

Master Thesis

Grigor Dimitrov

29 November, 2016

Abstract

This paper investigates the relationship between consumers' personality traits and the extent travel related online behavior. The results prove the positive relationship between general risk and general uncertainty attitudes and online travel related behavior associated with “unique domains”, “time” and “pageviews”. While the results on mobile show the opposite direction of the effect associated with uncertainty attitude and no significant relationship between the investigated metrics and risk attitude. The novelty of the paper lays in the methodological contributions more familiar in the practice as “data fusion”, the methodology utilizes behavioral data to gather information about the observable online behavior and survey data to assess consumers' unobservable personality traits. Finally, the results reveal a significant difference between desktop and mobile online behavior with should be further investigated.

Contents

1	Introduction	3
2	Literature review	7
2.1	Travel related activities and information search	7
2.2	Risk and uncertainty	10
2.3	Risk and uncertainty in tourism	11
2.4	Big Five Factors Inventory (BFI)	13
2.5	Personality traits and tourism information search	14
2.6	Hypotheses	15
3	Methodology	16
3.1	Introduction	16
3.2	Behavioral data & Technology	17

3.3	Steps	17
4	Results	26
4.1	Survey Data	26
4.2	Main results	28
4.3	Conclusion	32
5	Conclusions	34
5.1	Discussion	34
5.2	Contribution and Implications	36
5.3	Limitations & Recommendations	37
	References	39

1 Introduction

In the last few years we created 90% of the world's data, this claim has been made in 2013 and it still consistent today. More precisely it is estimated that every two years in the last three decades the amount of data increases by ten times. Chatfield (2016) . The advantages of such "big data" include possibilities to uncover heterogeneities and subtle population patterns" which are difficult to obtain using self-reported data Fan, Han, and Liu (2014). However, "big data" has been underutilized due to computational and statistical challenges Ammu and Irfanuddin (2013) Tole and others (2013) Labrinidis and Jagadish (2012). Moreover, Nunan and Di Domenico (2013) claim that there are high costs associated with analyzing and storing big data. One of the instances of "big data" is called behavioral data. In the context of this research "behavioral data" refers to recording human behavior, more precisely I am referring to "online behavioral data" which is clickstream data from subject's browser interactions and the Internet.

The survey questionnaires are foremost used when conduction primary market research data collection. Since the emergence of Web 2.0, the online survey made its way as it is a cheap and convenient form or data collection Evans and Mathur (2005). With the increased penetration of the internet, there were more people online and this availability gave survey market research a huge reach. For example, in 2002 500million dollars were spent on online surveys while this figure doubled in 2004 Evans and Mathur (2005).

However, 20 years ago when the survey plunged, our economic behavior wasn't focused that much online and the Internet was used mainly for information purpose. Surveys conducted back then explored topics related to our consumer offline behavior replacing standard paper and pencil interviewing popular in the 80's. Now more than ever our economic activity is focused online, our decision-making process is reflected online and it happens that online we can observe and capture information with ease. Consequently, there is an increasing availability of data that reflects this behavior and our online activities are becoming more valuable to researchers. Due to the economic value that online behavior can drive, there are emerging passive data collection technologies. Such technologies allow us to observe and capture all the online consumer behavior.

Behavioral data is purely observational, such data is more accurate in terms of describing actual behavior i.e. information gathering, research of alternatives and transactions which are the ultimate reflection of preference. On the other hand, survey data is

relevant when we have to assess unobservable information such as people's attitudes, motivations, and opinions. Furthermore, it is primary data collected for the specific research goals Hox and Boeije (2005) Glass (1976). These two types of data sources have their limitations and researchers are utilizing them separately, however, up to date, there is a limited amount of research that combines both sources of data. With the aid of passive data collection panels, this project provides a rare opportunity to match an actual online behavior with a subject behind it. That is, by combining passive online behavioral data with survey data, in this research I will be able to access information on both consumers' actual online behavior as well access unobservable information about consumers' personality traits.

The context of this research is the online travel industry amongst Dutch consumers. Tourism is highly competitive and fragmented market. It has been disrupted by the Internet at large, the disruption process has started with the emergence of the Internet and continues today. Before the penetration of the Internet the market was dominated by "high street" travel agencies, afterward the market was revolutionized by online travel agencies and direct distribution of travel services as accommodation and transportation.

We are witnessing dynamic segments of consumers emerged because of the technological advancements with constantly changing needs (Xiang, Magnini, and Fesenmaier 2015). This domain has been chosen due to the fact is one of the fastest growing online industry. For example, in the Netherlands for 2 out of 3 trips (69%) consumers book accommodations online according to Eurostat data (Eurostat 2016) Moreover, 40% of the Europeans used the internet for travel related purposes, in the Netherlands this number is higher than the average. Online presence of accommodation businesses in 2015 accounts for 95% of all enterprises compared to 75% for the whole economy. According to Euromonitor data, the sector is one of the fastest growing, rising from 899 million trips in 2009 to 1.4 billion trips in 2019 (Euromonitor 2015). Furthermore, there is an increasing availability of options and all sorts of services bringing the consumers and offers closer together manifesting the ongoing process of disintermediation. Xiang, Magnini, and Fesenmaier (2015) notes that there is growing "bifurcation" or a split among the traditional online travelers to users of traditional travel products and people seeking deeper and authentic experiences. The authors point out that understanding how contemporary travelers use the internet is an important foundation for building successful communication strategies by the business stakeholders. The importance of the internet has been attributed to three main factors, the extensive amount of

travel related information, the development of social networks, travel related social networks and peer-to-peer travel offerings where the user can exchange travel services and experiences and the mobile computing, smartphones in particular (Euromonitor 2015).

The increasing availability of travel related options online and increasing disintermediation lowers the costs of the travel related services and activities. However, this also makes the decision-making process more demanding as consumers themselves are primarily responsible for choosing the best option rather than using an advice from a travel service provider. Thus, the decisions related to travel involve a lot of risk and uncertainties. Arguably one of the risk and uncertainty reducing instruments that consumers use in decision making related to travel is an information search. Urbany, Dickson, and Wilkie (1989) introduces a body of research to support the link between uncertainty and information search. When it comes to perceived risks, in a meta-analysis Gemünden (1985) review one hundred empirical findings to conclude that information search is one of the risk reducing instruments consumers rely on when committing to purchases of complex products.

In an exploratory analysis, Roehl and Fesenmaier (1992) identify seven types of perceived risks related to leisure travel. Namely, equipment risks, financial risks, physical risks, psychological risks, satisfaction risks, social risks and time risks. Lepp and Gibson (2003) look at the general perceptions of risk and uncertainty from the perspective of travel motivations for novelty against familiarity. Their findings suggest that familiarity seekers are more risk averse comparing to less risk averse novelty seekers. With regards to risk attitude and tourism information search, Quintal, Lee, and Soutar (2010) examine the impact of risk and uncertainty avoidance on tourists' information search. The team distinguishes risk from uncertainty and investigate whether the two constructs have a different impact on the information search in terms of a number of sources. Their findings suggest that uncertainty avoidance has a positive significant relationship with the extent of information search while this is not valid for the risk averseness.

The following research further explores the relationship between risk and uncertainty attitudes and different instances of online information search behavior. More precisely, this research explores the relationship between risk and uncertainty attitudes, and information search behavior in tourism context while controlling for trip characteristics and demographics. The main objectives of this paper are to investigate the relationship

between different risk and uncertainty attitudes, captured by means of self-reported data along with activities on travel related domains such as depth, breadth of the information search, travel micromoments and total time spend on travel-related websites captured by means of online passive behavioral data.

One of the main motivations of this paper lays in the methodological and managerial contributions. Namely, demonstrating a full cycle of research, where the research design utilizes observable online behavioral data collected via passive tracking technology combined with subjects' unobservable characteristics collected via online questionnaire, would provide guidelines to managers in how to best process and exploit their tracking data to extract valuable insights. Up to date, there is a limited amount of literature exploring these constructs in relationship to one another and in a relationship with non-self-reported metrics of information sought. Moreover, the research has theoretical contributions as it combines literature streams that have been researched only using self-reported data. Self-reported is prone to biases as opposed to passive data, which is further discussed in the next chapter.

To improve the research quality, I combine both self-reported, recalled information with objective, passively measured data. Using passive metering technology, I can capture the full footprint of consumers across their devices. This data can help understand the time spent planning the travel, the amount of information search with great accuracy and detail. Although prone to biases, survey data is still important as it can capture unobservable information which is still important and valuable. Merging the two datasets can provide an even deeper layer of insights into consumers' travel behavior.

The study findings are relevant to all travel stakeholders such as businesses, travelers, and researchers. The increased availability of data from consumers and its right exploitation can have a huge impact on the design of travel related products, their personalization and cross-platform usage, in-line with the trends reviewed in the industry.

