Understanding customer satisfaction with services by leveraging big data: the role of services attributes and consumers' cultural background

Article

Accepted Version

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Understanding Customer Satisfaction with Services by leveraging Big Data: the Role of Services Attributes and Consumers' Cultural Background

MARCELLO MARIANI[1], GIUSEPPE DI FATTA[2], (Member, IEEE), MARCO DI FELICE[3]

[1]Henley Business School, University of Reading, Whiteknights, Reading, RG6 6AY, United Kingdom (e-mail: m.mariani@henley.ac.uk)
[2]Department of Computer Science, University of Reading, Whiteknights, Reading, RG6 6AY, United Kingdom (e-mail: G.DiFatta@reading.ac.uk)
[3]Department of Computer Science and Engineering, University of Bologna, Bologna, Italy (e-mail: difelice@cs.unibo.it)

Corresponding author: Marcello Mariani (e-mail: m.mariani@henley.ac.uk).

ABSTRACT User-generated content and online reviews are becoming an increasingly relevant source of information for online customers that use them for purchasing decisions. This study examines the impact of services attributes and consumers' cultural background on customer satisfaction with services in an online setting using big data. First, almost half a million Expedia.com hotel online reviews related to hotel properties located in five different countries (United States, United Kingdom, Italy, Spain, Russia) were retrieved. Second, the resulting dataset was used to investigate if and to what extent the overall customer satisfaction with a service is affected by the evaluation of specific hotel services attributes (operationalized based on an established typology of attributes) and by the consumers' cultural background (operationalized by means of Hoftstede's framework). A comprehensive multivariate regression analysis is carried out to test the literature-driven hypotheses formulated. In particular, the analysis reveals that critical service attributes such as hotel condition, room comfort, service and staff, and cleanliness positively affect the overall online satisfaction ratings. The cultural dimensions of power distance, individualism and uncertainty avoidance negatively affect overall online satisfaction, while long-term orientation and indulgence positively affect online satisfaction. Masculinity seem not to play a significant role. We also observe that reviews' text length exerts a negative impact on online ratings. Theoretical and practical implications are discussed.

INDEX TERMS Big Data, Online Reviews, Online Rating, Expedia.com, Hotel, Cultural Differences

I. INTRODUCTION

THE development of digital technologies and digital platforms has brought about a proliferation of data in the form of user-generated content (UGC) and online reviews (ORs). The latter ones constitute a relevant source of information for both companies and consumers [1], [2] and have now become the second most relevant information source after the traditional Word of Mouth (WOM) from friends and relatives [3]. This is particularly true not only for consumer electronics products, but also for travel, hospitality and tourism services [4] where ORs inform consumers' decisions and online consumer behaviour. For instance, the development and widespread adoption of online travel sites such as TripAdvisor.com and Online Travel Agencies (OTAs) such as Expedia.com and Booking.com are generating a significant amount of data related to hospitality and tourism services and experiences, which is available to virtually any consumer with Internet access [5]. Moreover, ORs are increasing in volume at a tremendous velocity: the number of reviews on TripAdvisor have increased from 1 million at the beginning of 2005 to more than 570 million in 2017 (TripAdvisor, 2005, 2018) and an analogous exponential growth can be detected also for OTAs such as Expedia.com and Booking.com. Consequently, this hyper-production of reviews is likely to generate information loads for both consumers and decision makers in firms and companies. A recent review on big data and business intelligence in the hospitality and tourism management field [6] has pointed to the fact that the fast hyper-production of ORs is posing significant challenges to managers, as they have to deal increasingly with big data

analytics to create value for their customers and improve firm's performance and profitability [7]. While the hospitality and tourism management and marketing research fields are trying to catch up with extant gaps in big data analysis, much remains to be done. A number of studies over the last few years have focused on the analysis of guest reviews of hotels [8], but literature leveraging huge amounts of data retrieved and analysed through big data techniques is still scant. Furthermore, most of the studies conducted so far have focused on individual tourism destinations (i.e., a city or a town or a few towns within a country), and therefore have not been able to broaden the scope of the analysis to generalize their findings. Accordingly, our study makes a first contribution which is methodological as it is the first study comprehensively examining online customer behaviour in the hospitality sector in five distinctively different countries (United States, United Kingdom, Italy, Spain, Russia) by gathering the highest number of certified ORs across countries thus far, from the leading OTA Expedia.com. Therefore, we expect our findings to be robust and generalizable across a number of different cities and countries. Secondly, our study is distinctive as, to the best of our knowledge, it is one of the first attempts to analyse conjointly the impact of hotel service attributes and online reviewers' cultural background on their online satisfaction based on huge volumes of data. This is particularly relevant in a service sector, as the online evaluation of services does not only depend on the attributes of services but also on encounters between service providers and service consumers often from different cultures. We innovatively operationalize cultural differences in big data studies through the Hofstede framework of cultural dimensions [9], [10]. Third, the study generates knowledge that is actionable by managers and decision makers active in one of the leading industries in the world accounting for 10.4% of global GDP and 9.9% of total employment in 2017 [11]. The industry is also relevant besides its dimension as it is growing at a very fast pace also due to technological advancements in transportation and ICTs: for instance international tourist arrivals have increased exponentially over the last sixty years thus reaching the historical record of 1.32 billion in 2017 [12]. The paper is organised as follows. Section II critically reviews the relevant social sciences and marketing management literature and develops the research hypotheses. Section III illustrates the empirical setting and the research methodology. Section IV presents the research findings. Section V discusses the major managerial implications. Finally, Section VI summarizes the conclusions, discusses the limitations and identifies a few further research questions worth exploring.

II. LITERATURE REVIEW AND HYPOTHESES FORMULATION

A. ONLINE REVIEWS AS BIG DATA TO GENERATE CONSUMERS' KNOWLEDGE IN THE HOSPITALITY SECTOR

The establishment and diffusion of two-sided and multi-sided platforms empowered by digital technologies has made it possible for companies such as Amazon, TripAdvisor and Expedia to connect a number of different market players including groups of suppliers, producers, intermediaries and customers. Most of these platforms are matchmakers, as they match one group of stakeholders with another group by reducing the transaction costs of each match [13]. The economic and business properties of those platforms and the role they play in industry competition have been extensively studied by economists [14] and increasingly also in management and marketing sciences [15], [16]. A subset of the aforementioned platforms have emerged as the de facto digital intermediaries in many industries: they include Amazon for consumer goods, and Expedia and Booking for travel and hospitality services. More specifically, OTAs such as Booking and Expedia today intermediate on their platforms an increasing share of all the transactions related to travel, tourism and hospitality products and services [17]. According to the latest market research, OTAs' share of the US online travel market has reached 39% in 2016 and will continue to increase to 41% by 2020, totalling US$81.4 billion in online gross bookings [18]. Almost 80% of the online bookings is performed through the OTAs Booking and Expedia and almost one third of leisure travellers and a half of business travellers use OTAs to compare the prices of available alternatives [17]. Besides representing relevant sources of information for travel planning, OTAs provide additional value to their customers as they allow them to write and read ORs. In the marketing management science literature ORs have been termed as electronic word-of-mouth (eWOM) whereby eWOM is "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" ([1] p. 39). The antecedents of eWOM (i.e., ORs generating factors) and its impact on either consumers or companies have been largely examined in the hospitality management literature [5]. Among the antecedents of ORs, we can recall the desire to express satisfaction or dissatisfaction with the hotel service, a helpful attitude towards other travelers, the degree of involvement with the purchase, socio-demographics characteristics, etc. Among the impact of eWOM on other consumers there are features related to the review such as valence, polarity, quality, trustworthiness (e.g., [19]). From a company perspective, eWOM has been found to be a relevant determinant for firm performance, strategies and reputation (e.g., [7]). As noted in a recent systematic quantitative literature review of business intelligence and big data in the domains of hospitality and tourism management [6], ORs (and therefore eWOM) represent today an important type of Big Data that are generated in large volumes, at high velocity (typically within a few days or weeks after the end of a trip or hotel stay) and in a variety of formats (i.e., written texts that is temporally and spatially referenced). Accordingly, ORs comply with the three major characteristics synthesized in the so-called "3Vs" of volume, velocity and variety [20], [21].

