Does Gender, Age and Usage Matter in Big Data's Perception Applied in Online Tourism?

Jean-Luc Pradel Mathurin Augustin[1] & Shu-Yi Liaw[2]

[1]Department of Tropical Agriculture and International Cooperation, National Pingtung University of Science and Technology, Pingtung City, Taiwan

[2] Management College, Computer Center, National Pingtung University of Science and Technology, Pingtung City, Taiwan

Correspondence: Shu-Yi Liaw, National Pingtung University of Science and Technology, No 1, Hsueh-Fu Road, Neipu Township, Pingtung City, 91201 Taiwan. E-mail:syliaw@mail.npust.edu.tw

Abstract

It is often asserted that big data can potentially help firms make data-oriented decisions. The rise of big data completely revolutionized the tourism industry, explaining partly why this research intends to explore the relationship between big data applied in online tourism and online tourists' behavior. This study will tackle the question in relation to big data's perception based on gender, age and usage's interaction, factors identified throughout the literature as important. It is found that both advantages and disadvantages of big data positively impact online consumer behavior. Although surprising, this result is justified in the fact that consumers always consider the trade-off between advantages and disadvantages which explains their willingness to still use online services. It is also found that usage is a determinant factor influencing big data's perception. One major take away is that disadvantages of big data do not necessarily translate into a negative behavior. Moreover, online tourism website designers can tailor their products in a way to appeal to light users of online tourism services, considering heavy users are likely to buy in no matter what.

Keywords: online tourism, big data perception, advantages and disadvantages of big data, online consumers' behavior

1. Introduction

Tourism is an industry that heavily relies on availability of information for its survival and growth. All actors of the industry need reliable information to reach decision-making stages. Consumers need information to learn more about destinations, compare prices in order to properly plan their trips; firms and businesses need information to understand consumers' behaviour, provide the right offers to the right consumers, establish a sound recommender system and make data-driven business decisions; governments as well as policy-makers need information in order to pass and implement appropriate policies in support of the industry to boost national economies. It is worth highlighting that many countries around the world rely on tourism to develop their economy. Progress in information and communication technologies over the last three decades saw the birth and rise of internet technologies, social media as well as mobile technologies. Impacts of these technologies on the tourism industry are well documented. With these technology improvements, increasing and enormous amounts of data are streamed from various sources and devices into contemporary businesses and organizations. The daily and continuous flow of information and data being streamed led to the appearance of big data and big data analytics.

Big data refers to data that is large, fast or complex and it's difficult or impossible to process using traditional methods. Accessing and storing large amounts of information for analytics has been around a long time, but the concept big data gained momentum in the early 2000s when Doug Laney articulated the now mainstream definition of big data as the three Vs: volume, velocity and variety. SAS later proposed their own version by adding variability and veracity to Laney's definition. It was reported the world would generate 1.8 zettabytes (1.8 x 1021 bytes) of data by 2011 and forecasted that this figure would increase to 35 zettabytes by the year 2020 (Le & Liaw, 2017). E-commerce later appeared and, to gain more insights on potential consumers, online businesses had to turn to big data to gather and analyze as much information as possible to better target

customers and achieve profit maximization from sales. E-commerce refers to traditional commercial activities conducted through the new medium: the internet (Liu, Lin, Lee & Deng, 2013). Online tourism falls in the category of e-commerce, and by definition, it refers to the buying, selling or consultation of tourism services on the internet. Export.gov, a platform helping US companies, in an overview of Taiwan's e-commerce market noted in 2018 Taiwan had the highest proportion of e-commerce shoppers in Asia (67%) and the highest average revenue per e-commerce user. It also noted an increasing number of young shoppers preferred online shopping due to its convenience. Tourism services are one of the top ten most-purchased products as Taiwan's e-commerce market continues to grow. The internet penetration in Taiwan (87.9% in 2018) might have been a helping factor to the continuous evolution of e-commerce in the country, but big data might just have been the engine behind it.

Notable technology advancements over the last decades have impacted how business is generally conducted and led to an increase in online commercial activities. The assumption that gender, age and frequency of usage are factors that influence perception of new technologies has always been put forward. In the literature however, there is no consensus on the influence of gender on new technologies' adoption and use considering the contradictory results of previous studies. Internet technologies were also often referred to as a young people's medium but that trend has been progressively changing (Sorce, Perotti & Wildrick, 2005). Van derKaay and Young (2012) discussed the similarities in perceived technology use and skills with those technologies between two groups of community college faculty while also noting the major difference being that younger ones used the technologies slightly more often than the older ones. In the age of big data and online tourism however, very little is mentioned in the literature regarding the perception of new technologies used to improve online customers' experience and convenience. Therefore, this study will attempt to bridge that gap by exploring the relationship between big data used in online tourism and self-reported online consumer behavior, and by determining the relative impact of gender, age and usage on big data's perception when applied in online tourism.

The present study will seek to provide answers to the following research questions: 1) what is the impact of big data applied in online tourism on online behavior? How does the effect of gender, age and usage rate of online tourism services impact the perception of big data (advantages and disadvantages) when applied in online tourism?

The present study is divided into five sections. The first section introduces the study. The second section provides a review of the literature. The third section presents the study's methodology. The fourth section presents and discusses the results while providing managerial implications. The last section is the conclusion of the study.

2. Big Data and Online Consumer Behavior

Whether businesses or firms are adjusting in order to capitalize on big data has been a major point of discussion; however, the big data era has arrived and technology is moving forward from this point on. The term big data, extensively discussed in recent literature, refers to extremely large data sets that are structured, semi-structured or unstructured. They are generated on a day to day basis as they inundate businesses and firms. However, the final use of the data will ultimately determine its importance. It is not enough to have the data, but analyzing it and taking insights from it for data-driven decision-making and strategic business moves is a key to business expansion, profit maximization, etc. When businesses are able to get insights from collected and analyzed data, it provides them with a market-based competitive advantage (Kuntz et al., 2017; Line, Dogru, El-Manstrly, Buoye, Malthouse & Kandampully, 2020).