2 Literature review

In this chapter, I will review the main literature stream relevant for the hypnotized relationships in this paper. First, I will review online travel related activities and information search. Second, I will focus on risk and uncertainty attitudes and expand to risk and uncertainty attitudes related to travel and tourism. Third, I will briefly explore the big five inventory and its utilization in tourism research. In the final section of this chapter, I will derive the hypotheses of this research.

2.1 Travel related activities and information search

Tourism related information search and planning has been a widely researched topic. Internet search behavior in tourism context has been examined primarily from the perspective of demographic variables, motivation and prior knowledge about the destination (Jani, Jang, and Hwang 2014). In a paper exploring longitudinal data for 12 years, Xiang, Magnini, and Fesenmaier (2015) identified as a key trend number one that the internet penetration among consumers using the Internet for travel planning already reached the level of saturation. However, in a research exploring the online behavior from a generation perspective, H. Kim, Xiang, and Fesenmaier (2015) noted that there is no sufficient amount of research with regards to how the newly emerged segments of consumers behave online and use the internet for travel planning.

Xiang, Magnini, and Fesenmaier (2015) turn special attention on travel planning process as a specific type of information search that is an important component of the decision making in tourism. During the planning, consumers obtain information based on which they choose their destinations and form their expectations. There is a substantial amount of research that looks at travel planning from different perspectives, some aim to identify the characteristics of the travelers' demographic characteristics (e.g. D.-Y. Kim, Lehto, and Morrison (2007)), other investigate the way travelers are conducting purchases and navigate in the information stream (e.g. Jun, Vogt, and MacKay (2007)), and recently social media and its influence on travel have been in the focus of researchers as well (e.g. Xiang and Gretzel (2010)).

From a practitioner standpoint, the more consumers are being active online the more prerequisites this creates for the tourism stakeholders marketers to reach them during

their decision-making process. Therefore, identifying the travel planning process i.e. the decision-making journey is a critical step for the brands to intervene and influence the consumers in their direction (Chatfield 2014). Practitioners define that the decision-making process related to the tourism as an array of micro moments which often are not even conscious to the consumers. A micromoment is essentially a user session with the particular goal of obtaining information or committing to a purchase.

For the purpose of this research, it is important to define the main components of the travel planning. In a meta-analysis, Jun, Vogt, and MacKay (2007) point out that there is a consensus among researchers that travel planning cannot be simplified to a single goal-oriented rational action but it is rather viewed as a complex task involving multiple goals and decisions around the different goals and characteristics of the trip. The authors define a conceptual model for travel planning which has three main sequential interrelated components, pre-trip, during trip and post-trip. This research focuses on pre-trip phase. Pre-trip phase itself consists of information search and planning (decision making), furthermore, travel related purchases also occur at the end of this phase as well as during the trip itself.

Jun, Vogt, and MacKay (2007) define travel plan as a complex decision involving an assessment of multiple alternatives organized around the travel goals in mind. The planning process includes setting goals and considering alternatives in order to achieve that goals including an evaluation of different alternatives' outcomes. Planning is dependent on the all information search behavior, utilization of the obtained information, purchase behavior and activities including past experience. Pan and Fesenmaier (2006) define vacation planning over the internet as an interaction between the user and the "online space" related to destinations and tourism. The online space contains content provided by diverse sources and the technology that facilitates the communication. User's "situation, knowledge and skills" combined with the "online space" contribute to the effective search.

It is important to be noted that trip planning is an important and enjoyable part of the vacation experience itself (Stewart and Vogt 1999) and it is likely to be high costs and high involvement decision (Bonn, Furr, and Susskind 1998). Furthermore, travel information search behavior explains travel purchase behavior (Woodside and MacDonald 1994). Quintal, Lee, and Soutar (2010) review numerous aspect of from which information search has been researched including the amount of search, the

number of sources, the search process, involvement, socio-demographic differences, culture etc.

Pan and Fesenmaier (2006) review the consumer vacation planning process from micro level perspective. Their research is motivated by the fact that previous research has been mostly focused on exploring planning and information search on macro level i.e. motivation, need, determinants, and outcomes. Their research focuses on a “snapshot” of travel planning where subjects make choices regarding a hypothetical holiday trip to San Diego. Using such setting it was possible to observe different chapters containing many episodes on how consumers adapt in their online search and come up with a final decision. These episodes or micromoments defined above represent one of the main variables of interest of this research.

Additionally, I will use the metrics defined by Ho and Liu (2005) in “An exploratory investigation of web-based tourist information search behavior”. The researchers explored and characterized the online behavior of 96 subjects in laboratory research by assigning them one tourism related task and recording their desktop activity. Ho and Liu (2005) empirically analyzed the recorded videos and reported on different measures, representing the information search behavior.

Additionally to micromoments, in this paper, I will adapt measures utilized in Ho and Liu (2005). The full list of dependent variables representing the internet search behavior includes:

- Total number of micromoments on travel related websites
- Total number of unique travel related domains (breadth of search)
- Total number of pageviews on travel related domains (depth of search)
- Total time spend on travel related domains
- Total sum of all travel related micromoments (this measure has been introduced due to one particular limitation of the mobile measurement discussed in the limitations section)

Ho and Liu (2005) conducted the research in laboratory environment exploring a single goal-directed session of activity. However, the whole information search planning and purchasing usually happen on many occasions within a longer period. Planning is fragmented on many episodes each characterized with its own purpose reflecting a specific problem (Pan and Fesenmaier 2006). In this research, I will use real-life data

representing the actual subjects' behavior. In this sense, the data is highly accurate and gives a detailed representation of the internet search behavior.

In conclusion, information search behavior is an important phase of the of the overall tourism behavior and more specifically the travel planning. From tourism stakeholders' perspective, it is a crucial phase where the consumer can be influenced by effective communication strategies and communication systems.

2.2 Risk and uncertainty

Risk attitudes are a central part of the economic theory. Classical economic theory of decision under risk states that the risk is related to the probability of the occurrence of the specific outcome. For example, according to expected utility a prospect with probability P to win x amount of money opposed to $1-P$ to win nothing, will be evaluated as follows: $p \cdot u(x) + (1-p)u(0)$ where u is the utility function of money. Risk attitudes are defined as follows, risk aversion is an attitude which is manifested by the preference of the sure outcome over a prospect with higher expected value that involves risk. Whereas, risk seeking attitude will occur when the prospect is preferred over the sure amount.

Later economic theory evolves by distinguishing individual level probability weighing and utility weighting by taking into account different psychological variables. (Woodside and MacDonald 1994) propose "Prospect theory" to explain choices among risky prospects that are inconsistent with the standard economic theory. In these recent developments, the risk attitudes evolve. In order to explain decision making under risk scholars explore choices involving a different amount of risk (high risk and low risk) and associated with outcome involving different monetary values (again high and low) as well gains or loses.

In a paper focused on risk measurement in consumer research Mandrik and Bao (2005) summarize that the measurement of risk attitudes typically has been assessed in three ways. The first method involved "choice dilemmas" where subjects are presented with several scenarios and asked for their preference between two courses of action, this results in computing an overall score which is used to determine respondents' risk attitude. The second method involves gambles. Subjects are asked to choose an amount in order to participate in a gamble. Finally, researchers use self-reported measures. These include the creation of different scales that are measure risk and uncertainty in specific decision

situations. The authors validate a novel self-reported scale which measures general risk attitude as a valid psychometrical measure. The construct proposed by Mandrik and Bao (2005) has been utilized in this research as it provides shorter and simple manner of assessing risk attitudes.

2.3 Risk and uncertainty in tourism

As noted above, one of the reasons researchers claims to cause the extensive information search is risk and uncertainty minimization. Stewart and Vogt (1999) attribute uncertainty as an implicit and universal characteristic of every planning process. Furthermore, the authors argue that in order to handle uncertainty the travelers prepare more than one plan for their trips. Sweeney, Soutar, and Johnson (1999) point out that consumers who are more sensitive to risk and uncertainty engage in more extensive search in order to avoid them. Sirakaya and Woodside (2005) claim that because of the intangible nature of tourism products the uncertainty in tourism is higher than comparing to other products or services.

An important remark related to risk and uncertainty is the difference between both constructs. The difference between them lays in the probabilities of their outcomes, while risk is associated outcomes with known probabilities, uncertainty is associated with outcomes with unknown probabilities. Quintal, Lee, and Soutar (2010) investigate the difference between risk and uncertainty on a country level using Hofstede (1980) uncertainty avoidance index (UAI) and risk scale measurement on tourists' information search. The researchers claim that many other papers do not make the distinction and this is especially problematic when researchers are using country UAI scores to explain individual-level behavior as individuals differ with regards to their attitudes of risk and uncertainty.