However, most of the studies conducted so far leveraging

ORs typically deploy sample sizes of less than 100,000 reviews and do not rely on BD-specific technologies such as noSQL databases (e.g., [22], [23]) or other BD techniques and tools for storage purposes and distributed processing of datasets on clusters of commodity hardware [6]. Moreover, all of the studies carried out so far have examined ORs pertaining to hotels located in a specific municipality or country, without generalizing their results by means of cross-country and cross-continental evidence. For instance, ORs have been used to capture online customer satisfaction [24] and reviews' features [25], mostly in individual countries: five cities in China [24], the Manhattan area in New York in the US [25]. For instance, [24] leverages on the highest number of ORs (i.e., 412,784) in hospitality management studies, if we exclude the methodological study carried out recently by [26]. The authors use TripAdvisor online reviews for 10,149 hotels from five Chinese cities to examine how tourists speaking different languages (namely English, French, German, Italian, Japanese, Portuguese, Russian and Spanish) give different online ratings to different hotel attributes such as "rooms", "cleanliness", "value". The study finds that Chinese tourists domestically exhibit distinct preferences for hotel attributes when compared to foreign tourists and interaction effect exist between the attributes "Rooms" and "Service" and between "Value" and "Service". However, the reviews sourced from TripAdvisor have a huge limit, as they are not certified like those present on OTAs such as Booking and Expedia [6]. In other terms, there is not a mechanism in place on TripAdvisor to double and cross check if the online reviews were written by real guests after an actual hotel stay. In [26] authors empirically examine the effects of the Booking.com rating system on the distribution of hotel ratings for the entire population of hotels located in London over a period of two years. Leveraging on 1.2 million online reviews, they illustrate that the overall distribution of hotel scores is significantly left-skewed.

The body of management and marketing literature within hospitality so far has mostly investigated to what extent the attributes of a hotel explain the overall level of satisfaction. For instance, [27], based on a small sample of 343 reviews of 187 Costa Rican lodges classify the attributes into critical, neutral, dissatisfiers, satisfiers based on the framework of [28]. Overall, it appears that existing studies have not sufficiently addressed the extent to which there is an interplay between the individual attributes of a hotel and cultural factors pertaining to the reviewer. Only a handful of studies has tried to analyze the role played by consumers' cultural background and its impact on online consumer satisfaction. For instance, [29] observe that readers' national orientation towards individualism or collectivism moderates the effect of ORs antecedents on the perception of information credibility. Additionally, the orientation towards individualism has been found to moderate negatively the relationships between information consistency/information rating and information credibility.

This study addresses the aforementioned research gap and examines if and to what extent services attributes and reviewers' cultural background affect conjointly online customer satisfaction. This study is distinctive among studies deploying big data within the hospitality marketing management literature in that: 1) it generates market intelligence about customers that might be treasured to enhance the competitive advantage of hospitality firms [30]; 2) it leverages on almost half a million online reviews related to hotels located in cities scattered across five different countries and three continents, thus allowing to achieve solid generalization to understand online customer satisfaction with hotel services; 3) it sources data from an OTA whose reviews are certified thus ensuring that data comply with an additional "V" related to "veracity" [6] to be juxtaposed to the 3Vs of big data [21] and to be conceived as reliability, validity and completeness of data; 4) it also sources data from the Hofstede center guaranteeing an accurate triangulation of data from diverse sources [31], [32]; 4) it deploys the Big Data technologies such as a mix of noSQL databases and other software tools developed for this specific study; 5) it assesses conjointly how individual hotel attributes and the reviewer's country of origin affect the overall satisfaction with a hotel service by leveraging on frameworks developed in offline settings [10], [28] and apply them in online settings. Overall, by clearly acknowledging the challenges brought about by Big Data [33]–[35] we contribute more broadly to the management and marketing science literatures by acknowledging that big data can make a difference to better understand firms' and customers' behaviors [36] and especially can assist firms in creating, delivering and capturing customer value [30].

In the next subsections, we discuss two theoretical frameworks adopted within management sciences to understand how services attributes and cultural factors can affect customer satisfaction in offline settings and we leverage on them to develop our hypotheses related to online settings.

B. THE ROLE OF HOTEL SERVICES ATTRIBUTES IN CONSUMER BEHAVIOUR WITHIN THE HOSPITALITY AND TOURISM SECTOR

In building the so called motivation-hygiene theory (M-H theory) to explore the underpinnings of job satisfaction, [37] underline that the presence of one set of job features or incentives leads to higher satisfaction at work, while the presence of another set of job characteristics leads to dissatisfaction. Drawing on the M-H theory [37], [28] put forward a typology of hotel attributes by leveraging on data stemming from a survey of managers and executives of the American Hotel and Motel Association conducted in 1978. In [28] authors suggest that several attributes (dissatisfiers) can be considered as salient in their potential to cause dissatisfaction while other attributes (satisfiers) can be considered as salient in their potential to cause high level of satisfaction. Some (criticals) have the capacity to cause both dissatisfaction and high satisfaction, and others (neutrals) are referred to very infrequently. An example of a dissatisfier from the aforementioned study is the accuracy of bill processing: there is

a minimum performance on this dissatisfier to be maintained (if there is a billing mistake, it should be corrected quickly) to avoid dissatisfaction. However, any efforts to achieve very high performance on this attribute (billing without any mistake) may not lead to customer satisfaction with the hotel. Employees' helpful attitude, establishment cleanliness, and variety of service are examples of a satisfier, a critical, and a neutral attribute respectively. Adopting a content analysis approach to the content analysis of a 343 Tripadvisor ORs, authors in [27] find that critical attributes relate to: ambiance, ecofriendliness, room and bathroom decor and layout, customer service, tour, food quality, room rates. Among the dissatisfiers they list: grounds and surroundings, noise, lodge amenities, room and bathroom facilities, room amenities, insect problems, reservation process, management policies, natural attractions. Among the satisfiers they include: other guests, nature-based activities. Last, among the neutral attributes they include: accessibility, closeness to town, closeness to attractions, extra service, restaurant service, food and drink price, other prices, weather. Consistently with those findings obtained in offline settings [28] and qualitatively with small data content analysis ([27], we can hypothesize that the attributes available on the Expedia OTA at the time of data collection might affect overall customer satisfaction differently based on the cluster they belong to (critical, satisfiers, dissatisfiers, neutral). More specifically out of the 8 attributes that were available on the platform at the time of data retrieval, i.e., cleanliness, service and staff, room comfort, hotel condition, convenience of location, neighborhood, value for money, room quality, the first four are critical attributes while the latter four are neutral ones based on previous studies [28]. We therefore hypothesize that those attributes that will make most of the difference to explain online customer satisfaction are precisely cleanliness, service and staff, room comfort, hotel condition. We therefore hypothesize what follows.

- **H0**: Most of the overall satisfaction with a hotel can be explained by the satisfaction with the critical attributes of hotel condition, cleanliness, service and staff, room comfort. This holds regardless of the context/country where the hotel is located.
- **H1.a**: Hotel guests satisfied with the hotel cleanliness will give high overall online ratings, regardless of the context/country where the hotel is located.
- **H1.b**: Hotel guests satisfied with the hotel service and staff will give high overall online ratings, regardless of the context/country where the hotel is located.
- **H1.c**: Hotel guests satisfied with the hotel room comfort will give high overall online ratings, regardless of the context/country where the hotel is located.
- **H1.d**: Hotel guests satisfied with the hotel condition will give high overall online ratings, regardless of the context/country where the hotel is located.