The appearance of the internet and mobile technologies have brought countless opportunities to serve the convenience of consumers as well as the need for profit maximization at the firm level. New ways of reaching customers were created once e-commerce emerged. E-commerce or electronic commerce refers to the activities of buying and selling (traditional commerce) taking place through the new medium that is the internet. It is discussed that since the appearance of electronic commerce, more consumers acquire goods from online sources on a yearly basis, indicating the continuous growth of consumer-oriented electronic commerce (Van Slyke, Bélanger, Johnson, & Hightower, 2010). Big data, since its appearance, changes the way companies and firms make decisions and do business. Many e-commerce and e-tourism organizations rely on big data insights to make smarter data-oriented business decisions. Big data provides detailed insights on consumers' behaviors and preferences which could help businesses tailor their products and services to meet potential customers' needs and preferences (Ghandour, 2015). Being aware of the appropriate content to communicate to each group of customers will benefit companies as they will end up minimizing their costs and, eventually, increase their profit margin by delivering high value to consumers' (Ghandour, 2015). Due to the heavily consumer-oriented

electronic commerce, consumers are able to experience a set of benefits from the application of big data in any type of online commerce including online tourism summarized into the following: a sound recommender system, a convenient information search function, a dynamic pricing system and an improved customer service (Tang & Wu, 2015; Le & Liaw, 2017; Augustin & Liaw, 2020). Alongside these advantages, there have been specific issues that consumers' have been subjected to. Most notably, the privacy and security concerns are top of the list followed by and increasing phenomenon of shopping addiction arising from the convenience of online shopping and concerns with group influences as well as the known negative effects of group influences (Le & Liaw, 2017; Augustin & Liaw, 2020).

2.1 Online Consumer Behavior

Augustin and Liaw (2020) studied online consumer behavior using the hierarchy of effects model to which they presented an extension of. They established that online consumer behavior follows a three-layer pattern inclusive of Attention and Intention (AI), Desire and Action (DA) and Sharing and Social Awareness (SS). Their model is in line with previous studies based on the AIDA model (Ehrenberg, 2000; Lee, Lin, Liao & Yeh, 2013; Le & Liaw, 2017). In the literature, it is recommended that a scale should be reliable and internally consistent (Edwards & Bagozzi, 2000; Petter, Straub & Rai, 2007) as it was the case with the second order reflective model presented by Augustin and Liaw (2020). Based on their definition of online consumer behavior, the concept can be seen as the process a customer goes through as he is pushed to make a purchase decision either by a marketing agent, a sales agent, or an advertisement campaign that is designed to trigger a certain behavioral response (Augustin & Liaw, 2020). But ultimately, it means that online sellers of tourism products should be able to draw attention from customers, raise their interest, have them desire the products and services on display and lead them to the purchase decision. At the same time, by offering a good service quality, they will expect the consumer to share their experiences with their social connections. That phase is critical in attracting new customers as well as inducing customers repurchase intention.

2.2 Impacts of Big Data Advantages on Consumer Behavior

As noted by Buhalis (1998), information technologies are a major contributor to competitiveness and competitive advantage, and they systematically changed the way businesses and companies compete. The use of big data technologies in particular offers ways to diversify products and/or services while also enabling a tighter targetting of customers. Big data technologies also offer new ways to reach customers while suggesting products addressing their specific needs. As the online tourist browses the internet, he conveniently finds a large amount of well-filtered information answering precisely his queries. As a result, he will not be required to filter and sort the information of relevance to him. Added to a convenient information search system, big data also provides a recommender system used by companies and firms to customize offerings to prospective customers. Moreover, big data allows companies to apply a dynamic price discrimination strategy where prices are set to match a given customer's willingness to pay. Added to the previous advantages, big data also helps companies provide high-quality customer service targeting consumers' satisfaction in a timely manner.

H₁: Advantages of big data are positively associated to online consumers' behavior

2.3 Big Data Disadvantages Effects on Consumers' Behavior

Big data allows companies and firms to reach out to consumers while they are online. But online companies want customers to remain on the internet for as long as possible in order to recommend products and services that could draw their attention and raise their interest. Once a customer spends excessive amount of time online, this behavior arguably becomes addictive, and the negative impacts of addiction are well documented in the literature. Moreover, being on the internet and, ultimately, purchasing products or services from e-vendors might put a customer at risk of seeing his personal information stolen from him due to known security issues coming along with internet's use. On top of that, companies collect an overwhelming amonnt of information from customers (knowingly or not) that one could argue is a violation of privacy. Considering these issues, whe propose the following hypothesis:

H₂: Disadvantages of big data are negatively associated to online consumers' behavior

2.4 The Effects of Gender, Age and Usage

Latest technology developments are often directed toward users' convenience. E-commerce does not fall short of that expectation as it brings the shopping experience to the customers no matter where they are. Mobile technologies as well as social media platforms are major contributors to the shopping experience being brought at the customers' fingertips. Everything is just a few clicks away nowadays given that with a cell phone or a portable computer, one can easily buy almost anything online.

In the literature, it is discussed that men and women behave differently when it comes to the adoption and use of internet technologies. As an example, it is argued that men and women differ in their views of online shopping (Van Slyke, Comunale & Belanger, 2002) and they also behave differently as consumers (Mitchell & Walsh, 2004). Previous studies have concluded that gender is a very important and determinant factor when it comes to consumers' behavior (Vijaya Lakshmi, Aparanjini Niharika & Lahari, 2017; Van Slyke, Bélanger, Johnson & Hightower, 2010). It is also discussed that men are more influenced by the utility than women. We know that perceived utility is closely linked to perceived advantage; therefore, one can expect perceived advantage to be stronger for men (Morris, Venkatesh & Ackerman, 2005). Moreover, there is evidence from previous studies that men's shopping experience is straight forward and time-saving oriented (Mitchell & Walsh, 2004). However, throughout the literature, studies that addressed gender differences contrast in their results and there is not an overwhelming amount of studies going in one unique direction with the influence of gender.

