, one must consider the odds ratio.

Let's start with the negative variable coefficients. 'Household Income' displays an odds ratio of 0.832, meaning that respondents are 0.83 times less likely to show a positive ABG after a one-unit increase in the income dimension while keeping all other variables constant. In the same vein, respondents experiencing a one-unit growth in the WTP a price premium, are 0.77 times less likely to misalign their attitudes and behaviour. Similarly, altruistic respondents are 0.86 times less likely to mismatch their attitudes and behaviour in terms of sustainable plastic consumption.

Moreover, the 'Perceived CC risk' variable displays an odds ratio of 0.702. Recall, that the variable expresses the perceived impact of climate change on extreme weather events. Hence, a respondent convinced that climate change will lead to a frequency rise in intense weather days is 0.70 times more likely to act accordingly to its intentions. Lastly, respondents that value their peers' opinions and behaviour towards climate change (descriptive norm) and categorise themselves as hardworking and eager towards routines (personality traits) are 0.84 times more likely to behave accordingly to their attitudes. All in all, among the variable substantially decreasing the likelihood of an ABG, the 'Altruism' variable is the most impactful whereas the 'Perceived CC risk' variable has the lowest effect among the negative coefficient variables.

**Table 5.9:** Logistic regression results

| Model | Standardised Coefficients | | T-statistic | p-values | Significance level[1] | Odds ratio |
|---|---|---|---|---|---|---|
| | Beta | Std. Error | | | | |
| Intercept | -0.163 | 0.040 | -4.051 | 5.09e-05 | ⋆⋆⋆ | |
| Household Income | **-0.184** | 0.040 | -4.593 | 4.36e-06 | ⋆⋆⋆ | **0.832** |
| Env. Education | **0.334** | 0.037 | 8.945 | <2e-16 | ⋆⋆⋆ | **1.397** |
| SDR | **0.352** | 0.042 | 8.318 | <2e-16 | ⋆⋆⋆ | **1.422** |
| Altruism | **-0.145** | 0.037 | -3.897 | 9.73e-05 | ⋆⋆⋆ | **0.865** |
| Env. Communication | **0.178** | 0.039 | 4.506 | 6.60e-06 | ⋆⋆⋆ | **1.195** |
| WTP | **-0.262** | 0.039 | -6.716 | 1.87e-11 | ⋆⋆⋆ | **0.770** |
| Social Influence | **-0.174** | 0.041 | -4.241 | 2.22e-05 | ⋆⋆⋆ | **0.840** |
| Perceived CC risk | **-0.354** | 0.040 | -8.884 | <2e-16 | ⋆⋆⋆ | **0.702** |

[1] Significance codes : $'\star\star\star' \; p \leq 0.001$ ; $'\star\star' \; p \leq 0.01$ ; $'\star' \; p \leq 0.05$ ; $'.' \; p \leq 0.1$

Similarly, to the interpretation of the negative variable estimates, the positive coefficients exhibit the increased likelihood in an ABG presence. Hence, the hereafter interpreted variables contribute to the ABG, increasing a mismatch between individuals' attitudes and behaviour. First, 'Environmental Education' increases the ABG likelihood by 39.7%, whereas 'Environmental Communication' increases the probability of an ABG by 19.5%. Second, the variable measuring the socially desirable response tendency indicating respondent's willingness to comply with social norms ('SDR'), inflates the ABG by 42%. To conclude, the 'SDR' variable shows the most significant positive relationship with a positive ABG whereas the 'Environmental Communication' variable has the lowest impact on an ABG increase compared to the remaining positive coefficient variables.

# Chapter 6 – Discussion

This penultimate chapter discusses and interprets the results of the conducted study. The results are compared to the developed hypotheses laying the theoretical foundation for the conceptual model presented in chapter 3. Besides the evaluation of the hypotheses, the research questions unfolded in section 1 are answered.

The present analysis was conducted to gain a better understanding of the attitude-behaviour gap among adolescents when it comes to plastic consumption. The objective of the research was to unfold the variables having the most significant effect in the formation of the inconsistency between attitudes and behaviour. Moreover, these findings should increase the predictability of the misalignment and be introduced for future recommendations and interventions to promote sustainable plastic consumption. The main research questions of the paper were the following;
*"How can psychological and social factors on sustainable attitudes and behaviour measured through a preferendum survey be used to explain an attitude-behaviour gap?"* complementary to the question on *"Which factors have a statistically significant effect on the attitude-behaviour gap?"*.

The answer to the first research question was obtained by adoption of an inductive research design starting with an exploratory research approach on the available survey data. Prior literature and research designs in social research and psychometrics laid the foundation for the work towards a new conceptual model supporting the findings on the attitude-behaviour gap (Field, 2013). Through principal component analysis (PCA), new artificially created dimensions capturing the largest variability in the data set were generated. The eight resulting components are the following; household income, environmental education, socially desirable responding (SDR), Altruism, environmental communication, willingness-to-pay (WTP) for a premium price, social influence, and the perceived climate change risk. The latter factors are the most appropriate to reduce the dimensionality of the original data while maximising the ability to explain and predict the investigated attitude-behaviour gap.

## 6.1. Dimensions influencing the attitude-behaviour gap

The second research question, aiming to detect the statistically most impactful factors explaining the gap, is answered by implementing a logistic regression on the principal component covariates. The inductive research approach facilitated the identification of the developed hypothesised assumptions in the principal component dimensions, which allowed for their straightforward testing and evaluation. The findings, outlined in chapter 5, show that all tested components have a statistically significant effect on the attitude-behaviour gap. Summary table 6.1 provides a synthesised overview of the tested hypotheses along with their respective impact on the attitude-behaviour mismatch.