C. THE ROLE OF CONSUMERS' CULTURAL BACKGROUND IN CONSUMER BEHAVIOUR WITHIN THE HOSPITALITY AND TOURISM SECTOR

The study of the role played by culture and cultural factors has a long tradition in the wide social sciences as well as general management studies [9], [38], [39] while it is relatively more recent in hospitality management studies [40]. Hospitality contexts especially in internationally renowned destinations are characterized by a high level of multi-culturalism whereby inter-cultural interactions between hotel service providers/employees and guests take place often and it is paramount for companies to deal with the needs, perceptions and expectations of consumers from different cultures [41]. All of the aforementioned aspects are becoming increasingly relevant in a globalised economy [42] where companies can achieve a global competitive advantage by both recruiting, motivating and retaining the best international talents [43] and by addressing effectively the needs, wants and preferences of a heterogeneous number of potential customers belonging to different cultures [44], [45]. Perhaps the two most influential frameworks deployed in tourism and hospitality management studies within offline settings are those developed by [39] and [9]. Based on [39], cultures can be distinguished in relation to several communication features: in some cultures (high-context communication), a non-verbal mode of communication is preferred, while in other cultures (low-context communication) direct and explicit communication is preferred. Building on a monumental research carried out in the seventies on IBM employees in a number of different subsidiaries scattered around the world, Geert Hofstede developed in the late seventies a comprehensive framework characterizing national cultures based on different dimensions such as power distance, masculinity, uncertainty avoidance, and individualism [9]. The framework has been recognized as the most influential in marketing management studies nowadays [46] and will be discussed in detail in the next subsection.

1) The Hofstede framework of cultural dimensions and its use in hospitality and tourism management

Among the existing theoretical frameworks adopted to investigate cultural aspects and issues in the wide management literature, certainly Hofstede's (1980) one is the most widely adopted. [10] defines culture as "the collective programming of the mind distinguishing the members of one group or category of people from others" ([10], p. 9). In the original version of the framework [9] defines four dimensions of culture: power distance (PD), individualism (IDV), masculinity (MAS), and uncertainty avoidance (UA). Power distance (PD) is "the degree to which the less powerful members of a society accept and expect that power is distributed unequally" (https://www.hofstede-insights.com/models/national-culture/). Individualism (IDV), relates to "a preference for a loosely-knit social framework in which individuals are expected to take care of only themselves and their immediate families" (ibidem). Masculinity

(MAS) refers to "a preference in society for achievement, heroism, assertiveness, and material rewards for success" (ibidem). Uncertainty Avoidance (UA) is defined as the "degree to which the members of a society feel uncomfortable with uncertainty and ambiguity" (ibidem). After further research conducted in the 90s in Asian contexts and cultures, Hofstede came up with two additional dimensions that have been added to the former ones: long-term orientation and indulgence [47]. Long-term orientation (as opposed to short-term orientation) gives value to societal change which is seen with suspicion in short term cultures that express a preference to "maintain time-honoured traditions and norms" (ibidem). Indulgence (IND) as opposed to restraint relates to societies that value and allow "relatively free gratification of basic and natural human drives related to enjoying life and having fun" (ibidem). All of the aforementioned six cultural dimensions have been operationalized by means of indexes that cover a large number of countries, with the exception of the latest two dimensions (i.e., long-term orientation and indulgence) whose coverage is less extensive. Despite a few criticisms of Hofstede's theoretical framework [48], several authors emphasize its relevance as a clear and useful framework for research aimed at capturing the effects of cultural factors in management and business studies. The Hofstede framework has been reasonably used by hospitality and tourism management scholars to interpret results of their inter-cultural researches in offline settings starting from the second half of the nineties. For instance, in a study conducted in two high-class hotels in Singapore, [41] finds that there are statistically significant differences between Asian and Western hotel guests in their evaluation of the hotel service encounter and overall service quality. The author discovers that Asian leisure travellers give a lower evaluation of the service encounter (as well as overall service quality) than their Western counterparts and the differences are explained by leveraging Hofstede's power distance (PD) dimension. [49] examine the complaint behavior of Western and Asian hotel guests about hotel services and find that Asian tourists are less likely to generate negative word-of-mouth than non-Asians. Moreover, if they do, they demonstrate a rather unexpansive public complaint behavior compared to non-Asians that is in line with Hofstede's model ([9], [47]). Limited attention has been paid so far to the impact that cultural factors might exert on the evaluation of hotel services in online settings: this implies that there is a need to understand if and to what extent cultural factors affect online hotel consumers' behaviours. In the following part of this section we elaborate on existing research using the Hofstede framework to interpret results in cross-cultural and cross-country studies and derive the rest of our hypotheses.

Power distance (PD) - [50] point out that in services industries (banking, consulting, hospitality, etc.) the relevance of the different service quality dimensions varies depending on the relative power of service providers and customers. Consequently, power distance assumes different meanings across industries: in banking and consulting, the provider of the service is more powerful based on the asymmetries due to its expertise, while in activities such as hospitality the provider has a low status and low power [41]. Accordingly, hospitality services' consumers from high power distance cultures expect services to be delivered with the highest level of quality [51], [52] as clear from evidence showing that the evaluation of hotel guests from Western countries (low in power distance) are higher than those of guests from Asian countries (high in power distance). Mirroring what happens in offline settings, we therefore hypothesize that reviewer's power distance (PD) is negatively related to online review ratings. Moreover, we are interested in understanding if the hypothesis holds in different countries (i.e., US, UK, Italy, Spain, Russia), where service providers display different values across cultural dimensions. Consequently, we hypothesize the following.

- **H2.a**: The reviewer's power distance (PD) affects negatively online ratings, so that reviewers from cultures with higher power distance (PD) give lower ratings regardless of the context/country where the hotel is located.

Individualism (IDV) - Within consumer research related to services industries, the dimension of individualism-collectivism is the one that has received a significant attention by scholars so far (e.g., [41], [53]). Most of the current services management and marketing literature seems to suggest that consumers from individualistic cultures have higher levels of expectations compared to customers espousing collectivistic cultures ([50], [53]–[55]). Furthermore, they and rely more on outcomes rather than processes in the hospitality sector [41]: for instance, Western customers (more individualistic consumers) give more importance to physical environment then their Asian counterparts and (not significantly) also to the service dimension. Moreover, consumers from highly individualistic cultures (e.g., the US, the UK and Canada) tend to complain more than their counterparts from collectivistic cultures (such as Singapore, China, and Korea). For instance, [56] show that American guests are more likely than Japanese to complain to hotel managers. As a result, we formulate the following research hypothesis.

- **H2.b**: The reviewer's individualism (IDV) affects negatively online ratings, so that reviewers from cultures with higher individualism (IDV) give lower ratings, regardless of the context/country where the hotel is located.

Masculinity (MAS) - The dimension of masculinity (MAS) has been investigated by a handful of scholars [50], [57]. The latter have found that high masculinity national cultures demonstrate less tolerance towards service failures [57] as they possibly pay more attention to responsiveness and reliability. Individuals from masculine societies are more likely to complain about poor service quality and can confront service providers for unsatisfactory experience or discontinue the experience [58]. Furthermore, cross-cultural eWOM studies seem to suggest that service customers with

low-masculine national orientation (e.g., Chinese) show a positive emotional attitude in their reviews [59]. Consequently, we hypothesize that in online settings:

- **H2.c**: The reviewer's masculinity (MAS) affects negatively online ratings, so that reviewers from cultures with higher masculinity (MAS) give lower ratings, regardless of the context/country where the hotel is located.