One common belief is that older adults are barely technology users. But this is discussed as a misconception and recent studies asserted that younger and older adults use the same technologies, but the major difference in their patterns is the usage rate compared to their younger counterparts (Olson, O'Brien, Rogers & Charness, 2011). Another study investigated age-related differences in technology usage among community college faculty and noted similarities in technologies used and degree of perceived skills with those technologies among the two groups (Van derKaay & Young, 2012). That study concluded overall technology used by older faculty was slightly less than younger faculty, but older ones were no less likely to use technology than younger ones.

Based on previously stated results found in the literature, we proposed the following hypothesis as well as the research model in Figure 1 where gender, age and usage are used as control variables and their differences will be assessed in order to analyze their effect on big data applied in online tourism:

H_{3-1}: the interaction effect between gender, age and usage influences the perception of big data advantages used in online tourism

H_{3-2}: the interaction effect between gender, age and usage influences the perception of big data disadvantages used in online tourism
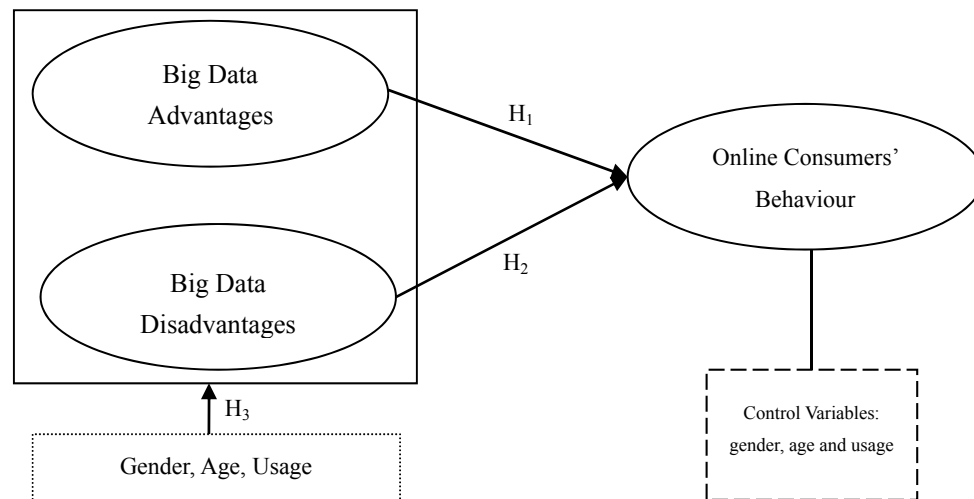

Figure 1. Research model

3. Methodology

The study will first and foremost assess the relationship between big data applied in online tourism (advantages and disadvantages) and online consumer behavior. Through assessment of that relationship, hypotheses 1 and 2 will be tested. This step will be followed by a three-way ANOVA assessing the combined effect of gender, age and usage of online tourism services on big data's perception when applied in online tourism, testing hypothesis 3-1 and 3-2.

3.1 Sample Size

It is suggested to aim at the desirable goal of 20:1 as a ratio between number of cases and number of free parameters; but a 10:1 ration seems a more realistic target (Kline, 2015). Although such a suggestion was made,

there are no absolute standards in the literature about sample size and path model complexity (Le & Liaw, 2017). Following the aforementioned recommendation, this study used more than 200 respondents while targeting young online tourists from Taiwan. The questionnaires were posted online and filled by the respondents between August and October 2018. After all statistical univariate outliers were removed, 226 questionnaires were deemed suitable to be used for the analyses. The demographics are presented in Table 1. However, they were not investigated any further due to their lack of significance to the study.

Table 1. Demographic characteristics of the sample

Variables		Frequency	Percentage (%)
Gender	Male	63	27.9
	Female	163	72.1
Age	Less than 20	101	44.7
	20 to 35	125	55.3
Usage	Low	76	33.6
	High	150	66.4

3.2 Data Collection and Measurement

Based on relevant literature, the questionnaire used in the study was designed to reflect the opinions of Taiwanese online tourists and their self-reported perceptions regarding the use of big data technologies in online tourism. An online tourist is an individual using the internet for touristic purposes such as gathering travel-related information, exploring new locations, buying or comparing tourist products, looking for all kinds of tourism related information. Reliability and validity of the constructs were tested through a pilot test on a sample of 66 online tourists. The questionnaire was posted on social media platforms where respondents had the opportunity to key in their answers. It is worth noting the questionnaire had two main sections, the first one measuring respondents' perception of each construct and the second one collected demographic characteristics such as gender, age, etc. A seven-point Likert scale was used, ranging between 1 (strongly disagree) and 7 (strongly agree).

Table 2. Constructs and Items of the questionnaire

Constructs	Variables	Items	Measurements
Big data advantages	Recommendation System (ARE)	A1	I experience the advantages of a recommendation system function
		A2	I receive suggestions that are cheaper and time saving
		A3	I am recommended complementary packages
		A4	Recommendations are appropriate to my situation
	Information Search function (AIS)	A5	I experience the advantages of an information search function
		A6	I can easily look for relevant information
		A7	Search results are detailed, complete and match my needs
		A8	Search results are fast, precise and accurate
	Dynamic Pricing function (ADP)	A9	I experience the advantages of a dynamic pricing
		A10	I am suggested different options with different prices
		A11	I am suggested complementary packages with different prices
		A12	Offers are customized and personalized
	Improved Customer Services (ACS)	A13	I experience the advantages of an improved customers' service
		A14	I can refer to previous customers' feedbacks
		A15	I can track my orders while they are being processed
		A16	I am offered a personalized experience thanks to the improved customer service
Big Data Disadvantages	Privacy and Security (DPS)	D1	My personal information might be disclosed to other companies
		D2	My personal information might be stolen
		D3	My banking information might be stolen by cybercriminals
	Shopping Addiction (DSA)	D4	I spend too much time reviewing products
		D5	I will take advantage of cheap offers I have no immediate need for
		D6	I buy packages before adjusting my schedule when they are cheap
	Group Influences (DGI)	D7	I felt guilty of buying a package because it seemed unreasonable
		D8	I buy packages to follow my friends' travel plans
		D9	I will use travel websites based on their popularity
Online consumer behavior	Attention	AI1	The browsing experience catches my attention
		AI2	I pay attention to information on websites used by my friends
	Intention	I1	Search results tailored to my needs raise my interest
		I2	I am interested in websites my friends are using
	Desire	De1	I will search websites where results are suitable to my needs
		De2	I will use websites with better and customized services
	Action	Ac1	I will search for travel and tourism information
		AC2	I will buy packages I need
	Sharing and Social awareness	SS1	I will fill comments and reviews
		SS2	I will share my experience with family and friends
		SS3	I will rate packages I purchased since my friends do the same
		SS4	I will share the websites on my social media platforms
		SS5	I will share the websites I use with my friends
		SS6	I will help my friends make travel plans from my experience