### 6.1.1. Environmental Education and Communication

Starting with the first and second hypotheses, treating the effect of environmental education and communication. As unfolded in the literature review, the impact of environmental education and communication on the response variable has been the object of mixed findings (Joshi & Rahman, 2015; Kollmuss & Agyeman, 2002). According to the derived hypotheses in the conceptual model, the latter variables were expected to have a statistically positive influence on the attitude-behaviour gap. Compared to the obtained results, the two first hypotheses must be rejected. At first, it seems counterintuitive, that environmental education and communication respectively increase the likelihood of an attitude-behaviour gap. However, as stated by White and her colleagues (2019), information overload can lead to the misconception, misunderstanding, and miseducation on climate change issues among young adolescents. These explanations are also applicable to environmental communication which additionally suffers from fake news, which contributes to misguiding climate change conceptions (Hung, 2014).

Appropriate appeals and message framing constitute a possible solution to tackle the here before described trade-off between lack of information and over information, misinformation and accurate information. Conveying information on the consequences of non-durable behaviour (e.g. non-recycling or reusing plastic bags and bottles) can have a positive effect on individuals' (waste management) behaviour. Prompts are especially persuasive when they are short, easy to understand and closely positioned to the location where the action takes place (e.g. residual waste containers, recycling plant, on plastic bags and bottles). Eco-labelling, certifications, framing, and packaging constitute a practical implementation to enhance consumer engagement to encourage eco-friendly and durable behaviour schemes. For marketers, companies, and entrepreneurs, these findings are highly relevant for the product pre-processing, and the marketing and communication strategy development. Hence, the form and the shape of a product, marketing or communication campaigns, certification, or labelling have a possible real effect on respondent's attitudes and behaviour (White et al., 2019). Notwithstanding, the proposed solutions to tackle a (mis)information overload must cautiously be implemented, since exaggerated these can rapidly lead to altered and unintentional outcomes, possibly contributing to an inflated attitude-behaviour gap.

### 6.1.2. Willingness to Pay Price Premium and Household Income

The third and fourth hypothesised assumptions could be combinedly discussed since they are both formulated with respect to a respondent's financial capacities and propensity to spend. Both assumptions are in line with the findings defined in chapter 3. A higher income level and a larger WTP for a premium price, respectively, decrease the probability of an ABG existence. An individual's willingness-to-pay for a price premium for alternative products is closely related to its income level. The larger the income, the less price-sensitive, the higher the possible WTP an extra price for sustainable products (Tse, 2001). However, one must consider that these hypotheses are exclusively applicable to the consumption of concrete and material goods. Hence the result can only be interpreted in terms of sustainable consumption and cannot be generalised for sustainable behaviour since durable behaviour is not restricted to physical consumption.

### 6.1.3. Perception of Risk

The fifth hypothesis is related to a respondents' perceived risk of climate change. As the assumption is supported, the larger the perceived risk, the higher the probability to behave in a sustainable manner which subsequently reduces the likelihood of an occurring attitude-behaviour gap. Considering that individuals view the consequences and risks of climate change as a temporally distant phenomenon, bringing its seriousness and urgency closer to the public's opinion is crucial to enhance sustainable behaviour (Slovic, 2000; Lowe et al., 2006; Hans & Cox, 2015). To make individuals realise that climate change is no far-off concept, concrete and present-oriented communication is crucial. By illustrating the immediate consequences of the environmental issue at a more local level (e.g. for a specific city, neighbourhood) in combination with current issues (e.g. increased frequency, severity and longevity of extreme weathers and natural catastrophes) increases the likelihood of sustainable-oriented actions and hence reduces inconsistencies between attitudes and behaviour (Monirul Qader Mirza, 2003; White et al., 2019).

### 6.1.4. Social Influence

The sixth hypothesised assumption, stating that descriptive norms have a significant effect on the response variable, is supported by the preferendum data. The present hypothesis is captured by the 'social influence' variable of the principal component construct. By having a closer look at the original survey items captured by that variable, one depicts elements related to descriptive norms, namely the ratio of peers aware of the climate change issue ('Young_seriousness_C') and the rate of youngsters willing to spend money to tackle that issue ('Young_spend_CC'). However, besides descriptive norms, personality traits are seized in the 'social influence' component, indicating whether a respondent is hardworking and enjoys routines. Recall, in chapter 3, the influence of a respondent's personality traits on the attitude-behaviour gap has been defined as being part of the 'Altruism and Personality traits' hypothesis.

Consequently, one can accept the sixth hypothesis declaring that descriptive norms have a substantial positive effect on limiting the attitude-behaviour inconsistency. Nevertheless, the positive impact, as displayed in chapter 5, is not exclusively assigned to a respondent's willingness to comply with unwritten social standards or guidelines. Indeed, the personality traits captured in the 'social influence' construct to some extend are accountable for the statistically significant effect. Since both effects cannot be decomposed, one cannot accurately affirm whether both concepts are equally important in the 'social influence' variable or whether one of both notions is dominating the other one. Further analysis and more specific survey items should be introduced to assess both assumptions separately and assess the real impact of social norms on the investigated attitude-behaviour misalignment.

### 6.1.5. Altruism and Personality traits

Hypothesis seven of the conceptual model asserts that personal norms and individuals' sense of responsibility has a significant positive effect on the attitude-behaviour gap. According to Moser (2015), personality traits as defined by the Big Five personality traits theory are seized in the personal norm notion. Compared to the analysis results in chapter 5, one can confirm the here before defined hypothesis. The 'Altruism' component has a substantial positive effect on reducing the likelihood of an attitude-behaviour presence among respondents prioritising the community over the self. The original survey items captured by that variable are all reflecting individuals' standards, sense of responsibility and degree of collectivism. Since the embodied items are exclusively quantifying the latter concepts and do not take into account a respondent's personality traits, the decisive variable effect can fully be assigned to individual's social altruism degree and sense of responsibility.

### 6.1.6. Socio-Demographic Characteristics

The penultimate hypothesis claims that respondents' socio-demographic traits have a significant impact on the explored attitude-behaviour gap. The eighth assumption on socio-demographics is divided into four sub-hypotheses, each on one demographic aspect measured by the preferendum survey.