Uncertainty avoidance (UA) - Existing service research has underscored that consumers from low uncertainty avoidance cultures have higher tolerance for ambiguity than their counterparts and can cope with uncertainty in the short term ([51], [53], [54]). In more detail, [54] find that customers from cultures characterized by higher uncertainty avoidance engage more with search of information about services and therefore display higher expectations. As individuals from cultures with high levels of uncertainty avoidance tend to be risk-adverse, it is likely that they will give lower ratings both when the service doesn't and does meet expectations. Interestingly, another study shows empirically that consumers from high uncertainty avoidance cultures perceive service quality as lower than their counterparts in low uncertainty avoidance cultures [51]. Therefore, we formulate the following research hypothesis:

- **H2.d**: The reviewer's uncertainty avoidance (UA) affects negatively online ratings, so that reviewers from cultures with higher uncertainty avoidance (UA) give lower ratings, regardless of the context/country where the hotel is located.

Long-term orientation (LTO) - The temporal orientation dimension in the Hofstede framework points to the idea that individuals from cultures with a LTO might be willing to give up a part of their present benefits to improve their future benefits and rewards. In services-related studies, a few scholars have encapsulated and/or controlled for the temporal dimension into their analyses. For instance, studies in offline settings suggest that individuals in high LTO cultures (e.g., Korea) enjoy high trust while the opposite holds in low LTO cultures (e.g. US) and therefore LTO culture [60]. On the contrary, individuals from short-term-oriented cultures form higher expectations about service providers thus being more critical [61]. In practice it has been found that individuals from LTO cultures value loyalty with the service provider [62], [63]. Therefore, we formulate the following research hypothesis:

- **H2.e**: The reviewer's long-term orientation (LTO) affects positively online ratings, so that reviewers from cultures with higher long-term orientation (LTO) give higher ratings, regardless of the context/country where the hotel is located.

Indulgence (IND) - The relevant literature emphasizes that individuals from cultures characterized by a high level of indulgence are happier than those living in restrained ones, and they develop a more positive attitudes and optimism due to the fact that as they are more likely to remember positive emotions [64]. In online settings, individuals from cultures with a high level of indulgence are more prone to use online social networks than individuals from restraint-oriented cultures [65]. As a consequence, we hypothesize the following:

- **H2.f**: The reviewer's indulgence (IND) affects positively online ratings, so that reviewers from cultures with higher indulgence (IND) give higher ratings, regardless of the context/country where the hotel is located.

III. EMPIRICAL SETTING AND RESEARCH METHODOLOGY

A. EMPIRICAL SETTING: AN ATTEMPT TO GENERALIZE FINDINGS DESPITE GEOGRAPHICAL LOCATION OF THE HOTELS

This study's design has been conceived to allow for generalization of the findings. Indeed we have focused on hotels located in five different countries (United States, United Kingdom, Italy, Spain, Russia) covering three different continents (America, Europe, Asia). The reason why the countries were chosen are as follows: 1) all of them feature among the Top Ten countries in terms of both international tourist arrivals and international tourist receipts [12]; 2) all of them include cities that individually attract a considerable amount of tourists annually [66]; 3) all of them record a relevant amount of room nights annually (see the reports from the individual national statistical offices); 4) the cultural backgrounds of both hotel customers and hotel providers in those countries are very diverse and the countries hosting the hotels themselves differ significantly across the Hofstede dimensions, allowing therefore to generalize the findings not only across customers but also across hotel service providers.

B. DATA AND METHODOLOGY

In the ensuing three subsections, we elucidate the main features of our research methodology with a focus on data collection, variables operationalization and data analysis.

1) Data collection

As far as data retrieval, understanding, cleaning, processing and analysis are concerned, we adopted big data science techniques [67]–[69]. Data was retrieved from two different sources: the Online Travel Agency Expedia.com and the "Hofstede Center" (https://www.hofstede-insights.com/product/compare-countries/). The former source was used to retrieve online review data; the latter one was used to collect data related to the cultural dimensions and related variables. As far as online hotel reviews are concerned, they were extracted from the OTA Expedia.com. The platform, founded in 1996, is now part of the Expedia Group which is a leading international company owning 200 travel booking sites (in addition to Expedia also Hotels.com, Egencia, etc.) and has a presence in more than 75 countries worldwide and a coverage of a very high number of properties, attracting simultaneously 75 million monthly flight

shoppers. Expedia was chosen because it hosts the second largest share of certified online reviews worldwide after Booking.com [26] and it was preferred over TripAdvisor exactly because Expedia ORs are certified while TripAdvisor ones are not: indeed Expedia enables hotel customers to post reviews only if they have actually stayed at a hotel. Moreover, while TripAdvisor is certainly the most popular and perhaps the largest ORs website for travel [70], it has been shown that a large number of ORs posted on TripAdvisor are fake, promotional and/or written even by individuals who have never consumed a specific service [71]. We collected data by means of a tool specifically developed for this work. The tool consists of three modules acting on a pipeline: (i) an extractor module, which performs data queries on the Expedia platform using the public Hotel Reviews and Hotel Search public APIs, which are available as REpresentational State Transfer (REST) web services; (ii) a parser module (written in Python language), which processes the results of each query (in JSON data format), and computes per-review statistics; (iii) an analyser module, which computes aggregated metrics over a time window defined by the user. The dataset was collected between April and December 2016. The software tool was used to retrieve all available information on a selected number of British, Italian, Russian, Spanish, US hotels marketed through Expedia. Instead of creating relational databases, data was collected through MongoDB, a NoSQL database, document-oriented, and therefore able to deal with heterogeneous and complex data in the within the Hadoop framework. MongoDB works on databases organized in collections and the collections include lists of documents where each document is a collection of fields. In a Relational Data Base Management System (RDBMS) collections correspond approximately with tables, documents with rows and fields to the column of a row. Overall, we gathered 500,175 ORs covering 31,580 hotels. At the company level, we retrieved information including the name of the hotel, its address, the hotel class, the overall rating as well as the rating for convenience of location, neighborhood, value for money. At the individual online review level, we collected all the data related to each individual review encompassing the timestamp, declared name of the reviewer, her/his country of origin/residence, the overall rating of the hotel, the ratings for each of the eight hotel service attributes (cleanliness, service and staff, room comfort, hotel condition, convenience of location, neighborhood, value for money, room quality). In order to enrich the above data, we obtained the measures of the six cultural dimensions (power distance, individualism, masculinity, uncertainty avoidance, long-term orientation, indulgence) based on data from the Hofstede center. As the data related to the country of origin and residence are a "free field" in several cases the places/locations were written with spelling errors and sometimes only the city of origin was mentioned. Therefore, we had to extend the dataset following these steps: 1) data cleansing ; 2) automatic match of the "free field" with the country of origin (ISO 3166-1 alpha-3 codes) of the online reviewer; 3) systematic manual validation and further matching. This laborious activity was required to improve the quality and quantity of the data related to the country of origin, which is necessary to relate the Hofstede's dimension to each OR. The automatic match was carried out by means of a comprehensive dictionary of countries, states, counties, provinces and cities. The manual matching helped to identify and use recurring patterns in nicknames of cities (e.g., 'Big Apple' for New York and 'Big Easy' for New Orleans). The final pre-processed dataset includes 499,991 reviews written by reviewers from 96 countries for 31,580 hotels. However, the Hofstede framework displays a significant number of missing values for the dimensions of LTO and IND for several countries. Thus, we created a subsample of the former dataset where missing values of the Hofstede dimensions led us to delete observation thus generating a second dataset including 495,633 reviews. We used the first dataset (499,991 reviews) to test a full model encompassing the six Hofstede dimensions and a second dataset (495,633 reviews), which is a subset of the first dataset, to test a partial model including only the four Hofstede dimensions for which there are no missing values. Overall, the samples deployed not only consist of international tourists visiting five countries spread across three continents, but it also includes significant cultural diversity, which constitutes a necessary condition for this analysis.

2) Variables operationalization

Our dependent variable is the overall Expedia.com rating of the hotel which varies from 1.0 to 5.0. Among the independent variables, we included first the four critical hotel attributes received from the literature (cleanliness, service and staff, room comfort, hotel condition) and secondly the Hofstede dimensions discussed in our literature review: power distance (PD), individualism (IDV), masculinity (MAS), uncertainty avoidance (UA), long-term orientation (LTO), indulgence (IND). In addition, we included also a further variable, such as the length of the text. The description of the variables are illustrated in the Table 1; while descriptive statistics for the relevant variables are reported in Table 2.