3.3 Statistical Data Analyses and Descriptive Statistics

To yield the desired results, the Statistical Package for Social Sciences (SPSS 22.0) and the Analysis of Moment Structures (Amos 22.0) were used to compute and analyze the data. Different analyses were thus conducted to obtain the results: correlation analysis, exploratory factor analysis, reliability analysis, independent-samples *t*-test as well as three-way ANOVA.

The factorability of the big data items was examined and the existence of significant correlations presented in Table 3 suggested reasonable factorability. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.779, higher than the recommended value of 0.6 (Hair, Black, Babin & Anderson, 2014) and Bartlett's test of sphericity was statistically significant ($p < 0.001$) suggesting factor analysis' suitability. Based on eigenvalue and scree plot criteria, the two-factor' solution was selected to determine the latent constructs using principal component analysis with varimax rotation (Table 4). The factors explained 46% and 30% of the variance, with all seven indicators scattered over the two factors. Harman's single-factor test suggested common method variance was not an issue in the study (Podsakoff, Mackenzie, Lee & Podsakoff, 2003).

Table 3. Correlation among variables of the latent constructs

	Mean	Std	ARE	AIS	ADP	ACS	DPS	DSA	DGI	AI	DA	SS
ARE	5.223	0.753	1									
AIS	5.506	0.773	0.714***	1								
ADP	5.509	0.696	0.654***	0.666***	1							
ACS	5.615	0.714	0.695***	0.735***	0.667***	1						
DPS	5.524	0.925	0.107	0.001	0.102	0.119	1					
DSA	4.620	0.858	0.110	0.037	0.056	0.094	0.679***	1				
DGI	4.263	0.813	0.243***	0.139*	0.224**	0.244***	0.493***	0.661***	1			
AI	5.287	0.578	0.531***	0.409***	0.493***	0.617***	0.193**	0.274***	0.364***	1		
DA	5.298	0.731	0.512***	0.557***	0.563***	0.573***	0.230***	0.269***	0.402***	0.648***	1	
SS	4.880	0.694	0.269***	0.231***	0.273***	0.364***	0.103	0.339***	0.386***	0.603***	0.637***	1

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4. Rotated components factors matrix

Components		Factor Loading	
Big Data Advantages	ARE	0.868	
	AIS	0.895	
	ADP	0.846	
	ACS	0.879	
Big Data Disadvantages	DPS		0.845
	DSA		0.914
	DGI		0.812

Note. Factor loading less than 0.50 are not shown; extraction method: principal component analysis; rotation method: varimax rotation with Kaiser Normalization

Table 5 shows the descriptive statistics of each factor (gender, age and usage) in relation to both independent variables: big data advantages and big data disadvantages. An independent-samples t-test was conducted to compare each group's mean, as well as the mean and standard deviation for each specific advantage and disadvantage. Moreover, the reliability of each scale was also investigated. As the reliability of the measurement items for each construct was assessed, it is found that all the scales were internally consistent with Cronbach α values ranging between 0.823 and 0.898 (Augustin & Liaw, 2017). No better internal consistency could have been achieved if items were deleted from any of the scales as seen with the "Cronbach α if item deleted" function in SPSS. Therefore, the variables within the constructs of the model were all deemed reliable and internally consistent.

Table 5. Mean comparisons of gender, age and usage

	Gender			Age			Usage		
	Male (n=63)	Female (n=163)	t (p)	< 20 (n=101)	20 – 35 (n=125)	t (p)	Low (n=76)	High (n=150)	t (p)
Advant.	5.255 (0.556)	5.546 (0.657)	-3.106**	5.418 (0.685)	5.503 (0.607)	-0.973	5.295 (0.673)	5.550 (0.612)	-2.864**
ARE	5.044 (0.689)	5.301 (0.766)	-2.324*	5.114 (0.896)	5.322 (0.601)	-1.999*	5.033 (0.911)	5.328 (0.639)	-2.529*
AIS	5.258 (0.681)	5.601 (0.787)	-3.047**	5.480 (0.794)	5.526 (0.759)	-0.442	5.385 (0.720)	5.567 (0.794)	-1.676
ADP	5.357 (0.511)	5.568 (0.748)	-2.415*	5.426 (0.722	5.576 (0.669)	-1.620	5.345 (0.645)	5.592 (0.707)	-2.545*
ACS	5.361 (0.691)	5.713 (0.701)	-3.400**	5.651 (0.753)	5.586 (0.683)	0.679	5.418 (0.716)	5.715 (0.695)	-3.008**
Disadvtg.	4.875 (0.795)	4.617 (0.606)	2.328*	4.648 (0.733)	4.722 (0.619)	-0.820	4.803 (0.711)	4.631 (0.646)	1.822
DPS	5.661 (0.988)	5.470 (0.897)	1.395	5.518 (0.953)	5.528 (0.906)	-0.810	5.557 (0.913)	5.507 (0.934)	0.385
DSA	4.836 (1.054)	4.536 (0.757)	2.065*	4.637 (0.860)	4.605 (0.860)	0.275	4.746 (0.949)	4.556 (0.804)	1.578
DGI	4.492 (0.827)	4.174 (0.792)	2.675**	4.122 (0.718)	4.376 (0.869)	-2.358*	4.294 (0.754)	4.247 (0.843)	0.411

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Standard deviation is in parentheses.