Hypothesis 8.1 considers whether gender influences the attitude-behaviour gap. According to the examined literature, compared to men, women are more engaged in the environmental cause (White et al.,2019). This statement is testified and verified in chapter 5. Men are more likely to mismatch their attitudes and behaviour, leading to an increased probability of a positive attitude-behaviour gap.

Likewise, hypothesis 8.2 claims that a higher accomplished educational level has a positive effect on the attitude-behaviour gap. According to prior literature, the greater the knowledge about climate change, the higher the awareness of climate change consequences on the environment and hence the larger the probability of acting in an environmental-friendly way. (White et al., 2019). However, the results yielded in chapter 5 do not support the here before described assumption. There is no linear relationship between educational level and ecological consciousness. Even though it has been demonstrated that most people do not have enough knowledge about climate change and its consequences, this does not mean that those people do not engage in sustainable behaviours. It has been proven that basic education is sufficient for individuals to switch to more durable and long-lasting behaviour patterns (Kollmuss & Agyeman, 2002).

Hypothesis 8.3 states that young consumers are more likely to be pro-environmentally engaged since those individuals are often more liberal and open for changes concerning their consumption habits (White et al., 2019). From a statistical point of view, based on the preferendum data; this hypothesis can be accepted. However, it should be considered that only 25% of the total participants are older than 20 years (0.40% are older than 24 years). Hence, the conclusions drawn on the age variable are limited since the sample size is not large enough in terms of represented age ranges. Increasing the sample size and introducing more

mature and older respondents could be a possible solution to accurately interpret the influence of the age variable on the attitude-behaviour gap.

The last sub-hypothesis 8.4 regarding the income and development level of a respondent's residence country, is supported by the conducted analysis in chapter 5. Indeed, the results show that there is a significant difference in the attitude-behaviour gap between respondents living in high-income countries and the ones originated from low-income economies. Compared to high-income-level countries, respondents from low-income economies are more likely to display a positive misalignment between their attitudes and behaviour. From a statistical point of view, one can conclude that consumers originated from less developed and prosperous countries are less environmentally conscious (Kollmus & Agyeman, 2002). This finding is in line with Jambeck et al. 's (2015) study results stating that the top twenty plastic waste producers are middle-income countries. Indeed, these countries are characterised by fast-growing economies, populations and most probably waste management infrastructure unable to keep up with its population growth pace. These findings have been supported by Nguyen and his colleagues (2019), demonstrating that developing nations are at greatest risk from climate change. Moreover, the present hypothesis could be linked to the conclusions drawn on the effect of individuals' income level on the attitude-behaviour gap outlined in the third hypothesis. Indeed, the results support that a higher income level has a positive influence on the gap. Hence, there could be a possible correlation between the income and development level of a respondent's country of residence and the respondents' income level.

### 6.1.7. Socially desirable responding in self-reporting survey research

Lastly, the ninth hypothesis assesses whether the socially desirable responding (SDR) bias has a significant effect on the attitude-behaviour gap. The survey items comprised in the SDR component are measured by elements retrieved from the impression management subscale of the Balanced Inventory of Desirable Responding (BIDR) model (Steenkamp et al., 2010). The logistic regression in chapter 5 demonstrates that respondents adapted their answers to the preferendum survey to comply with social norms. Consequently, the assumption is verified, which implies that the investigated attitude-behaviour gap is overstated by respondent's willingness to paint a better picture of themselves (Auger & Devinney, 2007; Carrington et al., 2010). Hence, the attitude-behaviour gap must cautiously be interpreted, since it is inflated by respondents thriving to make a positive impression by positioning themselves as mindful and sustainable consumers while avoiding environmentally responsible actions. Socially desirable responding threatens the validity of the model since individuals do not unveil their real attitudes and behaviour. Hence there is a need to control for this measurement bias by quantifying the need for approval score. Börger (2012) defines different techniques (i.e. dichotomous and continuous scoring) to measure the SDR. However, those techniques rely on extreme answers on 5-or 7-point Likert Scale items. Since these survey item constructs are usually easily detected by respondents and consequently avoided to be answered by extreme answers, it remains very challenging to assess the real effect of SDR on the response variable (Börger, 2012).

**Table 6.1:** Summary of all tested hypotheses and their respective outcome

| Hypothesis | Description | Result |
|---|---|---|
| **H1** | Environmental knowledge and education positively influence the ABG | Rejected |
| **H2** | Trustworthy and credible communication sources on climate change shape the general perception and hence have a positive impact on the ABG | Rejected |
| **H3** | A household's income level positively influences the ABG | Supported |
| **H4** | Consumer's willingness to pay a price premium to tackle climate change has a positive effect on green consumption behaviour | Supported |
| **H5** | The higher the perceived risk of climate change, the higher the likelihood to engage in pro-environmental behaviour. | Supported |
| **H6** | Descriptive norms have a significant positive effect on the ABG | Supported |
| **H7** | Personal norms and responsibility have a positive effect on bridging the ABG | Supported |
| **H8** | Socio-demographic traits have a significant impact on the ABG | Mitigated |
| **H8.1** | Women are more environmentally engaged than men. Hence the female gender has a positive influence on the ABG | Supported |
| **H8.2** | Higher completed education is positively correlated with the ABG | Rejected |
| **H8.3** | Young consumers are more likely to engage in pro-environmental behaviour | Supported |
| **H8.4** | The income and development level of consumers' country of residence has an impact on the ABG | Supported |
| **H9** | The bias introduced by social desirability responding might amplify the ABG | Supported |

# Chapter 7 – Conclusion

The closing chapter of the present research comprises the limitations and future research recommendations which constitute the foundation for an enhanced expansion of the here defined conceptual framework. Besides, the last subsection of this chapter consists of the conclusion of the analysis.