Quintal, Lee, and Soutar (2010) explain the relationship between uncertainty and risk in tourism and information search to be translated in the following way. In the early stages of their research, consumers search for information extensively and the outcomes are associated with uncertainty because the rate of occurrences of certain threatening events or undesirable outcomes is not known. In a later stage of the decision-making process, when travelers have already selected possible alternatives, the risk attitude is more likely to have an influence as consumers can assign relative probabilities to a few

selected alternatives i.e. alternatives are being compared to one another providing a reference point.

Based on the relationship explained above, Quintal, Lee, and Soutar (2010) provide a hypothesis that the uncertainty attitude has an influence on the extent of the information search while holding risk attitude constant. With regards to the risk attitude, Quintal, Lee, and Soutar (2010) provide a hypothesis that risk attitude doesn't influence the extent of the information search, holding uncertainty attitude constant. This hypothesis, however, has been defended under the notion that all the information search has been performed during the early stage of the trip planning and in the later stage consumers only compare alternatives based on the information they gathered in the early stage. However, in a conceptual model of trip planning presented by Jun, Vogt, and MacKay (2007), the information search is a component of all phases of the trip planning, including pre-trip, during trip and post-trip. Therefore, I find it credible to hypothesize that both risk and uncertainty attitudes influence the extent of the information search.

I further extend my hypotheses, regarding the direction of the effect. My hypotheses are based on Urbany, Dickson, and Wilkie (1989) and Gemünden (1985) claiming that information search is an instrument consumers use to minimize respectively the uncertainties and risks anticipated with regards to future purchases. For example, Gemünden (1985) finds that when the complexity of the decision making is increasing, so does the information search. They attribute this to higher risks involved in that decisions. As travel planning is regarded as a complex task, I can expect the same relationship. Urbany, Dickson, and Wilkie (1989) outlines two types of uncertainty, choice and knowledge uncertainty, choice uncertainty is related to uncertainties choosing a particular alternative amongst many, while knowledge uncertainty is related to familiarity with the product features. Their findings suggest that choice uncertainty has a positive relationship with the extent of the information search.

Based on the literature review above about information search, travel planning and risk and uncertainty hereby I form the following hypotheses:

- H1. Risk seeking attitude decreases the extent of information search, keeping other factors constant.
- H2. Uncertainty seeking attitude decreases the extent of information search, keeping other factors constant.

2.4 Big Five Factors Inventory (BFI)

According to Leung and Law (2010), the usage of personality traits in travel related literature appears to be low even though the academics agreed upon their value in Marketing domain (Baumgartner 2002). Personality is a temperament or person's inherent qualities of mind and strategies according to which one behaves, dispositions and behavioral patterns that are stable across time and can be used to characterize one's behavior. The trait perspective has been frequently utilized in consumer research because their ease of application as a self-reported measure and the measurement outputs can be easily applied in the statistical analysis. (Jani, Jang, and Hwang 2014).

Big five inventory (BFI) is the most well-known and established factor structure to measure personality (Denissen et al. 2008). It has been validated on many occasions and widely utilized by researchers, creating prerequisites to compare findings across different studies. Another important factor for the popularity of BFI is the fact they are freely available to use along with their validated translations in many languages. BFI has been proposed as a fundamental lexical hypothesis by Galton in 1884 (Goldberg 1993), which is a language taxonomy of human temperament based on adjectives describing different personality traits. The theory was put into practice by Goldberg (1993) and it has been greatly developed ever since, leading to the construction of five broad factors. BFI are based on factor analysis where a large group of traits is shown to be correlated and grouped into five universal traits.

BFI consists of Extraversion, Neuroticism, Conscientiousness, Agreeableness and Open to experience. Openness to experience is related to the degree of curiosity, inventiveness, adoption of novelty on the one hand and consistency and cautiousness on the other. That is, persons with high openness tend to be open-minded, adventurous while low openness can describe individuals that are more pragmatic. Conscientiousness reflects on the tendency for one to be organized, non-spontaneous, organized and efficient. Extroversion is related to traits such as outgoing personality, sociability, talkativeness. Personalities exhibiting low extroversion, on the other hand, can be perceived as less open and reserved. Agreeableness is described as a person being more compassionate and cooperative. It measures whether a person can be trusted or not and if they are well-tempered. High agreeableness personalities are seen as more naïve, while low are seen as more dominative and competitive. Finally, neuroticism explores the emotional stability of individuals. That is a high need for stability results in individuals who

are calm and stable, while a low need for stability can describe emotionally unstable individuals.

These five factors are shown to be the main factors that drive human behavior, appear in different cultures, are relatively stable across the lifetime of subjects and have strong predictive validity (Jani, Jang, and Hwang 2014). Researchers using BFI often aim to measure big five personality dimensions using as less as possible items which lead to the development and validation of short scales of BFI [(Denissen et al. 2008).

2.5 Personality traits and tourism information search

In tourism domain, BFI has been examined with regards to general travel behavior, the undertaken activities in the travel destination, adventure tourism and pilgrim tourism (Jani, Jang, and Hwang 2014). There are two articles researching internet search behavior from the perspective of BFI. Jani (2011) proposes a model relating information needs and tourist information behavior from the perspective of BFI and travel personality. It defines tourists' information needs as reasons for collecting information. Information search behavior is defined as breadth and depth of information sources that consumers use to obtain information. Furthermore, Jani, Jang, and Hwang (2014), research the relationship between BFI and internet search behavior. Using self-reported survey data, they observe that travel related information sought online varies with regards to the different personalities. The authors conclude that some of the factors from the BFI can improve the information search behavior predictability.

Jani, Jang, and Hwang (2014) address the research question whether BFI can be used as a predictor of internet search behavior in terms of sources of information and the extent of the information sought. Using self-reported measures for internet information search and channels used, the authors confirm that personality traits can be used as a predictor of information search behavior.

As BFI in tourism has not been researched thoroughly, in this research I will commit to hypotheses concerning the relationship between personality traits and information search behavior. However, the BFI will be used in the research design and analysis of this paper along with trip characteristics as control variables.

2.6 Hypotheses

Based on the literature review, here I define the full list of hypotheses relating each of them to the main variables of interest in regards to travel related activities as listed above. The following relationships between observed behavior and unobservable personality traits are researched:

Hypotheses				
H1	Risk attitude	<i>Decreases</i>	the amount of	a. Micromoments b. Domains c. Pageviews d. Time e. Lenght
H2	Uncertainty attitude	<i>Decreases</i>	the amount of	a. Micromoments b. Domains c. Pageviews d. Time e. Length

As mentioned above I will run an exploratory analysis of the impact of the big five personality traits on online travel behavior. However, I will not commit to a hypothesis with regards to their effect or direction.

3 Methodology

3.1 Introduction

The data utilized for this research has been collected from panelists who participate in a large consumer panel in the Netherlands. The data has two main components, behavioral data, and survey data. The methodological section of this paper explains the sampling, data collection and analysis procedures that have been conducted, in the following way.

- Panel and sampling. Sampling based on behavioral data. A sample of respondents has been pulled out of the behavioral dataset based on their online behavior on popular tourism related websites in the Netherlands.
- Survey. An online survey has been administered among the sampled panelists to reveal the “unobservables” from the perspective of the available behavioral data. Namely, these include data points related to trip characteristics associated with respondents’ latest tourist related purchases as well as the question regarding their personality. The code related to this step is available in *Appendix #2 Code. CHUNK #5*, *Appendix #2 Code. CHUNK #6*
- Data processing. The full behavioral data of the respondents who successfully completed the survey has been sampled out of the full behavioral panel dataset. The code related to this step is available in *Appendix #2 Code. CHUNK #1*, *Appendix #2 Code. CHUNK #4*
- Categorization. All of the unique websites of the sampled behavioral dataset have been classified into categories. In such a way it was possible to assess whether a certain website was travel related or not. The code related to this step is available in *Appendix #2 Code. CHUNK #2A* and *Appendix #2 Code. CHUNK #2B*, *Appendix #2 Code. CHUNK #3A*, *Appendix #2 Code. CHUNK #3B*
- Analysis. The aggregated information from the behavioral data has been regressed over the survey data to reveal the impact the unobservable personality attitudes and traits on the tourism related online behavior while controlling for the trip characteristics. The code related to this step is available in *Appendix #2 Code. CHUNK #8*

In the following sections, I will first focus on explaining the behavioral data and the

technology behind it. Afterward, the 5 major steps of the methodology are explained in details. Then I proceed with the descriptive results of each of the datasets (survey and behavioral). Finally, I report the analysis results.

3.2 Behavioral data & Technology

The behavioral data, also referred to as observational data, reflects the online behavior of the consumers. It consists of records of the interaction made via consumers' digital devices and the Internet.

