Introduction

Grigor Dimitrov

September 27, 2016

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 @fan2014challenges. However, "big data" has been underutilized due to computational and statistical challenges @ammu2013big @tole2013big @labrinidis2012challenges. Moreover, @nunan2013market 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 @evans2005value. 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 500millon dollars were spent on online surveys while this figure doubled in 2004 @evans2005value.

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 @hox2005data @glass1976primary. 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 of 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, afterwards 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 [@Xiang2015]. This domain has been chosen due to 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 [@eurostat2016] 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 millon trips in 2009

to 1.4 billon trips in 2019 [@euromonitor2015].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. @Xiang2015 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 user can exchange travel services and experiences and the mobile computing, smartphones in particular [@euromonitor2015].

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. @urbany1989buyer introduces body of research to support the link between uncertainty and information search. When it comes to perceived risks, in a meta-analysis @gemunden1985perceived 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, @Roehl1992 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. @Lepp2003 look at the general perceptions or 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, @Quintal2010 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 different impact on the information search in terms of 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 huge impact on the design of travel related products, their personalization and cross-platform usage, in-line with the trends reviewed in the industry.