## 7.1. Limitations and Future Research Recommendations

The first limitation of the present analysis concerns the research settings of the preferendum survey, which is the basis of the entire paper. The target group of the worldwide youth survey is Gen Z and Y, so respondents aged between 13 and 24 years from all over the world. Hence, the research results obtained on that particular demographic sample cannot be extrapolated to a more diversified population in terms of age ranges. More generalised conclusions on the attitude-behaviour gap could be achieved by broadening the target audience by scrutinising larger age groups. The lack of generalisation linked to the niche target group can simultaneously be viewed as a limitation and a specificity of the present analysis. Indeed, there is a remaining lack in attitude-behaviour gap literature on that specific demographic sample which is of particular importance since it represents the consumers of tomorrow (Vermeir & Verbeke, 2008). In the same vein, the findings related to respondents' residence country cannot be postulated for all adolescents. Since some countries are largely overrepresented (e.g. India, United Kingdom, Indonesia), the present research did not proceed to a cross-country analysis. The current study synthesised the country variable based on their income level as benchmarked by the World Bank, to reduce the dimensionality of the data and enhance the reliability and accuracy of the results drawn on the unweighted country variable. Consequently, further research should be conducted on the correct weighting of the under and over-represented countries to achieve a cross-country analysis. Pairwise, the research sample should be enlarged and diversified to cover a more balanced country representation.

Secondly, the specific research perimeter of the attitude-behaviour gap in terms of plastic consumption and waste production is relatively uncommon. Although one can retrieve some valuable insights from such peculiar research, the survey should accordingly be entirely dedicated to the defined subject of attitudes and behaviour towards plastic use. In the present study, the focus of sustainable plastic behaviour is investigated by subsetting a survey item quantifying eco-friendly behaviour in diverse fields. Therefore, the usefulness and accuracy of the findings are limited for practical implementations. An extensive survey devoted to the defined subject areas should be developed to get a better and more generalized understanding of the main drivers of the attitude-behaviour gap in the niche sector of plastics use.

Thirdly, the online nature of the preferendum questionnaire implemented through the Qualtrics platform is accountable for some biases. The first bias concerns the reduced ability to control for the survey environment. Since the survey was only available through the Facebook platform and required the creation of a user account on the KidsRights' webpage, the participation ratio to the study was lower than expected. To increase the click rate, and thus broaden and diversify the sample size, the accessibility to the survey should be facilitated. Moreover, the mediation through Facebook platform should be avoided since

individuals are suspicious and sceptical when it comes to Facebook's data privacy protection (Feng and Xie, 2014). Moreover, making the survey available in a broad panoply of languages would have a positive effect on the survey's click ratio.

An additional bias introduced based on the survey's self-reporting nature, is the socially desirable responding (SDR) distortion. The SDR quantifies to what extend the attitude-behaviour gap might be skewed by the desire to comply with social norms and expectations. Even though the latter one is taken into account in the present research, the mediation effect of the SDR on each influencing factor explaining the gap could be assessed. The last introduced bias is related to the non-reverse-coding of negatively phrased survey items. Consequently, the reliability and internal consistency of the constructed principal component factors might be jeopardised (Pilotte & Gable,1990; Marsh, 1996). Hence, negative-keyed Likert-scale items should be reversed to the opposite direction to compensate for the model reliability loss introduced by the unawareness of this slight but powerful nuance.

Finally, the defined conceptual framework could be extended to get a more detailed understanding of the attitude-behaviour gap. Besides assessing and evaluating the developed hypotheses embodied in the framework, some mediation and moderation effects could be considered. For instance, as concluded, the willingness-to-pay for a premium price, individuals' income level, and the income and development level of a respondent's residence country similarly affect the attitude-behaviour gap. Hence, there might be an association between those factors affecting the response variable. The same statement holds for the respondent's highest achieved educational level and environmental knowledge. Consequently, by carrying out some extensive research on possible mediation and moderation effects, a much more detailed and meaningful conceptual framework could be investigated, yielding more robust and concrete conclusions.

## 7.2. Conclusion

The purpose of the present research was to study the relationship between the attitude-behaviour gap and the factors explaining this inconsistency. Moreover, the demographic characteristics of adolescents misaligning their attitudes and behaviour were investigated. Hence, the demographic profile of environmentally conscious consumers mismatching their intentions and behaviour can be drawn. Social, behavioural research enabled the investigation in the most influencing factors aiming to explain and predict the attitude behaviour gap, alongside with the exploration of the association between socio-demographic variables and the misalignment.

The generated dimensions derived through principal component analysis intervening in the attitude-behaviour gap are the following; income level, environmental education and communication, altruism, willingness-to-pay for a price premium, social influence, perceived climate change risk, and socially desirable responding. The research findings show that all enumerated component factors are statistically significant and contribute to the explanation of the attitude-behaviour gap. By comparing socio-demographic characteristics among adolescents showing a positive attitude-behaviour gap and those who do not, several conclusions can be made.

Firstly, compared to women, men are more likely to misalign their attitudes and behaviour. Secondly, the association between age and the attitude-behaviour gap must cautiously be interpreted. From a statistical point of view, young consumers are less likely to display an attitude-behaviour gap. However, since the considered population sample barely captures respondents older than 24 years, no generalised conclusion can be made on the effect of age on the misalignment. Thirdly, respondents from low-income countries are more likely to show a mismatch between their intentions and actions, compared to residents from high-income countries. However, this finding is uncertain since there is a considerable imbalance when it comes to country representation among the data. Fourthly, the respondent's highest educational achievement level is not associated with the attitude-behaviour gap. Hence, the respondent's educational background is not relevant to the investigated inconsistency. Finally, from the research results and based on the preferendum data, one can conclude that consumers misaligning their attitudes and behaviour in terms of plastic consumption are typically men aged between 20 and 23 years originated from a low- to lower-middle-income country.