4. Results and Discussions

A two-stage analytical procedure was used involving a confirmatory factor analysis (CFA) and the examination of the structural relationships among the constructs. Upon finding the results, a three-way ANOVA was used to assess the gender, age and usage differences in perceptions.

4.1 Measurement Model

The hypothesized model is tested employing structural equation modeling (SEM) with AMOS 22. The confirmatory factor analysis (CFA) tests a measurement theory based on overall model fit. The hypothesized model exhibits a good fit with the data collected as seen in Table 6.

Table 6. Model fit indicators

Goodness-of-fit Statistics	Criteria	Indicators
Chi-square / df	<5.000	1.937
Goodness of fit (GFI)	>0.900	0.963
Adjusted Goodness-of-fit Index (AGFI)	>0.900	0.915
Normed Fit Index (NFI)	>0.900	0.966
Comparative Fit Index (CFI)	>0.9500	0.983
Root Mean Square Residual (RMR)	<0.040	0.038
Root Mean Square Error of Approx. (RMSEA)	<0.080	0.065

Resource: Tabachnick, Fidell & Osterlind, 2001; Bagozzi & Yi, 2012; Hair, Sarstedt, Ringle & Mena, 2012.

All standardized factor loadings were significant and ranged between 0.710 and 0.860, higher than the suggested values. It is recommended that factor loadings should be greater than 0.70 (Hair et al., 2010; Hair et al., 2011). Convergent validity of the CFA is to be supported by the composite reliability (CR) and the average variance extracted (AVE). Estimates of CR and AVE should respectively be higher than 0.70 and 0.50 (Hair, 2010), and as can be seen in Table 7, values of CR ranged between 0.831 and 0.899 while those of AVE ranged between 0.623 and 0.690, higher than the suggested estimates.

Table 7. Factor loadings, Composite Reliability and Average Variance Extracted of the model

Construct	Items	Standardized Factor Loading	Composite Reliability (CR)	Average Variance Extracted (AVE)
Big Data Advantages	ARE	0.82***	0.899	0.690
	AIS	0.85***		
	ADP	0.79***		
	ACS	0.86***		
Big Data Disadvantages	DPS	0.72***	0.836	0.634
	DSA	0.93***		
	DGI	0.72***		
Consumers' Behavior	AI	0.79***	0.831	0.623
	DA	0.86***		
	SS	0.71***		

Note. *** p < 0.001

With the use of the latent variable correlation matrix, discriminant validity can be established. Determination of the discriminant validity is done by looking down the columns and across the rows of the latent variable correlation matrix which has the square root of AVE on the diagonal and the correlations among the off-diagonal measures (Fornell & Larcker, 1981). Discriminant validity is satisfactory because the diagonal elements are larger than the off-diagonal elements (Table 8).

Table 8. Latent variable correlation matrix: Discriminant validity

	BD Advantages	BD Disadvantages	Cons. Behaviour
BD Advantages	**0.831**		
BD Disadvantages	0.220	**0.796**	
Cons. Behaviour	0.587	0.406	**0.789**

Note. Square root of AVE is on the diagonal.

4.2 Structural Equation Model

The structural model's results and the standardized path coefficients indicating the effects between the constructs of the model are shown in Figure 2. There is a positive relationship between big data advantages (BDA) and online consumers' behavior (OCB) (H_1: $\beta_1 = 0.555$, $t = 11.055$, $p < 0.001$) and it indicates that the first hypothesis was supported. There is also a positive relationship between big data disadvantages and OCB (H_2: $\beta_2 = 0.192$, $t = 4.356$, $p < 0.001$) indicating that H_2 was not supported. Let's also note that gender, age and usage were used as control variables for the dependent variable and none of them significantly impacted online consumers' behavior (their significance levels being completely above 0.05).
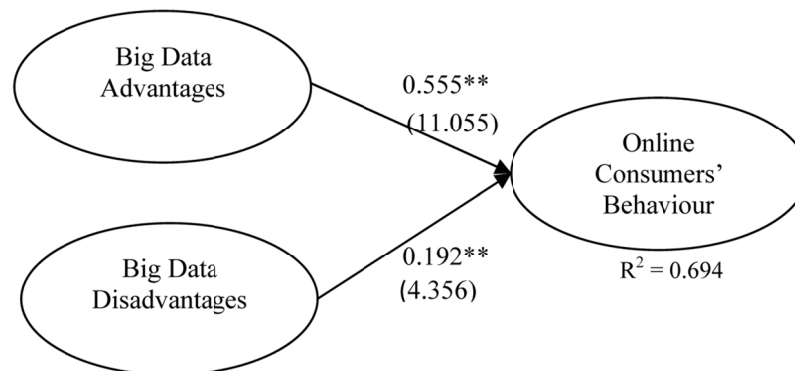

Figure 2. Research model results

Main disadvantages consumers are concerned with when big data technologies are used in e-commerce contexts are regrouped into three categories as seen above: privacy and security; shopping addiction and groups influences. Respondents usually claim that information privacy and security when surfing the internet are their main concern; however, they rarely behave accordingly (Kokolakis, 2017). The inconsistency of privacy attitudes and privacy behavior known as the "information privacy paradox" or "the privacy paradox" is very significant in terms of implications for e-commerce considering that e-commerce and online social network platforms are heavy collectors of personal information (Kokolakis, 2017). Lee, Park and Kim (2013), in their study assessing the influence of expected benefits and risks on users' intention to share personal information, concluded that users, despite concerns about expected benefits/expected risks, actively share personal information given they consider the risks as well as the expected benefits of sharing. It is however important to mention that in the context of big data, data that directly or indirectly carry personal information are considered personal information and in general, privacy information is the information that can be used to identify an individual. Users usually refer to private information as the information they provide about themselves. However, in the big data universe, private information can come either from personal data collected with users' acknowledgement or from online footprints created without users' awareness (Qing & Pivot, 2015). Very few users are aware of the real threat to their privacy and how their data is being collected; therefore, their trust in online platforms is based on their ability to change a few parameters of their privacy and security settings. Qing and Pivot (2015) indicated that online data has no border, it can always be collected and used, whether it is protected or not. Being unaware that your data is being collected makes a significant difference in any kind of behavior exhibited. Therefore, one can understand why privacy or security concerns do not negatively affect online consumer behavior as seen in our results.