The behavioral data has been collected via a technology developed and provided for this research by Wakoopa. The company is a provider of a tracking technology. The technology is utilized primarily for market research purposes. Similarly to the market research consumer panels, where panelists enroll to participate in online surveys for incentives, Wakoopa provides its technology to market research consumer panel companies that are interested in tracking the online behavior of their panelists. After enrolling into the panel and giving their consent to be tracked, panelists install an application on their devices i.e. desktop, mobile, and tablet. The tracking software collects every interaction of the panelists' devices on the Internet which consists of a path or the address the panelists are reaching and the duration of the visit. The software works in a different manner over the different operation systems platforms and devices, but the final result consists of recording raw data containing events. Each event has the address the participant accessed, the duration of the interaction and the client requesting the information i.e. browser, app etc.

3.3 Steps

3.3.1 Panel and sampling

Using Wakoopa's panel it was possible to 'pre-screen' relevant respondents for the purposes of the analysis, that could be invited to participate in the survey I have conducted. The respondents that I was looking for, should have been active on tourism related websites and should have conducted a purchase on such websites in the period of January 2015 to June 2016. Herein, I describe the process that was used to reach such

subjects. Upon starting of the project the panel used for the research had 6682 active panelists and 7103 active devices. The majority of the panelists were being active only on the desktop. See *Appendix #1. Panelists* and *Appendix #2. Devices* for reference.

First, I started looking into tourism related websites in the Netherlands. Initially, I used 300 domains to account for the majority of the tourism related internet traffic in the Netherlands according to the internet analytics company SimilarWeb. Afterward, based on this data I exported the activity of the whole panel over those websites ranging from 01/2015 to 06/2016. The data of this activity was manually analyzed looking for the end pages of the payments, also referred to as “confirmation” pages or “thank you” pages. Confirmation pages are the pages where a customer has been redirected after conducting a purchase at a company website. In general, the analysis consists of looking up keywords within the travel domains URLs and marking down the common patterns with the aid of regular expressions (See Mitkov (2005) for discussion). Identified were the patterns of “confirmation” pages and data for the participants visited from the period of 01/03/2016 to 16/06/2016 was exported. Using this information, it was possible to identify a sample composed of 949 respondents which were to be invited into the online survey as described in the second step below.

Given the estimation of the incidence rate provided by the panel supplier, 20%, and the initially desired sample of 500 respondents more respondents were needed. Therefore, an additional random sample of 123 respondents was selected based on whether they were active on the 300 initial domains but without evidence for their purchases from the data.

3.3.2 Survey

The selected 1039 panelists were invited to participate in an online survey, that aimed to reveal their attitudinal and personal characteristics as well as the trip characteristics of their last travel. The fieldwork was conducted during the last week of June 2016. Out of the 1039 invitations sent 872 started the survey, which resulted in 495 completed interviews. (Note to self, move this to a separate cleaning section). The data was further cleaned by accounting for speeders and flatliners. Speeders were panelists that have finished the survey within less than half of the average length of the interview. Flatliners refer to respondents who have answered all of the grid questions in a straight

line. Moreover, a few respondents attempted to participate in the survey multiple times and thus, they were excluded from further analysis. Thus, the final dataset consists of 426 observations.

The survey consisted of several parts. First, a screening criterion was used. It accounted for the number of travel-related purchases since March 2016. Respondents with no travel related purchases were not allowed to further proceed. Then, subjects who have not purchased neither flight nor accommodation were also excluded from the questionnaire. Business only travelers were not relevant for the analysis, thus they were also screened out, leaving only panelists who have gone on a leisure trip or both on a leisure and business trip. The second part was a demographics section, which consisted of question regarding age, gender, and income. The next part of the questionnaire was in regards to the trip characteristics, which was adapted from Roehl and Fesenmaier (1992) and contained information about moment of booking, destination, planning horizon, information sources used i.e. Internet, advice from friend and relatives, tourist information office, travel agent etc., products purchased online, duration of the trip or number of nights spent away from home, whether the destination was visited before and how many times, number of travel companions, indication whether there was children on the trip and whether subjects visited friends or relatives during their trip.

The next two sections aimed to reveal more about the subjects personality and risk and uncertainty attitude. Items assessing risk and uncertainty originate from Quintal, Lee, and Soutar (2010). Respondents reported their risk attitude on three item scale which the authors adapt from Donthu and Gilliland (1996). Uncertainty attitude has been assessed on four items scale which authors adapt from Yoo and Donthu (2002) and the scale is based on Hofstede (1980) UA items. I have chosen these scales as they are the shortest reliable scales for self-assessing the risk and uncertainty attitudes. Using such scales I can make sure to not tire respondents and keep them engaged. Personality traits related to openness, consciousness, extraversion, agreeableness and neuroticism have been assessed using the short scale from the big five inventory proposed by S. D. Gosling, Rentfrow, and Swann (2003).

3.3.3 Data processing

After having collected and cleaned all the data, I further proceed with data processing and analysis. All of the data processing and the analyzation tasks of this paper were done in R Studio. All of the code, including libraries can be seen in the appendix.

The data processing includes merging down the two main behavioral streams of data that originate from desktop and mobile. Additional to that it is needed to classify and derive all the tourism related domains that will be further used for analysis. This is an important step as it will allow to analysis what the relationship between the online travel behavior and the traveler’s attitudinal characteristics are. Categorization of the domains was done using machine learning classification algorithms provided by uClassify.com. The domains from the desktop and mobile behavioral data and all of the mobile apps are further classified using keywords.

Next step is, converting the raw data set onto micromoments. They are also further used in the analysis as I examine the relationship between one’s micromoment frequency and subject’s risk and uncertainty attitude. Micromoments are essentially user sessions where users were active at a given moment of time. For example, if a panelist access a certain domain and spend five minutes on it and then become inactive for more than 3 minutes, this will result in a factor variable grouping all of the observation within those five minutes of activity. If a travel related website has been visited during the micromoment, a dummy variable is assigned to this moment to indicate this. Furthermore, purchases of travel related products were included in the data based on the initial dataset used for pooling the sample out the panel for the survey. The final dataset includes aggregated information about panelists’ activity over two main levels. 1. top-level i.e. total activity 2. low-level micromoment level.

This section includes the data processing tasks done on the behavioral data including descriptive statistics and variable derivation, classification and aggregation on the different levels intended for further analysis.

Starting point of the analysis of the behavioral data include proccession the data and rendering it in a format suitable for running the analysis. In its initial form the data has been exported in a format containing the following variables: for the desktop data (sample data, see: *Appendix #3 Desktop data*) and for the mobile data (sample data, see: *Appendix #4 Mobile data*). The next step is adding “*host*” variable to both datasets.

The host variable contains domains and the subdomains that are going to be used for the categorization whether the domains were travel related or not. The next query on the data included identifying all of the micromoments in the data. First, the data has been subsetted on “*panelist_id*” level, sorted by the the timestamp “*used_at*” and assigned into a list where each element is the full data for each individual respondent. The sum of “*used_at*” variables and “*active_seconds*”, the duration in seconds respondent spend on the page, were compared to “*used_at*” variable of the next observation. If differences larger than five minutes were found all of the variables prior to this difference were grouped together under a common factor variable.

Table 3.1: Desktop dataset

<i>used_at</i>	<i>host</i>	<i>panelist_id</i>	<i>url</i>	<i>active_seconds</i>	<i>browser_name</i>	<i>mmid</i>	<i>Class_Travel</i>	<i>purchase</i>
----------------	-------------	--------------------	------------	-----------------------	---------------------	-------------	---------------------	-----------------

Table 3.2: Mobile dataset

<i>panelist_id</i>	<i>device_id</i>	<i>scheme</i>	<i>url</i>	<i>domain</i>	<i>app_id</i>	<i>app_name</i>	<i>used_at</i>	<i>connection</i>	<i>duration</i>	<i>host</i>
--------------------	------------------	---------------	------------	---------------	---------------	-----------------	----------------	-------------------	-----------------	-------------

3.3.4 Categorization

The categorization procedure includes using uClassify machine learning algorithms and also using keywords. The full activity coming from desktop and mobile devices resulted in 194,534 unique domains. The categorization algorithm has been responsible for classifying all domains that have more than ten visits or 47,818 domains in total works as follows. A web-scraper designed for these projects accesses a collection of domains and collects all of the information on the page, then removes the HTML elements along with the punctuation and renders down the information only to a part that is visible to the website visitors. Then, it passes this information to an application programmable interface (API) that returns the probability of this text being into sixteen different categories including travel.

The full list of categories includes:

NULL

Websites with content with the highest probability to be travel were assigned a value of a dummy variable 1 or 0 otherwise.