All of the variables display levels of skewness and kurtosis compatible with normal distributions and therefore can be deployed in multivariate regression analyses. The variable correlation matrix is included in Table 3.

3) Data analysis

Before running our analysis we focused on hotel attributes (both critical and not) and carried out a multivariate ordinary least squares (OLS) regressions to test our eleven hypotheses. OLS regression is suitable when variables display multivariate normality, like in the case analysed. The dependent variable, namely the rating (Overall satisfaction rating), was regressed on the critical hotel services attributes (Cleanliness, Hotel Condition, Room Comfort, Service and Staff) and the cultural dimensions (PD, IDV, MAS, UA, LTO, IND). Moreover, we considered also the length of the review as a further variable.

TABLE 1. Variables description, where the definition indicated with (*) are borrowed from https://www.hofstede-insights.com/models/national-culture/

Dimensions	Variable	Description
Online review features	Overall satisfaction	The overall rating of a hotel (from 1 to 5)
	Length of review	The length of the textual comment in number of characters
Critical services attributes	Cleanliness	The rating of the cleanliness of a hotel (from 1 to 5)
	Service and Staff	The rating of the service and staff of a hotel (from 1 to 5)
	Room Comfort	The rating of the room comfort of a hotel (from 1 to 5)
	Hotel Condition	The rating of the condition of a hotel (from 1 to 5)
Cultural dimensions	Power Distance (PD)	Degree to which the less powerful members of a society accept and expect that power is distributed unequally*
	Individualism (IDV)	Preference for a loosely-knit social framework in which individuals are expected to take care of only themselves and their immediate families*
	Masculinity (MAS)	Preference in society for achievement, heroism, assertiveness, and material rewards for success*
	Uncertainty Avoidance (UA)	Degree to which the members of a society feel uncomfortable with uncertainty and ambiguity*
	Long Term Orientation (LTO)	The degree to which the member of a society do not prefer to maintain time-honoured traditions and norms while viewing societal change without suspicion*
	Indulgence (IND)	Preference for a society that allows relatively free gratification of basic and natural human drives related to enjoying life and having fun*

TABLE 2. Descriptive statistics

	Mean	SD	Min	Max
Overall satisfaction rating	3.95	1.16	1	5
Length of review	351.99	305.73	0	2682
Cleanliness	4.2	1.12	1	5
Hotel Condition	3.88	1.18	1	5
Room Comfort	3.97	1.19	1	5
Service and Staff	4.23	1.08	1	5
PD	39.23	7.74	13	100
IDV	86.97	11.99	6	91
MAS	62.12	5.71	5	100
UA	44.09	9.87	8	100
LTO	35.68	12.85	4	100
IND	67.59	6.32	0	100

IV. ANALYSIS AND FINDINGS

The results of the analyses show that most of the variance of the rating of a hotel is explained, in order of relevance, by the critical hotel service attributes: hotel condition, room comfort, service and staff and cleanliness. The aforementioned attributes explain 83.95% of the variance for the full sample (including the six Hofstede dimensions) and 83.98% of the subsample without the Hofstede dimensions of long-term orientation and indulgence. Therefore, H0 is supported. The results of the multivariate regression analysis on the full sample and its subsample (with no missing values for the Hofstede dimensions of long-term orientation and indulgence) are illustrated in Table 4 and 5 respectively. In both the full sample and the subsample the coefficients for the critical hotel service attributes (hotel condition, room comfort, service and staff, cleanliness) are significantly positive (with $p<0.001$). These findings suggest that all of the critical hotel service attributes are positively related to the overall satisfaction online rating thus lending support to existing research conducted in offline settings [28] and content analysis conducted on a very small sample of 373 TripAdvisor online reviews [27]. This is partially reflected by the magnitude of the coefficients: 0.3741 for hotel condition, 0.2634 for service and staff, 0.2592 for room comfort and 0.1213 for cleanliness in the full sample (see Table 4). In the subsample (see Table 5) the magnitude of the coefficients is slightly different but their relative importance is consistent with that found in the full sample: 0.3746 for hotel condition, 0.2623 for service and staff, 0.2600 for room comfort and 0.1200 for cleanliness. Overall, hypotheses 1.a, 1.b, 1.c, and 1.d cannot be rejected across the two samples. Regarding the effects of cultural factors on the overall satisfaction ratings, we find differentiated results. The reviewers from countries characterized by higher power distance give lower online overall satisfaction ratings. This is consistent with surveys conducted in offline contexts and shows that high power distance individuals have high service expectations and are more critical in their evaluation process (see [52]). Therefore, H2.a is supported ($p<0.001$) in both the full sample and the subsample. This finding contributes to extend and generalize our findings across contexts such as the US and UK (which display a high level of individualism, and low levels of power distance and uncertainty avoidance) to contexts such as Italy, Spain, and Russia (which display lower levels of individualism, and higher levels of power distance and uncertainty avoidance). The online reviewers from cultures characterized by higher individualism tend to give lower online ratings. This is consistent with the service management literature indicating that customers from individualistic cultures tend to be more demanding and have high service expectations (e.g.., [50], [53], [55]). Consequently, they give low valuation scores to hospitality companies. Accordingly, H2.b is supported ($p<0.001$) in both the full sample and the subsample. Interestingly, the degree of masculinity is not associated significantly to the overall customer rating in the sub sample (indeed the coefficient is not statistically significant in both cases). Therefore, H2.c is not supported neither in the full model nor in the subsample. The results indicate that online customers from countries with higher levels of uncertainty avoidance provide lower online ratings. Indeed consumers that are risk adverse due to their cultural background could more easily share negative opinions when the service does not meet expectations [59]. Thus, H2.d is supported for both the full sample and the subsample. Addressing our attention solely to the subsample, we find that consumers from cultures characterized by high levels of long-term orientation and

TABLE 3. The variable correlation matrix

		1	2	3	4	5	6	7	8	9	10	11	12
1	Overall Rating	1											
2	Cleanliness	0.8	1										
3	Service & Staff	0.76	0.67	1									
4	Room Comfort	0.83	0.78	0.65	1								
5	Hotel condition	0.86	0.8	0.68	0.8	1							
6	PD	-0.04	-0.04	-0.04	-0.03	-0.03	1						
7	IDV	-0.01	-0.007	-0.007	0.003	-0.002	-0.72	1					
8	MAS	0.03	0.04	-0.05	0.03	0.03	-0.36	0.49	1				
9	UA	-0.05	-0.06	-0.05	-0.04	-0.04	0.61	-0.58	-0.55	1			
10	LTO	0.1	0.11	0.08	0.06	0.07	-0.13	-0.21	0.21	-0.35	1		
11	IND	0.01	0.01	0.01	0.01	0.01	-0.53	0.5	0.32	-0.4	-0.23	1	
12	Length	-0.17	-0.11	-0.16	-0.15	-0.14	-0.05	0.03	0.04	-0.07	0.07	0.016	1

TABLE 4. Effects of critical service attributes and the four Hofstede dimensions (original Hofstede framework) on online customer satisfaction ratings (full sample). (Notes: *p<0.10; **p<0.05; ***p<0.01; ****p<0.001; N = 499,991)

Independent Variable	Coefficient	Std Error
Cleanliness	0.1213****	0.0011
Hotel Condition	0.3741****	0.001
Room Comfort	0.2592****	0.001
Service and Staff	0.2634****	0.0008
PD	-0.0019****	0.0001
IDV	-0.0028****	0.00008
MAS	0.0009	0.0001
UA	-0.0017****	0.00009
Review length	-0.00008****	0
Constant	0.2106****	0.0144
R2	0.8403	
Adjusted R2	0.8403	

TABLE 5. Effects of critical service attributes and the six Hofstede dimensions on online customer satisfaction ratings (subsample without missing values for uncertainty avoidance and long-term orientation. (Notes: *p<0.10; **p<0.05; ***p<0.01; ****p<0.001; N = 495,633)

Independent Variable	Coefficient	Std Error
Cleanliness	0.1200****	0.0011
Hotel Condition	0.3746****	0.001
Room Comfort	0.2600****	0.001
Service and Staff	0.2623****	0.0008
PD	-0.0015****	0.0001
IDV	-0.0016****	0.0001
MAS	0.00006	0.0001
UA	-0.0005****	0.0001
LTO	0.0016****	0.00007
IND	0.0004****	0.0001
Review length	-0.00009****	0
Constant	0.0155	0.0235
R2	0.8409	
Adjusted R2	0.8409	

high levels of indulgence give higher ratings: indeed both the coefficients are statistically significant at p<0.001. Therefore, hypotheses H2.e and H2.f are supported. As far as the length of the written text is concerned, it exerts a negative impact on the overall satisfaction consistently with previous literature [72], [73]) in both the full sample and the subsample.