Online shopping addiction referenced as the compulsive buying present in e-commerce environments is often not considered a problem by the individual suffering from it or even by society. Shopping addiction is seen as the most socially accepted addiction individual might suffer from. Various reasons are presented as justifications for online shopping such as ease of search, lower prices, wider selection, time saved, ease of use, fun, promotions and impulse behaviors of the shopper (Günüç & Doğan Keskin, 2016). Not often seen as a problem, it is understandable that it positively affects consumer behavior. When one shops with the intent to identify himself or herself to a certain group, whether it leads to personal identity loss or not, it is also not seen by some as a problem. This helps understand as well the manner in which this big data disadvantage positively influences online consumer behavior.

Overall, one would expect that big data disadvantages would negatively impact online consumers' behavior; however, all the previously explained reasons elaborate on the rationale behind the fact that big data disadvantages positively influence online consumers' behavior.

4.3 Test of Gender, Age and Usage Differences in Perception of Big data

A three-way ANOVA test was conducted to assess whether gender, age and usage affected the perception of big data applied in online tourism and whether they are intertwined (see Table 8 and Figure 3). The results of the three-factor ANOVA indicated that gender, age and usage's combined effect influenced the perception of big data's advantages which leads us to conclude that H_{3-1} was supported. The same is not true of big data's disadvantages and H_{3-2} was not supported (but is at the margin of significance level). In contrast, the perception of big data's disadvantages is influenced by the combined effect of gender and usage. From the results, we observed that, in all instances, perception of big data is influenced only when usage is involved in any of the significant interactions (A*B*C for advantages and A*C for the disadvantages). A look back at the mean differences previously presented in the descriptive statistics, it is noted that usage differences are significant for all but one big data advantage (AIS). Therefore, the perception of big data is significantly different between light and heavy users because they enjoy the advantages of a recommendation system, a dynamic pricing system and particularly an improved customer care function which is very significantly different between light and heavy users. However, there is no significant difference between light and heavy users with regards to the big data disadvantages.

It is worth underlining gender being at the margin of significance level when it comes to the advantages ($p = 0.055$). It is estimated that with a larger sample size, a different scenario could have been observed. The same is noticed of the interaction of gender, age and usage on the disadvantages where the significance level of the interaction is also at the margin of significance level ($p = 0.062$).

Table 9. Three-factor analysis of variance of Big Data

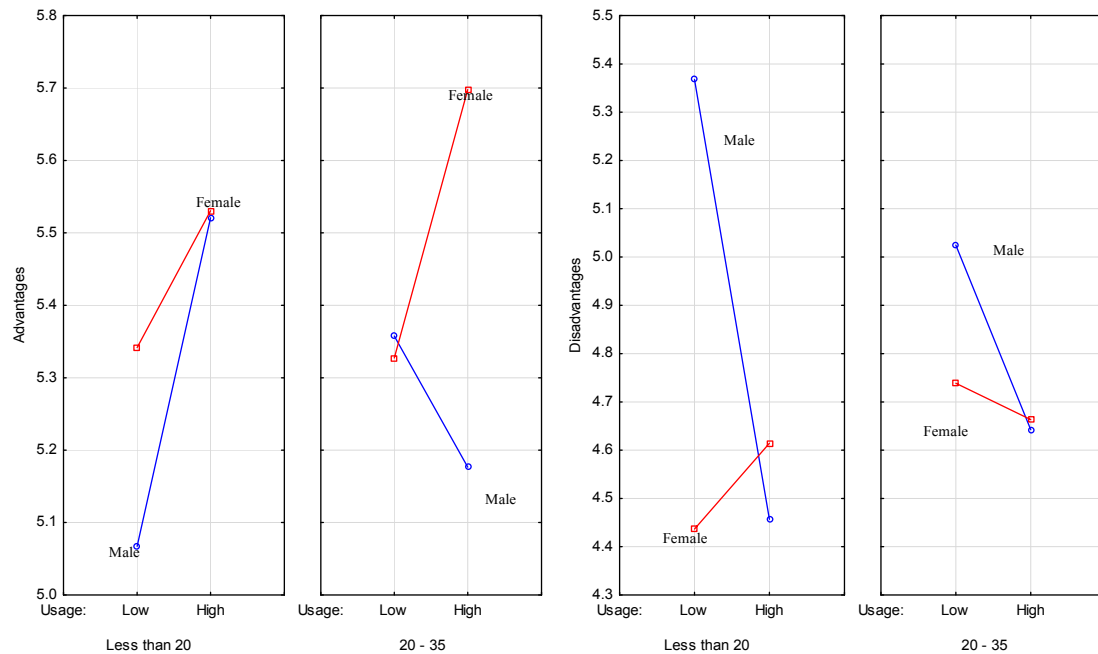|	Advantages					Disadvantages				
	df	SS	MS	F	p	df	SS	MS	F	p
Gender (A)	1	1.442	1.442	3.736	0.055	1	2.593	2.593	**6.220***	**0.013**
Age (B)	1	0.024	0.024	0.061	0.805	1	0.090	0.090	0.217	0.642
Usage (C)	1	1.664	1.664	**4.311***	**0.035**	1	3.404	3.404	**8.167****	**0.005**
A * B	1	0.103	0.103	0.266	0.607	1	0.620	0.620	1.488	0.224
A * C	1	0.198	0.198	0.513	0.474	1	4.680	4.680	**11.227****	**0.001**
B * C	1	0.493	0.493	1.278	0.260	1	0.182	0.182	0.437	0.509
A * B * C	1	1.600	1.600	**4.144***	**0.043**	1	1.468	1.468	3.522	0.062
Error	218	84.163	0.386			218	90.877	0.417		