Due to technical and time constraints the API service wasn't able to classify all of the domains. Therefore, the full dataset of unique domain names has also been scanned for keywords. Keywords include travel, tourism, accommodation, hotels, flight etc. (full list of the keywords can be found in the code in the appendix). If any of the keywords appear within the domain name, the respective domain has also been assigned to the list of travel domains. The same keyword approach has been used for classification of the application on mobile devices, where the apps have been classified using their names.

3.3.5 Analysis

This paper utilizes regression analysis, factor analysis, and stepwise regression. The functional form has been selected based on regression analysis. The underlying factors regard to risk and uncertainty attitudes and BGI have been analyzed by exploratory factor analysis. Finally, the main model has been compiled using stepwise regression. The analysis techniques utilized by this research please refer to the appendix. The "Analysis" chapter contains brief elaboration on regression model "OLS", factor analysis and all of the tests concerning the underlying assumptions of the "OLS" assuring the best model performance.

3.3.6 Model selection

3.3.6.1 Restricted Model

There are two distinct sources of data resulting in three data sets to be investigated.

1. Dataset from desktop
2. Dataset from mobile
3. Combined dataset from desktop and mobile

Furthermore, there are five dependent variables that are intended to be examined during this research. Namely,

1. Number of travel micro-moments
2. Number of unique travel domains visited
3. Number of travel pageviews

4. Total time in seconds spent on travel domains
5. Total length in seconds of travel micro-moments

The difference between the last two (4 and 5) is as follows. Total time in seconds (4.) has been measured once a panelist arrives at a certain website that has been classified as travel, whereas length (5.) is the total length (i.e. the difference between the start and end) of micro-moments amongst which a participant visited a travel related domain.

The restricted model for each of the three datasets takes the following form:

Table 3.3: Restricted model(s)

Model	Dependent variables	Independent variables	Control *	Control2
# 1	Number of travel micro-moments	Risk + Uncertainty + Interaction	Total micro-moments	Total purchases + Days active in the panel
# 2	Number of unique travel domains visited		Total domains	
# 3	Number of travel pageviews		Total pageviews	
# 4	Total time in seconds spent on travel domains		Total time	
# 5	Total length in seconds of travel micro-moments		Total micro-moments length	

* As there was only one purchase detected on mobile devices this term wasn't present in the mobile model

Running Ramsey-Reset test per each of the basic models showed evidence that the functional form of the models is not well specified, thus the models were rejected. Consequently, all of the dependent variables along with their corresponding controls

were transformed into logs accounting for additional five models per each dataset. The results from the consequent running of the Ramsey-Reset test were still not satisfying, as the test showed significant results indicating that the functional form of the model is not well specified. Consequently, the corresponding control variables, accounting for total activity and purchases, were transformed into binary variables with values 1-indicating a participant belongs to a highly active groups of participants (having values above the mean of the sample) and 0 vice versa. The latter transformation accounted for investigating ten more models per each dataset reaching the total number of thirty models to be reviewed. The functional form of the model has been selected based on the dataset containing desktop data as 1) this dataset is the most complete in terms of a number of observation and 2) the participants were selected based on their desktop behavior.

Among all tested models, based on the results from the Ramsey-Reset test, I select the following functional form of the model for Desktop, Mobile, and Combined datasets. Shapiro-Wilk test for normality also performs best for this functional form, yet the sample is large enough so we can relax the normality assumption.

Table 3.4: Restricted model(s) functional form

Model	Dependent variables	Independent		
		variables	Control *	Control2 **
# 1	log(#) of travel micro-moments	Risk + Uncertainty + Interaction	(D) Total micro- moments	(D) Total purchases + Days active in the panel
# 2	log(#) of unique travel domains visited		(D) Total domains	
# 3	log(#) of travel pageviews		(D) Total pageviews	
# 4	log(#) time in seconds spent on travel domains		(D) Total time	

Model	Dependent variables	Independent		
		variables	Control *	Control2 **
# 5	log(#) length in seconds of travel micro-moments		(D) Total micro- moments length	

* Note: as there was only one purchase detected on mobile devices this term wasn't present in the mobile model.

** Dummy variables take value 1 if the respondents' activity is above the average activity of the sample and 0 otherwise.

A table with the results from the restricted model along with the model performance metrics can be found in the appendix:

Table 4.1: Desktop data, restricted model *Table 4.2: Mobile data, restricted model* *Table 4.3: Combined data, restricted model* *Table 4.4: Restricted models tests*

Due to poor model performance is shown in *Table 4.4: Restricted models tests* row **11 to 15**, the model shown in *Table 4.3: Combined data, restricted model* has been rejected. See Appendix *Table 4.4: Restricted models tests* for reference.

3.3.6.2 Final Model

Based on the selected functional form of the model(s) I proceed with stepwise selection in order to select the final model across desktop and mobile. First, the models are presented then results of the OLS performance tests are shown in the appendix under *Table 7: Full model tests, Desktop* and *Table 9: Full model tests, Mobile*. The tests have shown evidence for heteroscedasticity. Therefore, in all of the tables of the results, the standard errors have been corrected using white standard errors. The procedure is explained in the appendix.

4 Results

4.1 Survey Data

The original data consisted of 495 observations. Before proceeding with analysis, the data was cleaned based on several considerations. Namely, I looked at whether 1) respondents were speeding throughout the survey, 2) subjects were flatliners, e.g. if they have answered the grid question with the same answer, and 3) if they had low activity less than 1000 observation in the behavioral data. 4) if they attempted to participate the survey more than once. The final dataset thus consists of 426 observations.

Looking at the demographic characteristics, we see that there is an equal split between males and females. Seeing the distribution among age, the population above 50 years old is overrepresented, accounting for 59% of the sample. In terms of income, the distribution is normal, most of the subject indicated yearly income between 33.000 €–66.000 euro.

All subjects were actively involved in the trip planning as the main decision maker or contributed to the decision, subject who didn't participate in the decision making were screened out. The business trip only travelers were screened out as well, those left in the sample are either leisure or leisure and travelers. Finally, we see that the majority of respondents indicated they have booked both flight and accommodation.

The next part of the survey explored the trip characteristics with regards to their general planning, online behavior, and destination. These data points were used as control variables for the analysis. On average it took respondents 9 weeks to plan the whole trip, for about 40% it took between 1 and 4 weeks to plan everything, while 15% indicated they needed more than 20 weeks of planning.

The majority (80%) visited countries within Europe, and 57% have not visited the country before. On average, people who have visited the country before went there for the 4th time. Typically together were traveling 2 people, the largest travel group consisted of 14 people and they have spent approximately 9 nights away from home. About 34% of the respondents indicated they have spent more than 10 days. Only 20% of the subjects have stayed with relatives.

The majority of subjects researched for the trip using the Internet, but many have also

asked for advice from friends or relatives or approached a travel agent. Moreover, the internet is widely used for booking accommodation and not so much for traveling or booking “entertainment” online.

4.1.1 Risk and uncertainty

For understanding risk and uncertainty attitudes, the seven-item scale was analyzed using exploratory factor analysis. According to the analysis results, I cannot accept the hypothesis that two factors are sufficient to explain the seven-item scale. However, exploring further, the results show that the factor loadings of the different variables have the highest loadings on the assumed two underlying factors i.e. risk and uncertainty. Furthermore, running factor analysis when assuming 3 factors results in separating item #4 as a distinct factor on its own with relatively low factor loading. According to Yong and Pearce (2013) choosing too few factors may lead to leaving out important variance out of the analysis. However, I accept the high factor loadings onto the two assumed factors as a sufficient reason to continue the analysis with two factors. Namely, items 1 to 3 accounting for underlying risk attitude and items 4 to 7 accounting for attitude towards uncertainty. Factor analysis results on risk and uncertainty items are available in the *Appendix 1: Factor analysis results*

4.1.2 BIG5

Running exploratory factor analysis on BIG5 items resulted in confirming the initially assumed personality traits. Furthermore, based on the results we can accept the hypothesis that 5 factors are sufficient to explain the variance of the 11 item scale. Factor analysis results on BIG5 items are available in the *Appendix 1: BIG5*

Next, with the aid of factor analysis respondents were grouped into 2x2 groups based on the mean score of the items per each factor: 1) risk “averse” and “seeking” and 2) uncertainty “averse” and “seeking”. With regards to “big five”, the respondents have also being allocated into two groups based on the mean score of items accounting per each factor i.e. “high” and “low”. The procedure has been adopted based on the approach from Jani, Jang, and Hwang (2014).

It is interesting to observe some of the descriptive statistics which can already provide

some insights on what the relationships between people’s risk and uncertainty attitude and their travel behavior based on the survey data. First of all, the majority of subjects are risk averse and uncertainty averse. The subject being risk seeking account for 30% of the population, while the uncertainty seeking are 40%.