V. CONTRIBUTIONS AND IMPLICATIONS

By adopting a big data science approach [69] to hospitality and tourism management [6], and leveraging on a large volume of data stemming from user generated content (namely hotel reviews) covering hotels located in 5 countries scattered across three continents, this study has tried to capture conjointly the effects of critical service attributes and cultural factors on online rating behaviors. We therefore applied simultaneously the typology proposed by [28] and the influential cultural framework developed by Hofstede ([9], [10]) to understand if and to what extent critical service attributes (firm-related and service-related variables) and the national culture (reviewer-related variables) affect individual reviewers behaviors. Three set of key findings emerge from the analysis. First, the critical service attributes identified in offline settings by [28] explain most of the variance of online customer satisfaction with hotels. Second, all of the critical service attributes affect positively online customer satisfaction. Third, cultural factors related to the reviewer (i.e., her/his country of origin) play a role. More specifically, reviewers from countries characterized by high power distance give low ratings in online settings. This holds in five different countries among which there are cultures (Italian, Russian, Spanish) that are characterized by lower levels of individualism and higher levels of power distance than the United States and the United Kingdom. Online hotel customers from highly individualistic countries appear to give low hotel ratings. This result appears consistent with previous literature revealing that customers from individualistic cultures tend to have higher levels of expectations compared to consumers embedded in collectivistic cultures ([50], [53]–[55]). Online hotel reviewers from highly masculine countries do not seem to give low hotel ratings as we expected. This finding seems to be in contrast with traditional survey-based literature illustrating that consumers from masculine societies are more critical towards service failures [57]. However, the effect of this dimension might be amplified or moderated by the actual gender that might be worth considering for future analysis as women tend to be more generous in online ratings. Online customers from countries with high levels of uncertainty avoidance tend to give low ratings; this finding is consistent with the

relevant literature pointing out that risk adverse consumers have low perceptions of service quality because they have low tolerance for ambiguity [51] and face difficulties in dealing effectively with uncertainty [51], [53], [54]. Moreover, consumers from cultures characterized by high long-term orientation and high indulgence give high online ratings. This is compatible with recent literature emphasizing that long-term oriented individuals are more patient and more loyal [62], [63] and that individuals from cultures characterized by a high level of indulgence develop a more positive attitude and are more likely to remember positive emotions [64] in addition to being more prone to using user generated content [65].

Overall, our findings lend support to both the classification of service attributes provided by [28] and largely support service management literature developed in offline setting using the Hofstede framework of cultural dimensions [10]. In so doing, this study leverages on big data to shed lights on the interplay between firm related attributes and consumers' characteristics in determining online consumer behaviors.

A. THEORETICAL CONTRIBUTIONS

This study contributes to management literature within hospitality and tourism contexts on a number of levels. First, to the best of our knowledge, this is the first research addressing simultaneously how critical service attributes and cultural background of hotel service consumers can conjointly influence consumer online satisfaction and rating behavior. Second, this study is among the few ones looking at how service attributes have been operationalized in marketing management literature [28] in offline settings often through customer surveys relying on small samples and apply the received cluster of attributes to online settings allowing to gather large volumes of data by means of big data retrieval techniques. Third, the analysis is distinctive because innovatively investigates how cultural differences across consumers can affect their online reviewing behavior based on a consolidated framework within social sciences: the Hofstede framework of national culture ([9], [10]). Accordingly, this study complements existing research in the hospitality and tourism management field that has examined only one of the cultural dimensions in online settings. The big data analysis carried out has allowed gathering a very large and culturally diverse set of customers whose analysis can shed light on the role played by national culture [9] on customers' evaluations of a hospitality service with a broad scope and allowing for generalizations, given that customers' perceptions of services have been retrieved from hotels based in five countries scattered across three continents. Interestingly enough the study finds that cultural background (translated into cultural dimensions) makes a difference in consumer behavior within online settings. This contribution certainly adds to previous knowledge on the relevance of cultural aspects in marketing management literature [74]. Fourth, this study provides certainly a methodological contribution as it is among the first attempts to generate managerial insights based on a large sample of data including almost half a million digital records. While the relevant literature has used mostly samples totaling less than 100,000 records, we analyse the effect of critical service attributes and the Hofstede cultural dimensions using a large dataset of respondents that cumulatively wrote half a million ORs scattered in 5 different countries covering three continents. Fifth, the study contributes to the eWOM research line within marketing management [1] by assessing conjointly how service characteristics and individual personal backgrounds can affect the way consumers produce eWOM. This is a relevant and distinctive contribution to the research line pertaining to the antecedents of eWOM [5]. Sixth, this study contributes to the thin yet emerging hospitality and tourism management research line dealing with big data [6] showing not only how big data analytical tools and techniques can be deployed to generate valuable market knowledge and intelligence [31] but also enriching the debate within management and marketing science about the relevance of big data studies as a scholarly phenomenon [33] that can help test and establish theories that were developed in offline settings by means of data-driven knowledge [75]. Last, this study is among the few ones whose research design draws on an in-depth and critical understanding of the sources, quality and nature of big data stemming from online review platforms. Certainly there are advantages associated with the use of big data, but also challenges [34], [36] that relate to the selection of the right sources of data based on their quality. For instance, in hospitality and tourism management studies most of research have a-critically deployed ORs of the online travel community TripAdvisor [26]. However, TripAdvisor's ORs are not certified: in other terms, anyone could write a review about a hotel even if s/he has never been a guest in the hotel. This is the reason why a few scholars have questioned the veracity (another "v" to add to the 3V identified by [21] of the data. By focusing on a specific category of data sources, namely ORs from OTAs such as Expedia or Booking, we make an important step forward in terms of accuracy of big data analyses in hospitality and tourism management. Deploying ORs from OTAs guarantees that the data used is overall authentic and reliable as they relate to real customer that have actually stayed at the hotel they review [6], thus decreasing the likelihood of collecting fake reviews.