Given the limitations encountered throughout the research, it is suggested for future research to analyze a more extensive and subject-specific survey. Indeed, the investigated preferendum survey embodies various possible research fields. Hence, the data was not precise enough on the plastic consumption and waste production to draw robust and generalised conclusions. Furthermore, the outlined conceptual model should be extended by testing possible mediation and moderation effects on the attitude-behaviour gap.

# ANNEXES

## ANNEX A

### A.1. Descriptive statistics per variable in the preferendum survey

**Table A.1**: Descriptive statistics per variable (socio-demographics excluded)

| Survey Items | Acronym | Measurement | Mean |
|---|---|---|---|
| Is CC a serious problem | Att1 | 7-point LS | 6.72 |
| Is extra action needed to tackle CC | Att2 | 7-point LS | 6.68 |
| Did you recycle/reuse plastic in the past 12 months | Bhv | Dichotomous | 0.87 |
| **How much did you learn about CC from** | | | |
| Teachers | Knwl_teacher | 7-point LS | 4.38 |
| Parents | Knwl_parents | 7-point LS | 3.96 |
| Friends | Knwl_friends | 7-point LS | 4.48 |
| Other people | Knwl_others | 7-point LS | 4.91 |
| News | Knwl_news | 7-point LS | 5.73 |
| Movies | Knwl_movies | 7-point LS | 4.50 |
| **How much do you trust in** | | | |
| Politicians | Trust_poli | 7-point LS | 2.31 |
| Media | Trust_media | 7-point LS | 3.44 |
| Scientists | Trust_scient | 7-point LS | 5.99 |
| Famous people | Trust_fam | 7-point LS | 3.41 |
| **How willing are you to spend money to** | | | |
| Reduce CC | Spend_CC | 7-point LS | 5.07 |
| Reduce CC, even if others don't | Spend_CC_ind | 7-point LS | 5.05 |
| **What % of young people in your country** | | | |
| Think CC is a serious problem | Young_seriousness_CC | 10-point LS | 5.43 |
| Are willing to spend money to tackle CC | Young_spend_CC | 10-point LS | 3.48 |
| Compared to other young people in your country, how much do you know about CC | Aware_CC | 7-point LS | 5.52 |
| **How many days with extreme weather** | | | |
| Did you experience in the past 12 months | Extr_weath_past | 6 factor levels | 4.09 |
| Do you expect for the next 12 months | Extr_weath_fut | 3 factor levels | 2.37 |
| **In general, how willing are you to** | | | |
| To give up smth today to benefit in the future | Concess_fut | 10-point LS | 8.18 |
| To take risks | Risktaking | 10-point LS | 7.82 |
| To help others without expecting smth in return | Helpful_noreturn | 7-point LS | 8.72 |
| **Household Income, in my family** | | | |
| We have enough money to buy things I want | Income_enough | 7-point LS | 4.12 |
| We don't need to worry too much about paying our bills | Income_nowor | 7-point LS | 3.90 |
| We won't have to worry about money in the future | Income_nowor_fut | 7-point LS | 3.88 |
| **Social desirability** | | | |
| I never hide mistakes | NeverHideMist | 7-point LS | 4.87 |
| I never take things that don't belong to me | NeverTakeThings | 7-point LS | 5.85 |
| I don't gossip about other people's business | DontGossip | 7-point LS | 5.09 |
| I always obey laws, even if I am unlikely to get caught | AlwaysObeyLaws | 7-point LS | 5.46 |
| When I hear people talking privately, I avoid listening | AvoidListPrivConv | 7-point LS | 5.19 |
| **Personality survey items** | | | |
| Men have more right to a scarce job than women | ScarceJobMen | 7-point LS | 2.26 |
| I mostly work very hard | Hardworking | Dichotomous | 0.77 |
| I eat more than I should | Eat_habit | Dichotomous | 0.34 |
| I like to visit new places | Visit_places | Dichotomous | 0.97 |
| Advertisements are silly | Ads_silly | Dichotomous | 0.38 |
| My days follow a routine | Days_routine | Dichotomous | 0.65 |

## A.2. List of discarded variables based on prior knowledge and intuition:

- **Q1.3. Were you born in the country you currently live?**
  Since the country of residence (Q1.1) is retained to the analysis, Q1.3 does not add any valuable value, since the country of birth is neither related to the actual country of residence nor the nationality of the respondent.
- **Q3.0 16 options on how to deal with climate change from which respondents should choose their favourite one**. (NB: only 4 options were displayed)
  Since the respondents did not see all 16 options, it does not represent their actual preferred option to tackle climate change. The question will be used for descriptive purposes in the conclusion of the present paper.
- **Q3.2 Do you want to submit other innovative ideas on how to tackle climate change?**
  **& Q3.2.1 If the response is yes, the subsequent submitted idea.**
  This question will be included in the conclusion of the paper to investigate some innovative ideas proposed by respondents.
- **Q5.1 How bad or good are you at mathematics?**
  **& How bad or good is your understanding of the English language?**
  Both questions were removed. Mathematical knowledge does not necessarily have an influence on respondents in terms of pro-environmental attitudes and/or behaviour. Concerning respondent's English level, since the survey was exclusively shown to respondents having the default language setting of their Facebook account set to English, one can assume them having a high enough understanding of English to take part in the survey.
- **Q5.3 Would you like to answer few more questions?**
  This question aims to partition the survey into the mandatory and the optional block. Since further analysis only considers complete answers on both blocks, the present question can be removed.

## A.3. Country classification by income level (World Bank)

The respondents of the preferendum survey reside in 116 different countries (out of 159 possible countries). However, the 116 country levels are not equally filled. For illustration, the majority of respondents (38.29%) originate from India. The second-largest represented country is the United Kingdom of Great Britain (14.45%) closely followed by Indonesia (10.55%). Hence, the three most represented countries account for 63.29% of the total respondents. Consequently, many country levels are represented by very few to single observations which can be noisy data to some extent. To avoid further techniques to be model and fitted on random noise and to overstate the weight of the overrepresented country levels, the country variable should be synthesised to increase understanding and interpretability of respondent's country of residence. The World Bank suggests a country classification by income level based on the threshold presented in Table A.1 (World Bank, 2019). This country classification benchmark will be used to replace the individual country levels in the original preferendum survey.