Table 4.1: Risk attitude

averse	seeking
324	170

Table 4.2: Uncertainty attitude

averse	seeking
299	195

Interestingly enough, given the data, it is observed that risk averse and uncertainty avoidant subjects actually prefer to go to a destination they haven’t visited before, while risk and uncertainty seeking respondents have a slightly higher preference to go to a known destination. Further looking at the data, I find that on average risk averse people tend to spend more time planning the trip than risk seeking people (10 days vs 7 days). This effect is significant at 5% significance level. However, uncertainty averse and seeking people exhibit similar behavior (9 versus 10 days), the mean difference is not significant at 5% significance level, and thus there is no difference in the planning horizon for uncertainty averse and seeking travelers.

4.2 Main results

4.2.1 Desktop Data

Looking at a table *Full model: Desktop, robust*, I can say that most of the hypotheses of this paper have been confirmed. The table with the main results is available below under *Table 5.3: Full model: Desktop, robust* while the table with the full results is available in the Appendix #1 *Table 5: Full model: Desktop, robust*

First, looking at the effect of risk seeking attitude towards travel behavior, I see that risk seeking attitude decreases individuals’ activities related to travel behavior such as

travel domains ($\log(\text{TD})$), travel pageviews ($\log(\text{TDPV})$), travel time ($\log(\text{TT})$) and total travel-related micromoment length ($\log(\text{TL})$). For example, risk seeking individuals tend to use 23% fewer travel domains on average compared to risk averse individuals, *ceteris paribus*. Moreover, risk seeking individuals spend 46% less time and have 37% fewer pageviews on travel related domains on average compared to risk averse individuals, *ceteris paribus*. Finally, when it comes to the two variables accounting for micromoments, we see that risk seeking attitude doesn't have a significant effect over the total number of travel related micromoments, while when it comes to overall micromoment time (containing travel related activity) it does have a significant effect. On average, risk seeking attitude decreases the overall time of micromoments by 26% compared to risk averse attitude, *ceteris paribus*.

Second, looking at the effect of uncertainty seeking attitude towards travel behavior, I see that uncertainty seeking attitude decreases individuals' activities related to travel behavior such as travel domains ($\log(\text{TD})$), travel pageviews ($\log(\text{TDPV})$) and travel time ($\log(\text{TT})$). Uncertainty seeking individuals tend to use 25% fewer travel domains on average compared to uncertainty averse individuals, *ceteris paribus*. Moreover, uncertainty seeking individuals spend 37% less time and have 34% fewer pageviews on travel related domains on average compared to uncertainty averse individuals, *ceteris paribus*. Finally, when it comes to the two variables accounting for micromoments, we see that uncertainty seeking attitude doesn't have a significant effect over both the total number of travel related micromoments and overall micromoment time (containing travel related activity).

Third, the interaction term between risk seeking and uncertainty seeking attitudes has a significant effect only on time spent on travel-related websites. Interestingly, being both risk seeking and uncertainty seeking increases the time spent on travel related domains by 76%, *ceteris paribus*.

Exploring the effect of personality traits on online travel behavior the following effects are revealed. Scoring low on personality trait "Openness to experience" decreases the number of pageviews on travel related websites by 32% and the total time spent on travel-related websites by 33%, *ceteris paribus*. Scoring low on personality trait "Consciousness" decreases the total time spent on travel-related websites by 20%, *ceteris paribus*.

Table 4.3: Full model: Desktop, robust

	<i>Dependent variable:</i>				
	D log(MM) (1)	D log(TD) (2)	D log(TDPV) (3)	D log(TT) (4)	D log(TL) (5)
Risk Seeking	-0.11 p = 0.36	-0.23* p = 0.08	-0.37*** p = 0.005	-0.46*** p = 0.0002	-0.26** p = 0.03
Uncertainty Seeking	-0.13 p = 0.26	-0.25** p = 0.04	-0.34*** p = 0.004	-0.37*** p = 0.002	-0.17 p = 0.12
Risk(x)Uncertainty: Seeking	0.09 p = 0.73	0.10 p = 0.69	0.39 p = 0.16	0.76** p = 0.02	0.30 p = 0.22
Days active	0.01*** p = 0.001	0.01*** p = 0.002	0.01** p = 0.02	0.01** p = 0.04	0.005*** p = 0.01
D Purchase	0.53*** p = 0.0000	0.63*** p = 0.0000	0.84*** p = 0.00	0.88*** p = 0.00	0.50*** p = 0.0000
D Micromoments	1.13*** p = 0.00				
D Domains		1.03*** p = 0.00			
D PageViews			0.83*** p = 0.00		
D Time				0.80*** p = 0.00	
D MM Length					1.36*** p = 0.00

Note: * p<0.1; ** p<0.05; *** p<0.01

4.2.2 Mobile data

Looking at the table *Full model: Mobile, robust*, I see a big difference in respondents behavior compared to the desktop behavior. Most of the hypotheses with regards to the mobile behavior of this paper have not been confirmed. The table with the main results is available below under *Table 5.4: Full model: Mobile, robust* while the table with the full results is available in the Appendix #1 *Table 6: Full model: Mobile, robust*

First, looking at the effect of risk seeking attitude I do not observe any significant relationship with the dependent variables. Second, looking at the effect of uncertainty seeking attitude towards travel behavior reveals some surprising results. Uncertainty seeking attitude increases individuals' activities related to travel behavior such as travel micromoments (log(MM)), travel domains (log(TD)), travel time (log(TT)) and overall micromoment time (log(TL)).

Third, on the other hand, the interaction term between risk seeking and uncertainty seeking attitudes decreases greatly individuals' activities related to travel behavior such as travel micromoments (log(MM)), travel domains (log(TD)), travel time (log(TT))

and overall micromoment time ($\log(TL)$).

Forth, scoring low on personality trait “Consciousness” increases the number of micromoments by 60% and travel domains by 53%, *ceteris paribus*. Scoring low on “Extraversion” increases the pageviews on travel related domains by 88%, *ceteris paribus*. Scoring low on “Agreeableness” significantly decreases travel micromoments ($\log(MM)$), travel domains ($\log(TD)$), travel pageviews ($\log(TP)$) and travel time ($\log(TT)$). Scoring low on “Neuroticism” increases travel domains ($\log(TD)$), travel pageviews ($\log(TP)$), travel time ($\log(TT)$) and overall micromoment time ($\log(TL)$).

Table 4.4: Full model: Mobile, robust

	<i>Dependent variable:</i>				
	D log(MM) (1)	D log(TD) (2)	D log(TDPV) (3)	D log(TT) (4)	D log(TL) (5)
Risk Seeking	0.16 p = 0.66	0.13 p = 0.61	0.53 p = 0.22	1.32 p = 0.17	1.09 p = 0.21
Uncertainty Seeking	0.72** p = 0.04	0.73** p = 0.02	0.57 p = 0.11	1.98* p = 0.06	2.05** p = 0.02
Risk(x)Uncertainty: Seeking	-1.76*** p = 0.003	-1.80*** p = 0.0000	-0.75 p = 0.34	-7.89*** p = 0.00	-7.37*** p = 0.0000
Days active	0.01*** p = 0.001	0.01*** p = 0.0001	0.004 p = 0.12	0.02*** p = 0.0000	0.02*** p = 0.0001
D Micromoments	1.03*** p = 0.004				
D Domains		1.77*** p = 0.0000			
D PageViews			2.44*** p = 0.0000		
D Time				2.06*** p = 0.01	
D MM Length					2.20*** p = 0.003
<i>Note:</i>					
*p<0.1; **p<0.05; ***p<0.01					

4.3 Conclusion

Based on the results I conclude that risk and uncertainty attitudes have an impact of online travel related behavior. I can confirm that the following hypothesized relationships exist.

Based on the analysis the following hypotheses are confirmed/rejected with regards to consumers' **Desktop** activity:

Hypotheses				
H1	Risk attitude	<i>Decreases</i> the amount of	a. Micromoments	<i>Rejected</i>
			b. Domains	Confirmed
			c. Pageviews	Confirmed
			d. Time	Confirmed
			e. Length	Confirmed
H2	Uncertainty attitude	<i>Decreases</i> the amount of	a. Micromoments	<i>Rejected</i>
			b. Domains	Confirmed
			c. Pageviews	Confirmed
			d. Time	Confirmed

Hypotheses
e. Length <i>Rejected</i>

Based on the analysis the following hypotheses are confirmed/rejected with regards to consumers' **Mobile** activity:

Hypotheses
H1 Risk attitude <i>Decreases</i> the amount of
a. Micromoments <i>Rejected</i>
b. Domains <i>Rejected</i>
c. Pageviews <i>Rejected</i>
d. Time <i>Rejected</i>
e. Length <i>Rejected</i>
H2 Uncertainty attitude <i>Decreases</i> the amount of
a. Micromoments Opposite effect
b. Domains Opposite effect
c. Pageviews <i>Rejected</i>
d. Time Opposite effect
e. Length Opposite effect

In the next chapter, the results are discussed thoroughly.