Note. *** p < 0.001, ** p < 0.01, * p < 0.05

Figure 3. Interaction between gender, age and usage

The previous figure illustrates the combined effect of gender, age and usage on the perception of big data advantages and disadvantages in online tourism. Although previously discussed, but the figure presents a clearer picture of the interaction between the three factors. It indicates that women, in both age groups, have a positive perception of big data advantages applied in online tourism as their usage increase from low to high. For their male counterparts, the scenario is contrasted. Younger and male heavy users have a more positive perception of big data advantages compared to high users in the second age category (20 to 35). Male light users in both age groups are more concerned with big data disadvantages than male heavy users in the two age groups as well as their female counterparts.

4.4 Managerial Implications

Application of big data imposed itself as the new innovation in the e-commerce environment which includes e-tourism. The results of our study highlighted that, although consumers express their concerns regarding the disadvantages associated with big data technologies, it doesn't necessarily affect their behavior in a negative way. Therefore, applying big data technologies will be rewarding because these concerns do not translate into negative behavioral patterns that would drive customers away. Evidence of this is found in the study of Le and Liaw (2017) on consumers' responses in the context of big data applied in e-commerce where they discussed for example that privacy and security have a very low impact on consumers' behavior. Here, we found that privacy and security still positively affect consumers' behavior for various reasons as discussed earlier.

Various factors might influence customers' behavior. Businesses should pay close attention to the combined effect of gender, age and usage when targeting a specific audience. Understanding differences in perceptions from each group will help managers adopt the most suitable strategy to address their needs. Better understanding of customers' needs leads to better offerings, better recommendations and possibly profit maximization.

Ultimately, how consumers perceive big data will help governments design policies that will tackle specific challenges and disadvantages. For example, governments should address consumers' privacy and security concerns because this is in the best interest of the overall economy and particularly the tourism industry. Adopting the right policies to deal with the issue will help establish consumers' trust and might ultimately help e-businesses increase their profit, which ultimately will help the overall economy.

Among all studied factors, usage of online services is definitely one companies might need to pay a closer attention to assuming heavy users will enjoy all big data functionalities more than light users. This might help them figure out innovative ways to better attract light users' attention as well by either creating contents that will have them spending more time online or by improving the overall experience for them.

5. Conclusion

The inquiry was to know if gender, age and usage are important factors when it comes to the perception of big data applied in online tourism. It is important to highlight that in the present context big data was referred to as the advantages and disadvantages attached to its use in online tourism. Moreover, the study intended to explore the relationship between big data (advantages and disadvantages) applied in online tourism and self-reported consumers' behavior online. The results and the discussion that followed indicated the interactive effect of gender, age and usage does matter in the perception of big data's advantages. At a margin, this could be true as well for big data's disadvantages. Although main effects can be misleading in an analysis of variance, a closer look indicated age might not be a significant factor in the perception of big data. Further results also explained the overall positive impacts of big data advantages on online consumer behavior. This result is similar to a previous study by Le and Liaw (2017) where it is found that consumers' responses were positively affected by positive factors of applying being data analytics in e-commerce contexts. We found disadvantages of big data applied in online tourism to positively impact online consumers' behavior, which contrasts with previous research (Le & Liaw, 2017). Although disadvantages on big data might be major concerns to consumers', it is argued that they pay closer attention to the benefits they get from the advantages. As stated by Lee et al. (2013), consumers consider the risks as well as the benefits; which implies that if the benefits are greater than the risks, consumers' behavior will hardly be negatively impacted.

Big data advantages are there for businesses to profit from and address consumers' need for convenience. However, the use of big data brings its own challenges. Ultimately, the decision sits with businesses and firms to weight on how to create value from the use of big data. It is clear that big data technologies carry various benefits for consumers, firms and businesses alike. However, aside from disadvantages affecting consumers', there are specific challenges associated with big data's adoption for companies. The heavy expenditures of using big data technologies is a major concern and companies will have to analyze whether the value they will extract from using the technology is worth the cost (in terms of infrastructures as well as trained and qualified resources to manage the technology) it requires.

5.1 Limitations and Future Study

Big data is one of the latest buzz words and there is frenzy among all industries and businesses to take advantage of its use. The present study focused on exploring Taiwanese consumers' perception of the big data technology being applied in online tourism contexts. The study relied on questionnaire survey and collected self-reported behavioral information which might not provide a very accurate picture considering that consumers' might behave differently in real-life situations where many external factors might influence their decision-making process.

Because of usage's common impact in the interaction terms as previously seen, it would be interesting to further investigate whether it could act as a mediator of the relationship between big data (advantages and disadvantages) and online consumers' behavior.

References

Augustin, J.-L. P., & Liaw, S.-Y. (2017). Tourism Competitiveness Index of the Asia-Pacific Region through Consistency Analysis. *Asia Pacific Journal of Tourism Research*, *12*, 1295−1307. https://doi.org/10.1080/10941665.2017.1391855

Augustin, J.-L. P., & Liaw, S.-Y. (2020). Exploring the Relationship between Perceived Big Data Advantages and Online Consumers' Behaviour: An Extended Hierarchy of Effects Model. *International Business Research*, *13*(6), 73−85. https://doi.org/10.5539/ibr.v13n6p73

Bagozzi, R. P., & Yi, Y. (2012). Specification, Evaluation and Interpretation of Structural Equation Models. *Journal of the Academy of Marketing Science*, *40*(1), 8−34. https://doi.org/10.1007/s11747-011-0278-x

Buhalis, D. (1998). Strategic Use of Information Technologies in the Tourism Industry. *Tourism Management*, *19*(5), 409−421. https://doi.org/10.1016/S0261-5177(98)00038-7

Edwards, J. R., & Bagozzi, R. P. (2000). On the Nature and Direction of Relationships Between Constructs and Measures. *Psychological Methods*, *5*(2), 155−174. https://doi.org/10.1037/1082-989X.5.2.155

Ehrenberg, A. S. (2000). Repetitive Advertising and the Consumer. *Journal of Advertising Research*, *40*(6), 39−48. https://doi.org/10.2501/JAR-40-6-39-48

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable variables and Measurement Error. *Journal of Marketing Research*, 39−50. https://doi.org/10.2307/3151312

Ghandour, A. (2015). Big Data Driven E-Commerce Architecture. *International Journal of Economics, Commerceand Management*, *3*(5), 940–947.