B. PRACTICAL IMPLICATIONS

A multi-faceted set of practical implications can be derived from this study, encompassing implications for managers, developers and managers of online review platforms, and customers of hotel services. As far as hospitality managers are concerned, first they could progressively juxtapose traditional survey-based methods based on small data to more advanced tools and techniques leveraging high volume of data sourced from online review platforms [67], [69]. A balanced evaluation of customer behavior in today's digital world would require triangulating customer perception data collected through traditional small data surveys with behav-

ioral data inferred from big data analytics [6], [68]. While this could be generally conducive to better managerial decision making [32], it might pose challenges for most of the hotel properties located in several geographical areas that are Small and Medium Enterprises (SMEs), with limited financial and human resources to invest in big data analyses [76]. Secondly, hotel managers might deploy the findings of this study to understand the conjoint impact of perceived service attributes with guests' cultural background. For instance, they may focus on the feedback provided by those guests from cultures exhibiting extreme values across each and every of the Hofstede cultural dimensions [10]. In so doing, they might segment and target more effectively potential customers based on their cultural background and tailor their services (and related critical attributes) to the characteristics (i.e., cultural background) of their customers [77]. The way to operationalize this aspect might be to track the country of residence of the potential customer based on geo-localization of IP addresses of the booking device used. Regarding developers and managers of online review platforms, the results of this research are particularly relevant for both independent online review platforms such as TripAdvisor and OTAs such as Expedia and Booking. As far as the former one is concerned, it could start developing algorithms addressing readers' attention to those reviews that have been written by reviewers of the same country or of a similar culture that ultimately could be more useful or helpful for them [78]. As far as the latter ones are concerned, these findings might help OTAs to optimize the algorithms they embed into their recommender systems to increase reservations on their websites and ultimately revenues and firm value [30]. At the moment of writing, these websites allow their users to filter reviews based on the language and country of residence of the reviewer. Nonetheless, recommender systems might be empowered by suggesting reviews from countries that are similar across one or more of the cultural dimensions of the Hofstede model [10]. Third, developers of OR platforms could build more comprehensive algorithms to enhance the helpfulness of online reviews for instance by allowing readers of ORs to filter the helpful reviews based on country of origin and setting as a default the helpful reviews written by guests sharing analogous cultural characteristics. As far as online customers are concerned, it is recommended that if they use independent online review platforms such as TripAdvisor, they may focus first on those reviews written by online reviewers from the same country or displaying a similar cultural background (in terms of dimensions such as individualism, masculinity, uncertainty avoidance, power distance, long-term orientation, indulgence). Secondly, if they use third party commercial websites such as Expedia or Booking, they could put some efforts in evaluating previous reviews and ratings controlling by a sufficiently large number of reviews written by reviewers from a similar cultural background.

VI. CONCLUSIONS AND LIMITATIONS

This study contributes to different research streams within management and marketing science and, in particular, to online consumer behavior and eWOM. By leveraging Big data analytical tools and techniques [67], [69] we use a large sample of almost half a million online consumer reviews related to hotels located in 5 countries scattered across three continents, to understand how service attributes and reviewers' cultural background affect online customer satisfaction with hotels. The findings of the study indicate that critical service attributes and reviewers' cultural background play a crucial role in affecting online customer satisfaction. We therefore generalize previous scattered empirical evidence of well rooted social sciences theoretical frameworks into online settings. More specifically, we find that all of the critical services attributes impact positively the overall customer satisfaction. As far as cultural dimensions are concerned, power distance, individualism and uncertainty avoidance negatively affect overall online satisfaction, while long-term orientation and indulgence positively affect online satisfaction. Masculinity seem not to play a significant role. We also observe that reviews' text length exerts a negative impact on online ratings. In general, we extend previous anecdotic evidence that the cultural background of consumers can affect consumers' attitudes and perceptions [74]. Interestingly, the use of big data from ORs created by consumers visiting hotels in multiple countries and continents allows us generalizing our findings in the online consumer behavior and eWOM fields [26]. Our study provides a robust and multi-faceted set of implications for hospitality managers, developers and managers of online review platforms, and hotel services' customers. This study is not without limitations. First, it was not possible to collect additional reviewer-level variables (such as age and gender) given that they are "free fields"; certainly, they might be included in future model specifications as moderators and/or mediators to understand their impact on the explanatory variables used to explain online customer satisfaction. Secondly, it was not possible to control by the type of hotel, i.e., independent vs. chained hotels due to practical difficulties in collecting this type of data. This data collection will certainly be part of our future efforts of enriching the already extensive dataset. Third, while this is one of the very few studies conducted on an OTA other than TripAdvisor (see also [26]), it would be worthwhile to compare findings across different platforms and to generate additional insights on online customer behavior. Fourth, we did not control by hotel category and by weather or not a hotel was part of a chain and this will be undertaken in future research. Last, using different platforms might shed light on how the layout and functionalities of the platform could generate different user behaviors translating in differentiated online reviewing behaviors. Certainly this would be an important step to further consolidate the idea that big data analysis and big data analytics are beyond a mere hype [79] but can make a difference for firms, managers and entrepreneurs in understanding consumer behavior.

REFERENCES

[1] T. Hennig-Thurau, K. Gwinner, W. Walsh, and D. Gremler, "Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet?" Journal of Interactive Marketing, vol. 18, no. 1, pp. 38–52, 2004.

[2] C. Dellarocas, X. Zhang, and N. F. Awad, "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," Journal of Interactive Marketing, vol. 21, pp. 23–45, 2007.

[3] The Nielsen Company. (2015) Global trust in advertising. winning strategies for an evolving media landscape. (accessed 17.05.2018). [Online]. Available: http://www.nielsen.com/us/en/insights/reports/2015/global-trust-in-advertising-2015.html

[4] ——. (2016) Global connected commerce (accessed 27.06.2018). [Online]. Available: http://www.nielsen.com/content/dam/corporate/us/en/reports-downloads/2016-reports/connected-commerce-report-jan-2016.pdf

[5] A. Cantallops and F. Salvi, "New consumer behavior: A review of research on ewom and hotels," International Journal of Hospitality Management, vol. 36, pp. 41–51, 2014.

[6] M. Mariani, R. Baggio, M. Fuchs, and W. Höpken, "Business intelligence and big data in hospitality and tourism: A systematic literature review," International Journal of Contemporary Hospitality Management ~~(in press)~~, vol. 30, no. ~~10~~ 2, 2018.

[7] Y. Yang, S. Park, and X. Hu, "Electronic word of mouth and hotel performance: a meta-analysis," Tourism Management, vol. 67, pp. 248–260, 2018.

[8] B. Sparks and V. Browning, "The impact of online reviews on hotel booking intentions and perception of trust," Tourism Management, vol. 32, no. 6, pp. 1310–1323, 2011.

[9] G. Hofstede, Culture's consequences: international differences in work-related values. Beverly Hills, CA: Sage Publications, 1980.

[10] ——, Culture's consequences: comparing values, behaviors, institutions, and organizations across nations (2nd ed.). Thousand Oaks, CA: Sage Publications, 2001.

[11] The World Travel and Tourism Council. (2018) World Economic Impact for the travel and tourism industry (accessed 25.07.2018). [Online]. Available: https://www.wttc.org/-/media/files/reports/economic-impact-research/regions-2018/world2018.pdf

[12] The World Tourism Organization. (2018) 2017 International Tourism Results: the highest in seven years. Press Release No.: 18003, 15 Jan 2018 (accessed 27.07.2018). [Online]. Available: http://media.unwto.org/press-release/2018-01-15/2017-international-tourism-results-highest-seven-years

[13] D. S. Evans and R. Schmalensee, "Matchmakers: the new economics of multisided platforms," Journal of Service Research, 2016.

[14] J. Rochet and J. Tirole, "Platform competition in two-sided markets," Journal of the European Economic Association, vol. 1, pp. 990–1029, 2003.

[15] T. Eisenmann, G. Parker, and M. Van Alstyne, "Strategies for two-sided markets," Harvard Business Review, October 2006.

[16] G. Parker and M. Van Alstyne, "Two-sided network effects: A theory of information product design," Management Science, vol. 51, no. 10, pp. 1494–1504, 2005.

[17] Google. (2014) The 2014 traveler's road to decision (accessed 28.07.2018). [Online]. Available: http://www.chessbase.com/newsdetail.asp?newsid=3595

[18] Rauch, M. (2017) Hotel and mobile boost ota online bookings, phocuswright. (accessed 22.07.2018). [Online]. Available: https://www.phocuswright.com/Travel-Research/Research-Updates/2017/Hotel-and-Mobile-Boost-OTA-Online-Bookings

[19] R. Filieri and F. McLeay, "E-WOM and accommodation. an analysis of the factors that influence travelers' adoption of information from online reviews," Journal of Travel Research, vol. 53, no. 1, pp. 44–57, 2014.