5 Conclusions

5.1 Discussion

The present study aimed at exploring the personality traits such as risk and uncertainty attitudes and their relationship with online travel behavior. The online travel behavior was viewed both from desktop and mobile perspective, in line with recommendations from Chatfield (2014) and Xiang, Magnini, and Fesenmaier (2015). It is also observed that mobile online behavior is one of the major trends happening in the online travel industry, therefore, nowadays it is essential to review the consumer journey on every channel. The results suggest that the same personality traits have a different impact on desktop and mobile online travel behavior.

The finding with regards to desktop online travel behavior is in line with the main hypotheses of the study. It has been found that risk seeking attitude (H1) decreases the aspect of online travel related behavior related to a number of travel related domains, pageviews on travel related websites, time spent on travel-related websites, overall length of travel micromoments, while the effect is not significant with regards to a number of travel micromoments. The uncertainty attitude has been found to influence a number of travel related domains, pageviews on travel related websites, time spent on travel related websites, while it does not have a significant relationship with micromoments and overall length of travel micromoments. These findings are in line with previous finds in the literature that information search is uncertainty reducing instrument (Urbany, Dickson, and Wilkie 1989). The findings confirm the relationship between uncertainty and the extend of information search studied by Quintal, Lee, and Soutar (2010). However, the analysis also reveals that there is a relationship between risk attitude and the extend of information search in travel context, which contradicts to Quintal, Lee, and Soutar (2010). This finding is in line with the literature that information search is risk reducing instrument (Gemünden 1985). The contradiction may come from the fact that Quintal, Lee, and Soutar (2010) uses self-reported data on different information sources used when making travel decisions. In a design, metrics on the extend of information searches such as time or pageviews cannot be accurately accessed.

When it comes to mobile behavior the findings are surprising. The risk attitude hasn't been found to have a relationship with the online travel related behavior (in line with

Quintal, Lee, and Soutar (2010)) while the uncertainty attitude has been found to have an opposite direction of the effect i.e. uncertainty seeking behavior increases the extend of online travel information search. Furthermore, the interaction term between risk and uncertainty seeking attitude has a huge impact (e.g. more than 700% of time). Such a big fluctuations and unusual results can be explained by the fact that the sample size on mobile is really small 109 respondents and there is no indication what kind of mobile device is the source of data e.g. smartphone or tablet which leaves a big part of the variance in the error term. Furthermore, there is a measurement error when measuring time on mobile (explained in the limitations). Nevertheless, this difference is interesting because of the rising role of mobile in the online travel industry and its worth further investigation.

With regards to big five personality traits, only a few effects found in previous research have been recognized. The desktop behavioral, suggest that the effect of scoring low on the factor “openness to experience” decreases the number of pageviews on travel related websites and the total time spent on them. These findings support the notion that “closed-mindedness” as opposite to “openness” is related to more reserved views towards new experiences. Being low in “consciousness” i.e. “carelessness” decreases the total time spent on travel related websites. Again, the intuition suggests that careless individuals would not spend more efforts obtaining information to decrease probabilities of undesired future outcomes compared to conscious ones.

The findings from the mobile behavior are different. Low in personality trait “consciousness” increases the number of micromoments and travel domains. Low in “extraversion” increases the pageviews on travel related domains. Low in “agreeableness” decreases travel micromoments, travel domains, travel pageviews and time. This contradicts the notions that “antagonism” opposed to “agreeableness” is related to more skeptical and analytical minded individuals which would be expected to spend more time obtaining new information. Low in “Neuroticism” increases travel domains, travel pageviews, travel time and overall micromoment time. Again, these findings are on the opposite side of the intuition which suggest that more “emotionally stable” or “confident” individuals opposed to “neurotic” would be less likely to spend time gathering information and researching alternatives online.

The results from the previous literature (Jani, Jang, and Hwang 2014) suggests that travel information search varies with the different factors except from “extraversion”,

and the online sources utilized for travel information search vary with the different factors except for “conscientiousness”. One of the most obvious reason for the difference in the research methodology, as Jani, Jang, and Hwang (2014) utilized survey data for reporting online activities which are shown to be not particularly accurate (Wakoopa 2016). However, as noted above, some of the findings are contra intuitive, therefore further research is needed in order to explore these relationships in details.

5.2 Contribution and Implications

Behavioral data gives us an amazing capacity to understand the online behavior of the consumers as never before. Although the level of understanding of the consumers reached by the academia and practitioners is still great before the age of behavioral data, the capability of zooming into details of the real consumer behavior and the robustness of the insights are truly fascinating. Practically behavioral data empowers researchers to answer every question concerning the observable online behavior that comes to mind in a way beyond the capacity even of the consumer himself. Due to the complexity and fragmentation, the consumer is unable to recall the information in such rich details as when it is recorded.

Due to this fact, the relationship under investigation has not been investigated before in such a detailed manner. This research confirms the finding from previous research in the area that the attitude towards uncertainty, but not the risk is responsible for choosing a number of sources used for travel decision making. Our findings contribute to previous research through passive metering by investigating the time people spend on travel websites, as well as the efforts in terms of pageviews. The results from the analysis strongly confirm the hypotheses that both risk and uncertainty attitudes account for changes in time and pageviews spend desktop devices.

This study contributes to 4 areas, it has been used as promotional material, methodologically to the area of behavioral research, to the stream of the literature of decision making under risk and uncertainty, to tourism research and makes managerial implications for better online travel related marketing activities.

The findings serve as a tangible example of behavioral data employed in practice. The work has been used as promotional marketing material to promote the business of Wakoopa. The company offers technology for behavioral data collection. This work

has been released as part of blogpost series: “How behavioral data sheds light on real consumers’ travel planning behavior”:

- **Part #1**
- **Part #2**
- **Part #3**

The main contribution of the research is adding methodologically to the emerging stream of research in the behavioral observational research. The research explores unobservable attitudes of one’s personality assessed via questionnaires in relation to a real worlds observational data. Even though it uses an established and well-known analysis techniques, the novelty in data collection techniques gives significant methodological contribution for practitioners and the academia.

The research contributes to the area of behavioral economics more precisely in decision under risk and uncertainty. It expands the area in relation to of online travel related behavior. It confirms the notion ones that attitude toward risk and uncertainty play a significant role in decision making concerning travel related information search and that individuals with plausible attitude toward both constructs spend less time in planning whereas those more concerned with the risk and uncertainty, in general, are more careful planners.

5.3 Limitations & Recommendations

1. Risk and uncertainty measurement. The risk and uncertainty attitudes have been measured using the shorted possible way for measurement assessing only the general attitude of these personality traits. Arguably there could be discrepancies between the general attitude and domain specific attitude. Therefore, the experiment can be reproduced using a more sophisticated measurement of risk and uncertainty attitudes such as gambles and lotteries as well as assessing perceived risks in travel related domains.
2. Classification. The classification used in the research has been used as it is provided, however, it is possible to employ better classification algorithms providing better fragmentation of the travel related behavior. Therefore, for future research, I suggest the use of better classification. Micro-moment, the length of a

micro-moment has been set to be 5 minutes without any reference or estimation whatsoever. There is a research that suggests how the best way cut-off point in order to measure user sessions can be estimated. Furthermore, we use the same cut-off point of 5 minutes across desktop and mobile device. As the activity of both type of devices largely differs it could be necessary to use different cut-off points per desktop and mobile. This would also mean that the comparison between desktop and mobile is harder as it is generally different.

3. Multiple users per device. It is known that some desktop devices are shared in the households, therefore we are recording data streams from different users and assessing these data as coming from a single user. Thus, the analysis of this data can be biased as we 1) cannot distinguish the different users and 2) cannot connect properly the survey data with the behavioral data.
4. Time is inferred. The measurement cut-off point is 2 seconds anything shorter than two seconds is not recorded. Crosstab browsing on desktop devices cannot be accessed. There is a measurement error on mobile devices which makes it possible to have events with time zero.
5. Selection bias. One should not overlook the possibility of selection bias as the people who agree to be tracked can be arguably different than the ones who are more privacy conscious and would not agree to share such information even for research purposes. Selection bias states that the selected sample who agrees to participate in the research may yield different characteristics than the sample of interest drawn from the general population.
6. Programming. The program written to process and analyze the data contains several thousand lines of code. Developing such a code in an absence of a quality assurance team and/or a proper quality assurance procedures is prompt to errors.
7. Survey measurement. There are two issues related to the survey methodology. First, the income scale was related to individuals' income, while the questions have been asked with regards to household income. This prevented me from using this variable in the research design. Second, the fourth item from the BFI has been wrongly included as it contains two items instead of one. Yet, they are both from the same traits, thus the impact of this error is potentially small.