Günüç, S., & Doġan Keskin, A. (2016). Online Shopping Addiction: Symptoms, Causes and Effects. *Addicta: The Turkish Journal of Addictions*, *3*, 353–364. https://doi.org/10.15805/addicta.2016.3.0104

Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science*, *40*, 414–433. https://doi.org/10.1007/s11747-011-0261-6

Hair, J. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis: Pearson New International Edition* (7th ed.). Edinburgh: Pearson Education Limited.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*. New York, NY, USA: Guilford Publications.

Kokolakis, S. (2017). Privacy Attitudes and Privacy Behavior: A Review of Current Research on the Privacy Paradox Phenomenon. *Computers & Security*, *64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Kuntz, W., Aksoy, L., Bart, Y., Heinonen, K., Kabadayi, S., Ordenes, F. V. et al. (2017). Customer Engagement in a Big Data World. *Journal of Services Marketing*, *31*(2), 161–171. https://doi.org/10.1108/JSM-10-2016-0352

Le, T. M., &Liaw, S.-Y. (2017). Effects of Pros and Cons of Applying Big Data Analytics to Consumers' Responses in an E-commerce context. *Sustainability*, *9*, 798. https://doi.org/10.3390/su9050798

Lee, H., Park, H., & Kim, J. (2013). Why do People Share their Context Information on Social Network Services? A Qualitative Study and an Experimental Study on Users' Behavior of Balancing Perceived Benefit and risk. *International Journal of Human-Computer Studies*, *71*(9), 862–877. https://doi.org/10.1016/j.ijhcs.2013.01.005

Lee, T. R., Lin, J. H., Liao, L. W., & Yeh, T. H. (2013). Managing the positive and negative characteristics of enterprise microblog to attract user to take action through the perspective of behavioural response. *International Journal of Management and Enterprise Development*, *12*(4–6), 363–384. https://doi.org/10.1504/IJMED.2013.056443

Line, N. D., Dogru, T., El-Manstrly, D., Buoye, A., Malthouse, E., &Kandampully, J. (2020). Control, Use and Ownership of Big Data: A Reciprocal View of Customer Big Data Value in the Hospitality and Tourism Industry. *Tourism Management*, *80*. https://doi.org/10.1016/j.tourman.2020.104106

Liu, W.-Y., Lin, C.-C., Lee, Y. S., & Deng, D.-J. (2013). On Gender Differences in Consumer Behavior for Online Financial Transactions of Cosmetics. *Mathematical and Computer Modelling*, *58*, 238–253. https://doi.org/10.1016/j.mcm.2012.08.010

Mitchell, V.-W., & Walsh, G. (2004). Gender Differences in German Consumer Decision-making Styles. *Journal of Consumer Behavior*, *3*(4), 331–346. https://doi.org/10.1002/cb.146

Morris, M. G., Venkatesh, G., & Ackerman, P. L. (2005). Gender and Age Differences in Employee Decision about New Technology: An Extension of the Theory of Planned Behavior. *IEEE Transactions on Engineering Managment*, *52*, 169–184. https://doi.org/10.1109/TEM.2004.839967

Olson, K. E., O'Brien, M. A., Rogers, W. A., &Charness, N. (2011). Diffusion of Technology: Frequency of Use for Younger and Older Adults. *Ageing International*, *36*, 123–145. https://doi.org/10.1007/s12126-010-9077-9

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in IS research. *MIS Quaterly*, *31*(4), 657–779. https://doi.org/10.2307/25148814

Podsakoff, P. M., Mackenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, *8*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Qing, T., & Pivot, F. (2015). *Big Data Privacy: Changing Perception of Privacy*(pp. 860–865). IEEE International Conference on Smart City. Chengdu: IEEE Computer Society. https://doi.org/10.1109/SmartCity.2015.176

Sorce, P., Perotti, V., &Wildrick, S. (2005). Attitude and Age Difference in Online Buying. *International Journal of Retail & Distribution Management*, *33*(2), 122–132. https://doi.org/10.1108/09590550510581458

Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). *Using Multivariate Statistics*. Upper Saddle River, NJ, USA: Pearson Education.

Tang, M., & Wu, Z. (2015). Research on the Mechanisms of Big Data on Consumer Behavior using C2C e-Commerce and Countermeasures. *African Journal of Business Management*, *9*, 18−34. https://doi.org/10.5897/AJBM2014.7560

Van derKaay, C. D., & Young, W. H. (2012). Age-Related Differences in Technology Usage Among Community College Faculty. *Community College Journal of Research and Practice*, *36*(8), 570−579. https://doi.org/10.1080/10668920903054865

Van Slyke, C., Bélanger, F., Johnson, R. D., & Hightower, R. (2010). Gender-Based Differences in Consumer E-Commerce Adoption. *Communications of the Association for Information Systems*, *26*(2), 7−35. https://doi.org/10.17705/1CAIS.02602

Van Slyke, C., Comunale, C., & Belanger, F. (2002). Gender Differences in Perceptions of Web-based Shopping. *Communications at the ACM*, *45*(7), 82−86. https://doi.org/10.1145/545151.545155

Vijaya Lakshmi, V., Aparanjini Niharika, D., & Lahari, G. (2017). Impact of Gender on Consumer Purchasing Behavior. *IOSR Journal of Business & Management*, *19*(8), 33−36.

Copyrights