**Table A.3**: World bank benchmark for country classification (2019-2020)

| Threshold | Acronym | Gini/Capita/$ |
|---|---|---|
| Low Income | LI | <1.206 |
| Lower-middle income | LMI | 1.026 - 3.995 |
| Upper-middle income | UMI | 3.996 – 12.375 |
| High income | HI | >12.375 |

## ANNEX B

**Table B.1:** Full Rotated PCA results

| Survey Items | RC4 | RC2 | RC3 | RC8 | RC5 | RC1 | RC6 | RC7 |
|---|---|---|---|---|---|---|---|---|
| Knwl_teacher | -0.04 | **0.61** | 0.03 | 0.06 | 0.20 | -0.08 | 0.07 | -0.05 |
| Knwl_parents | 0.09 | **0.69** | 0.10 | 0.03 | 0.07 | -0.05 | 0.07 | -0.02 |
| Knwl_friends | 0.04 | **0.60** | -0.03 | 0.03 | 0.04 | 0.08 | 0.19 | 0.15 |
| Knwl_news | 0.03 | 0.16 | 0.00 | 0.08 | **0.49** | -0.08 | 0.00 | 0.03 |
| Knwl_movies | -0.08 | **0.56** | 0.06 | 0.09 | 0.22 | -0.06 | -0.27 | 0.02 |
| Trust_poli | 0.08 | 0.10 | -0.03 | -0.13 | **0.58** | 0.09 | 0.19 | -0.20 |
| Trust_media | 0.02 | 0.13 | 0.05 | -0.04 | **0.76** | 0.00 | 0.07 | -0.10 |
| Trust_scient | 0.02 | -0.12 | 0.02 | 0.13 | **0.50** | 0.07 | 0.09 | 0.27 |
| Trust_fam | 0.00 | 0.26 | 0.05 | -0.07 | **0.60** | 0.10 | -0.07 | -0.05 |
| Spend_CC | 0.17 | -0.08 | 0.01 | 0.27 | 0.08 | **0.83** | 0.02 | 0.06 |
| Spend_CC_ind | 0.12 | -0.06 | 0.05 | 0.31 | 0.07 | **0.82** | -0.02 | 0.05 |
| Young_seriousness_CC | 0.04 | 0.38 | -0.06 | -0.14 | 0.01 | 0.31 | **0.64** | 0.05 |
| Young_spend_CC | 0.05 | 0.42 | -0.03 | -0.11 | 0.08 | 0.38 | **0.56** | -0.09 |
| Aware_CC | 0.11 | 0.06 | 0.02 | **0.49** | 0.10 | 0.06 | 0.02 | 0.29 |
| Extr_weath | -0.05 | 0.20 | 0.12 | 0.01 | -0.01 | 0.01 | -0.14 | **0.62** |
| Extr_weath_fut | 0.04 | -0.10 | -0.05 | 0.04 | -0.06 | 0.07 | 0.02 | **0.71** |
| Concess_fut | 0.08 | -0.02 | 0.11 | **0.66** | -0.01 | 0.25 | 0.00 | 0.10 |
| Risktaking | -0.02 | 0.11 | 0.07 | **0.74** | -0.06 | 0.16 | -0.07 | -0.06 |
| Altruism | -0.02 | -0.01 | 0.19 | **0.67** | -0.04 | 0.06 | 0.08 | -0.02 |
| Income_enough | **0.86** | -0.02 | -0.03 | 0.04 | 0.04 | 0.11 | 0.07 | 0.08 |
| Income_nowor | **0.91** | -0.01 | -0.02 | 0.02 | 0.04 | 0.07 | 0.04 | 0.02 |
| Income_nowor_fut | **0.83** | 0.06 | 0.02 | 0.05 | 0.05 | 0.06 | 0.01 | -0.07 |
| Social_Desi_1 | -0.04 | 0.22 | **0.54** | 0.22 | 0.03 | -0.05 | 0.00 | -0.11 |
| Social_Desi_2 | 0.03 | -0.12 | **0.66** | 0.04 | 0.02 | 0.08 | 0.11 | 0.09 |
| Social_Desi_3 | -0.02 | 0.09 | **0.72** | 0.06 | -0.05 | 0.04 | -0.03 | -0.06 |
| Social_Desi_4 | 0.07 | -0.09 | **0.55** | 0.11 | 0.14 | -0.05 | 0.25 | 0.04 |
| Social_Desi_5 | -0.07 | 0.14 | **0.70** | 0.01 | 0.00 | 0.01 | -0.10 | 0.04 |
| Masculinity | -0.05 | **0.41** | 0.17 | -0.19 | 0.14 | -0.04 | -0.18 | -0.35 |
| Hardworking | -0.05 | 0.00 | 0.07 | 0.28 | 0.01 | -0.10 | **0.51** | -0.05 |
| Visit_places | 0.01 | -0.02 | -0.03 | 0.25 | 0.07 | -0.21 | 0.20 | -0.07 |
| Days_routine | 0.10 | -0.07 | 0.13 | 0.00 | 0.14 | -0.12 | **0.44** | 0.00 |
| AVE | 0.12 | 0.09 | 0.08 | 0.06 | 0.05 | 0.04 | 0.04 | 0.04 |
| Cumulative Var | 0.12 | 0.21 | 0.29 | 0.35 | 0.40 | 0.44 | 0.48 | 0.52 |

Estimation method: maximum likelihood estimation
Rotation method: Varimax with Kaiser normalisation
AVE=average variance extracted