References

- Ammu, Nrusimham, and Mohd Irfanuddin. 2013. "Big Data Challenges." *International Journal of Advanced Trends in Computer Science and Engineering* 2 (1). Citeseer: 613–15.
- Baumgartner, Hans. 2002. "Toward a Personology of the Consumer." *Journal of Consumer Research* 29 (2). The Oxford University Press: 286–92.
- Bonn, M. a., H. L. Furr, and a. M. Susskind. 1998. "Using the Internet as a Pleasure Travel Planning Tool: an Examination of the Sociodemographic and Behavioral Characteristics Among Internet Users and Nonusers." doi:10.1177/109634809802200307.
- Chatfield, T. 2014. "Google Travel Survey 2014." *Google Travel Research*. Google/Ipsos MediaCT. https://storage.googleapis.com/think/docs/2014-travelers-road-to-decision_research_studies.pdf.
- . 2016. "The Trouble of Big Data Its Called Recency Bias." *BBC*. BBC. <http://www.bbc.com/future/story/20160605-the-trouble-with-big-data-its-called-the-recency-bias>.
- Denissen, Jaap JA, Rinie Geenen, Marcel AG Van Aken, Samuel D Gosling, and Jeff Potter. 2008. "Development and Validation of a Dutch Translation of the Big Five Inventory (Bfi)." *Journal of Personality Assessment* 90 (2). Taylor & Francis: 152–57.
- Donthu, Naveen, and David Gilliland. 1996. "The Infomercial Shopper." *Journal of Advertising Research* 36 (2). World Advertising Research Center Ltd.: 69–77.
- Euromonitor. 2015. "Travel Industry and Online Travel Global Overview." *Euromonitor*. Euromonitor.
- Eurostat. 2016. "Statistics on Ict Use in Tourism." *Eurostat*. Eurostat. http://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics_on_ICT_use_in_tourism.
- Evans, Joel R, and Anil Mathur. 2005. "The Value of Online Surveys." *Internet Research* 15 (2). Emerald Group Publishing Limited: 195–219.
- Fan, Jianqing, Fang Han, and Han Liu. 2014. "Challenges of Big Data Analysis." *National Science Review* 1 (2). Oxford University Press: 293–314.
- Gemünden, Hans Georg. 1985. "Perceived Risk and Information Search. a Systematic Meta-Analysis of the Empirical Evidence." *International Journal of Research in*

Marketing 2 (2). Elsevier: 79–100.

Glass, Gene V. 1976. “Primary, Secondary, and Meta-Analysis of Research.” *Educational Researcher* 5 (10). JSTOR: 3–8.

Goldberg, Lewis R. 1993. “The Structure of Phenotypic Personality Traits.” *American Psychologist* 48 (1). American Psychological Association: 26.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann. 2003. “A very brief measure of the Big-Five personality domains.” *Journal of Research in Personality* 37 (6): 504–28. doi:10.1016/S0092-6566(03)00046-1.

Ho, Chaang-ian, and Yung-ping Liu. 2005. “An exploratory investigation of web-based tourist information search behavior.” *Asia Pacific Journal of Tourism Research* 10 (4): 351–60. doi:10.1080/10941660500363645.

Hofstede, Geert. 1980. “Motivation, Leadership, and Organization: Do American Theories Apply Abroad?” *Organizational Dynamics* 9 (1). Elsevier: 42–63.

Hox, Joop J, and Hennie R Boeije. 2005. “Data Collection, Primary Vs. Secondary.” *Encyclopedia of Social Measurement* 1: 593–99.

Jani, Dev. 2011. “The influence of personality on tourist information behaviour.” *E-Review of Tourism Research* 9 (3): 88–95.

Jani, Dev, Jun-Ho Jang, and Yeong-Hyeon Hwang. 2014. “Big Five Factors of Personality and Tourists’ Internet Search Behavior.” *Asia Pacific Journal of Tourism Research* 19 (5): 600–615. doi:10.1080/10941665.2013.773922.

Jun, S H, C A Vogt, and K J MacKay. 2007. “Relationships between Travel Information Search and Travel Product Purchase in Pretrip Contexts.” *Journal of Travel Research* 45 (April): 266–74. doi:10.1177/0047287506295945.

Kim, Dae-Young, Xinran Y Lehto, and Alastair M Morrison. 2007. “Gender Differences in Online Travel Information Search: Implications for Marketing Communications on the Internet.” *Tourism Management* 28 (2). Elsevier: 423–33.

Kim, Heejun, Zheng Xiang, and Daniel R Fesenmaier. 2015. “Use of The Internet for Trip Planning: A Generational Analysis.” *Journal of Travel & Tourism Marketing* 32 (3): 276–89. doi:10.1080/10548408.2014.896765.

Labrinidis, Alexandros, and Hosagrahar V Jagadish. 2012. “Challenges and Opportuni-

ties with Big Data.” *Proceedings of the VLDB Endowment* 5 (12). VLDB Endowment: 2032–3.

Lepp, Andrew, and Heather Gibson. 2003. “Tourist roles, perceived risk and international tourism.” *Annals of Tourism Research* 30 (3): 606–24. doi:10.1016/S0160-7383(03)00024-0.

Leung, Rosanna, and Rob Law. 2010. “A Review of Personality Research in the Tourism and Hospitality Context.” *Journal of Travel & Tourism Marketing* 27 (5). Taylor & Francis: 439–59.

Mandrik, Carter A, and Yeqing Bao. 2005. “Exploring the Concept and Measurement of General Risk Aversion.” *Advances in Consumer Research* 32 (32): 531–39.

Mitkov, Ruslan. 2005. *The Oxford Handbook of Computational Linguistics*. Oxford University Press.

Nunan, Daniel, and MariaLaura Di Domenico. 2013. “Market Research and the Ethics of Big Data.” *International Journal of Market Research* 55 (4). The Market Research Society: 2–13.

Pan, Bing, and Daniel R. Fesenmaier. 2006. “Online Information Search. Vacation Planning Process.” *Annals of Tourism Research* 33 (3): 809–32. doi:10.1016/j.annals.2006.03.006.

Quintal, Vanessa Ann, Julie Anne Lee, and Geoffrey N. Soutar. 2010. “Tourists’ information search: The differential impact of risk and uncertainty avoidance.” *International Journal of Tourism Research* 12 (4): 321–33. doi:10.1002/jtr.753.

Roehl, W. S., and D. R. Fesenmaier. 1992. “Risk perceptions and pleasure travel: an exploratory analysis.” *Journal of Travel Research* 30 (4): 17–26. doi:10.1177/004728759203000403.

Sirakaya, Ercan, and Arch G Woodside. 2005. “Building and Testing Theories of Decision Making by Travellers.” *Tourism Management* 26 (6). Elsevier: 815–32.

Stewart, Susan I, and Christine A Vogt. 1999. “A Case-Based Approach to Understanding Vacation Planning.” doi:10.1080/014904099273165.

Sweeney, Julian C, Geoffrey N Soutar, and Lester W Johnson. 1999. “The Role of Perceived Risk in the Quality-Value Relationship: A Study in a Retail Environment.”

Journal of Retailing 75 (1). Elsevier: 77–105.

Tole, Alexandru Adrian, and others. 2013. “Big Data Challenges.” *Database Syst J* 4 (3): 31–40.

Urbany, Joel E, Peter R Dickson, and William L Wilkie. 1989. “Buyer Uncertainty and Information Search.” *Journal of Consumer Research* 16 (2). The Oxford University Press: 208–15.

Wakoopa. 2016. “Mind the Gap.” *Wakoopa*. Wakoopa. <https://blog.wakoopa.com/mind-the-gap-a-case-study-about-reported-and-actual-behavior-on-mobile-devices-2/>.

Woodside, Arch G, and Roberta MacDonald. 1994. “General System Framework of Customer Choice Processes of Tourism Services.” *Spoilt for Choice*. Kulturverl: Thaur, Germany, 30–59.

Xiang, Zheng, and Ulrike Gretzel. 2010. “Role of Social Media in Online Travel Information Search.” *Tourism Management* 31 (2). Elsevier: 179–88.

Xiang, Zheng, Vincent P. Magnini, and Daniel R. Fesenmaier. 2015. “Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet.” *Journal of Retailing and Consumer Services* 22. Elsevier: 244–49. doi:10.1016/j.jretconser.2014.08.005.

Yong, An Gie, and Sean Pearce. 2013. “A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis.” *Tutorials in Quantitative Methods for Psychology* 9 (2): 79–94.

Yoo, Boonghee, and Naveen Donthu. 2002. “Testing Cross-Cultural Invariance of the Brand Equity Creation Process.” *Journal of Product & Brand Management* 11 (6). MCB UP Ltd: 380–98.