## ANNEX C

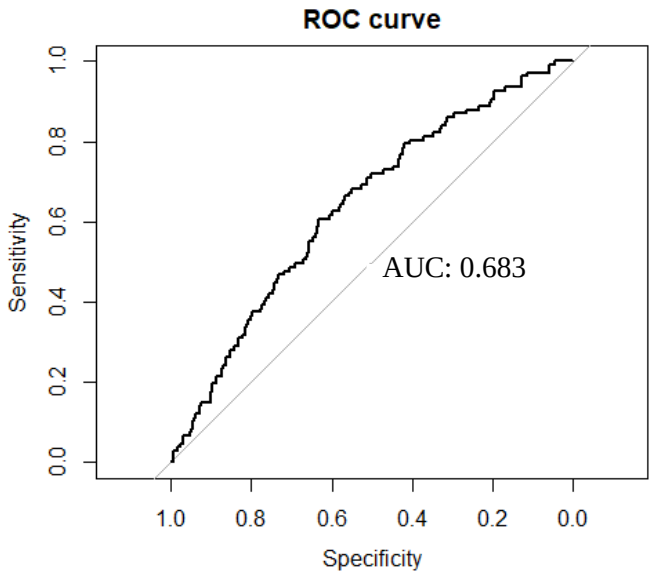


**Figure C.1**: ROC curve based on the binary logistic predictions

**Table C.2**: Confusion matrix on the logistic regression

| | **N=743** | **Actual** | |
|---|---|---|---|
| **Predicted** | | 0 | 1 |
| | 0 | TN=411 | FN=44 |
| | 1 | FP=216 | TP=63 |

$$AccuracyRate(ACC) = \frac{TP + TN}{TP + TN + FN + FP} = 0.646$$

$$Sensitivity(SN) = \frac{TP}{TP + FN} = 0.589$$

$$Specifity(SP) = \frac{TN}{TN + FP} = 0.655$$

## REFERENCES


Adane, L., & Diriba, M. (2011). Survey on the usage of plastic bags, their disposal and adverse impacts on environment: A case study in Jimma City, Southwestern Ethiopia. *Journal of Toxicology and Environmental Health Sciences* 3.8 (2011): 234-248.

Aguilera, A. M., Escabias, M., & Valderrama, M. J. (2006). Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis*, *50*(8), 1905-1924.

Ajzen, I. (1991). The theory of planned behaviour. *Organisational Behaviour and Human Decision Processes,* 50(2), 179-211.

Auger, P., & Devinney, T. M. (2007). Do what consumers say matter? the misalignment of preference with unconstrained ethical intentions. *Journal of Business Ethics*, *76*(4), 361–383.

Azen, R., & Traxel, N. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioural Statistics*, *34*(3), 319-347.

Blaikie, N. (2003). *Analysing quantitative data: From description to explanation*. Sage.

Boucher, J., & Billard, G. (2019). The challenges of measuring plastic pollution. *Field Actions Science Reports. The journal of field actions*, (Special Issue 19), 68-75.

Boucher, J., & Friot, D. (2017). *Primary microplastics in the oceans: a global evaluation of sources* (pp. 2017-002). Gland, Switzerland: IUCN.

Boulstridge, E., & Carrigan, M. (2000). Do consumers really care about corporate responsibility? Highlighting the attitude-behaviour gap. *Journal of communication management*, *4*(4), 355-368.

Börger, T. (2012). *Social desirability and environmental valuation* (p. 265). Peter Lang International Academic Publishers.

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of personality assessment*, *85*(1), 17-32.

Campbell, B., Khachatryan, H., Behe, B., Hall, C., & Dennis, J. (2016). Crunch the can or throw the bottle? Effect of "bottle deposit laws" and municipal recycling programs. *Resources, Conservation And Recycling*, *106*, 98-109.

Carrington, M. J., Neville, B. A., & Whitwell, G. J. (2010). Why ethical consumers don't walk their talk: towards a framework for understanding the gap between the ethical purchase intentions and actual buying behaviour of ethically minded consumers. *Journal of Business Ethics*, *97*(1), 139–158.

Chapman, C., & Feit, E. M. (2015). *R for marketing research and analytics* (p. 195e223). New York, NY: Springer.

Carrington, M. J., Neville, B. A., & Whitwell, G. J. (2014). Lost in translation: Exploring the ethical consumer intention–behaviour gap. *Journal of Business Research*, *67*(1), 2759-2767.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, *58*(6), 1015.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98.

Cowe, R., Williams, S. (2000). Who Are the Ethical Consumers?. The Co-operative Bank, Manchester. Retrieved May 18, 2020, from:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.458.5207&rep=rep1&type=pdf

Evans, G. W. (2019). Projected behavioral impacts of global climate change. *Annual review of psychology*, *70*, 449-474.

Fang, W. T., Ng, E., Wang, C. M., & Hsu, M. L. (2017). Normative beliefs, attitudes, and social norms: People reduce waste as an index of social relationships when spending leisure time. *Sustainability*, *9*(10), 1696.

Feng, Y., & Xie, W. (2014). Teens' concern for privacy when using social networking sites: an analysis of socialisation agents and relationships with privacy-protecting behaviours. *Computers in Human Behaviour*, *33*, 153–162.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.

Fishbein, M., Ajzen, I. (1980). Understanding attitudes and predicting social behaviour. Englewood Cliffs, NJ: Prentice Hall

Gaudelli, J. (2009). The greenest generation: The truth behind millennials and the green movement. *Advertising Age*.

Hadi, N. U., Abdullah, N., & Sentosa, I. (2016). An easy approach to exploratory factor analysis: Marketing perspective. *Journal of Educational and Social Research*, *6*(1), 215.

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). Multivariate Data Analysis (6th ed.). New Jersey: Pearson Educational, Inc.

Hansen, A., & Cox, R. (Eds.). (2015). *The Routledge handbook of environment and communication*. Routledge.

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

Harpe, S. E. (2015). How to analyse Likert and other rating scale data. *Currents in Pharmacy Teaching and Learning*, *7*(6), 836-850.

Hauck Jr, W. W., & Donner, A. (1977). Wald's test as applied to hypotheses in logit analysis. *Journal of the american statistical association*, *72*(360a), 851-853.

Heaney, J. (2007). Generations X and Y's internet banking usage in Australia. Journal of Financial Services Marketing, 11(3), 196–210.

Heidbreder, L., Bablok, I., Drews, S., & Menzel, C. (2019). Tackling the plastic problem: A review on perceptions, behaviours, and interventions. *Science of the Total Environment*, *668*, 1077-1093.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hume, M. (2010). Compassion without action: Examining the young consumers consumption and attitude to sustainable consumption. Journal of World Business, 45, 385–394

Hung, C. C. (2014). *Climate change education: Knowing, doing and being*. Routledge.

IBM, (2019). KMO and Bartlett's Test. Retrieved June 12, 2020, from https://www.ibm.com/support/knowledgecenter/SSLVMB_23.0.0/spss/tutorials/fac_telco_kmo_01.html

Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., Law, K.L. (2015). Plastic waste inputs from land into the ocean. *Science*, *347*(6223), 768-771.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, pp. 3-7). New York: springer.

Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical education*, *38*(12), 1217-1218.

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *31*(3), 300-303.

Jolliffe, I. T. (1986). Principal components in regression analysis. In *Principal component analysis* (pp. 129-155). Springer, New York, NY.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Joshi, Y., & Rahman, Z. (2015). Factors affecting green purchase behaviour and future research directions. *International Strategic management review*, *3*(1-2), 128-143.

Kaiser, H. F. (1991). Coefficient alpha for a principal component and the Kaiser-Guttman rule. *Psychological reports*, *68*(3), 855-858.

KidsRights. (2019). First ever worldwide youth preferendum on climate. Retrieved June 4, 2020, from https://kidsrights.org/news/first-ever-worldwide-youth-preferendum-on-climate/

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political analysis*, *9*(2), 137-163.

Kolkailah, S.K., Aish, E.A., & Bassiouny, N. (2012). The impact of corporate social responsibility initiatives on consumers' behavioural intentions. International Journal of Consumer Studies, 2, 19-35

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behaviour?. *Environmental education research*, *8*(3), 239-260.

Lowe, T., Brown, K., Dessai, S., de França Doria, M., Haynes, K., & Vincent, K. (2006). Does tomorrow ever come? Disaster narrative and public perceptions of climate change. *Public understanding of science*, *15*(4), 435-457.

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of personality and social psychology*, *70*(4), 810.

Matsunaga, M. (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International journal of psychological research*, *3*(1), 97-110.

Mirza, M. M. Q. (2003). Climate change and extreme weather events: can developing countries adapt?. *Climate policy*, *3*(3), 233-248.

Moser, A. K. (2015). Thinking green, buying green? Drivers of pro-environmental purchasing behaviour. *Journal of consumer marketing*.

Nunnally, J. C. (1978). Psychometric Theory. McGraw-Hill Book Company, pp. 86-113, 190-255.

Ocean Conservancy International Coastal Cleanup. Retrieved from: https://oceanconservancy.org/trash-free-seas/international-coastal-cleanup/annual-data-release/ (accessed on 11 May 2020).

Pallant, J. (2013). SPSS Survival Manual. A step by step guide to data analysis using SPSS. McGraw-Hill Education (UK).

Parker, L. (2019). The world's plastic pollution crisis explained. Retrieved May 11, 2020, from https://www.nationalgeographic.com/environment/habitats/plastic-pollution/

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, *96*(1), 3–14.

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, *49*(4), 974-997.

Pilotte, W. J., & Gable, R. K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, *50*(3), 603-610.

PlasticsEurope. (2019). Plastics - The facts 2019 An analysis of European plastics production, demand and waste data analysis of European plastics production, demand and waste data

R Core Team (2019). R: A Language and environment for statistical computing. R - - Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Salas-Eljatib, C., Fuentes-Ramirez, A., Gregoire, T. G., Altamirano, A., & Yaitul, V. (2018). A study on the effects of unbalanced data when fitting logistic regression models in ecology. *Ecological Indicators*, *85*, 502-508.

Schreiber-Gregory, D. N. (2018). Logistic and Linear Regression Assumptions: Violation Recognition and Control. *Henry M Jackson Foundation*.

Schwarz, N. (1999). Self-reports: how the questions shape the answers. *American Psychologist*, *54*(2), 93–105.

Slovic, P. E. (2000). The perception of risk. *Earthscan publications* (pp.511)

Steenkamp, J. B. E., De Jong, M. G., & Baumgartner, H. (2010). Socially desirable response tendencies in survey research. *Journal of Marketing Research*, *47*(2), 199-214.

Stevens, J. (2002). Applied Multivariate Statistics for the Social Sciences: Vol. 4th ed. *Lawrence Erlbaum Associates, INC*

Suárez-Alvarez, J., Pedrosa, I., Lozano Fernández, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273-1296.

Tse, A.C.B. (2001), "How much more are consumers willing to pay for a higher level of service? A preliminary questionnaire", *Journal of Services Marketing,* Vol. 15 No. 1, pp. 11-17 74

Vermeir, I., & Verbeke, W. (2008). Sustainable Food Consumption Among Young Adults in Belgium: Theory of Planned Behaviour and the Role of Confidence and Values. *Ecological Economics 64(3)*, 542–553.

White, K., Rishad, H., & Hardisty, D. (2019). How to SHIFT Consumer Behaviours to be More Sustainable: A Literature Review and Guiding Framework, *Journal of Marketing*, *83*, *3*, 22-49.

World Bank. (2019). The World Bank's classification of countries by income (English). Washington, D.C.: World Bank Group. Retrieved June 24, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Zwick W. R., & Velicer W. F. (1982). Factors influencing four rules for determining the number of components to retain. Multivariate Behavioural Research, 17, 253–269. http://dx.doi.org/10.1207/s15327906mbr1702_5