[20] M. Chen, S. Mao, and Y. Liu, "Big data: A survey, mobile networks and applications," Decision Support Systems, vol. 19, no. 2, pp. 171–209, 2014.

[21] Laney, D. (2001) 3D data management: Controlling data volume, velocity and variety, META group research note, 6. [Online]. Available: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

[22] T. White, Hadoop: The Definitive Guide, 4th Edition, Storage and Analysis at Internet Scale. O'Reilly Media, Sebastopol, CA, 2015.

[23] K. Chodorow and M. Dirolf, MongoDB: The Definitive Guide, 1st ed. O'Reilly Media, Inc., 2010.

[24] Y. Liu, T. Teichert, M. Rossi, H. Li, and F. Hu, "Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews," Tourism Management, vol. 59, pp. 554–563, 2017.

[25] Z. Xiang, Q. Du, Y. Ma, and W. Fan, "A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism," Tourism Management, vol. 58, pp. 51–65, 2017.

[26] M. M. Mariani and M. Borghi, "Effects of the booking.com rating system: bringing hotel class into the picture," Tourism Management, vol. 66, pp. 47–52, 2018.

[27] W. Lu and S. Stepchenkova, "Ecotourism experiences reported online: Classification of satisfaction attributes," Tourism management, vol. 33, no. 3, pp. 702–712, 2012.

[28] E. Cadotte and N. Turgeon, "Dissatisfiers and satisfiers: suggestions from consumer complaints and compliments," Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, vol. 1, pp. 74–79, 1988.

[29] C. Luo, J. Wu, Y. Shi, and Y. Xu, "The effects of individualism-collectivism cultural orientation on eWOM information," International Journal of Information Management, vol. 34, no. 4, pp. 446–456, 2014.

[30] P. Verhoef, E. Kooge, and N. Walk, Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. Routledge, London, 2016.

[31] R. M. Chang, R. J. Kauffman, and Y. Kwon, "Understanding the paradigm shift to computational social science in the presence of big data," Decision Support Systems, vol. 63, pp. 67–80, 2014.

[32] D. J. Power, "Using 'big data' for analytics and decision support," Journal of Decision Systems, vol. 23, no. 2, pp. 222–228, 2014.

[33] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," Information, Communication and Society, vol. 15, no. 5, pp. 662–679, 2012.

[34] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," National Science Review, vol. 1, no. 2, pp. 293–314, 2014.

[35] D. A. McFarland and H. R. McFarland, "Big data and the danger of being precisely inaccurate," Big Data and Society, vol. 2, no. 2, pp. 1–4, 2015.

[36] G. Gerard, E. Osinga, D. Lavie, and B. Scott, "Big data and data science methods for management research," Academy of Management Journal, vol. 59, no. 5, pp. 1493–1507, 2016.

[37] F. Herzberg, B. Mausner, and B. Snyderman, The motivation to work. Wiley, New York, NY, 1959.

[38] T. Parsons, Introduction, to Part IV, Culture and the Social System in Talcott Parsons et al (eds). Theories of Society, New York: Free Press, 1961.

[39] E. Hall, Beyond Culture. Garden City, NJ: Doubleday, 1976.

[40] A. Pizam, Managing Cross-cultural Hospitality Enterprises. The International Hospitality Industry: Organizational and Operational Issues. John Wiley, New York, 1993.

[41] A. Mattila, "The role of culture and purchase motivation in service encounter evaluations," Journal of Services Marketing, vol. 13, no. 4-5, pp. 376–389, 1999.

[42] T. Levitt, "The globalization of markets," Harvard Business Review, vol. 12, no. 93, pp. 58–85, May 1983.

[43] T. Hout, M. Porter, and E. Rudden, "How global companies win out," Harvard Business Review, vol. 60, no. 5, pp. 98–108, September/October 1982.

[44] W. Keegan and M. Green, Global Marketing. Prentice-Hall, Inc., Upper Saddle River, NJ, 2000.

[45] P. Kotler, "Global standardization courting danger," Journal of Consumer Marketing, vol. 3, no. 2, pp. 13–16, 1986.

[46] N. Holden, "Why marketers need a new concept of culture for the global knowledge economy," International Marketing Review, vol. 21, no. 6, pp. 563–572, 2004.

[47] G. Hofstede, G. Hofstede, and M. Minkov, Cultures and Organizations: Software of the Mind (3rd ed.). McGraw-Hill, USA, 2010.

[48] A. Mattila, "The essentials of scholarship: A reply to Geert Hofstede," Human Relations, vol. 55, no. 11, pp. 1363–1372, 2002.

[49] W. T. Ngai, C. S. Heung, Y. H. Wong, and K. Y. Chan, "Consumer complaint behavior of asians and non-asians about hotel services: An empirical analysis," European Journal of Marketing, vol. 41, no. 11-12, pp. 1375–1391, 2007.

[50] O. Furrer, B. Liu, and D. Sudharshan, "The relationships between culture and service quality perceptions: basis for cross-cultural market segmentation and resource allocation," Journal of Service Research, vol. 2, no. 4, pp. 355–371, 2000.

[51] R. Ladhari, F. Pons, G. Bressolles, and M. Zins, "Culture and personal values: How they in?uence perceived service quality," Journal of Business Research, vol. 64, no. 9, pp. 951–957, 2011.
[52] A. Mattila, "The impact of culture and gender on customer evaluations of service encounters," Journal of Hospitality and Tourism Research, vol. 24, no. 2, pp. 263–273, 2000.
[53] P. Sharma, C. I.S.N., and S. Luk, "Exploring the role of indŰcol as a moderator in the comprehensive service evaluation model," Journal of International Consumer Marketing, vol. 24, no. 1-2, pp. 129–142, 2012.
[54] N. Donthu and B. Yoo, "Cultural influences on service quality expectations," Journal of Service Research, vol. 1, no. 2, pp. 178–186, 1998.
[55] M. Laroche, L. C. Ueltschy, A. Shuzo, and M. Cleveland, "Service quality perceptions and customer satisfaction: Evaluating the role of culture," Journal of International Marketing, vol. 12, no. 93, pp. 58–85, 2004.
[56] J. Huang, C. Huang, and S. Wu, "National character and response to unsatisfactory hotel service," International Journal of Hospitality Management, vol. 15, no. 3, pp. 229–243, 1996.
[57] J. Crotts and R. Erdmann, "Does national culture influence consumers' evaluation of travel services? A test of Hofstede's model of cross-cultural differences," Managing Service Quality: An International Journal, vol. 10, no. 6, pp. 410–419, 2000.
[58] E. N. Torres, X. Fu, and X. Lehto, "Examining key drivers of customer delight in a hotel experience: A cross-cultural perspective," International Journal of Hospitality Management, vol. 36, pp. 255–262, 2014.
[59] H. Fang, J. Zhang, Y. Bao, and Q. Zhu, "Towards effective online review systems in the chinese context: A cross-cultural empirical study," Electronic Commerce Research and Applications, vol. 12, no. 3, pp. 208–220, 2013.
[60] S. Ryu and C. W. M., "Long-term orientation as a determinant of relationship quality between channel members," International Business and Economics Research Journal, vol. 1, no. 9, pp. 1–9, 2009.
[61] J. Meng and M. Venkatapparao, "Cultural influences on web service quality perceptions of e-retailing consumers," Journal of Marketing Channels, vol. 18, no. 4, pp. 303–326, 2011.
[62] B. Bartikowski, G. Walsh, and S. E. Beatty, "Culture and age as moderators in the corporate reputation and loyalty relationship," Journal of Business Research, vol. 64, no. 9, pp. 966–972, 2011.
[63] X. R. Li, C. Lai, R. Harrill, S. Kline, and L. Wang, "When east meets west: An exploratory study on chinese outbound tourists' travel expectations," Tourism Management, vol. 32, no. 4, pp. 741–749, 2011.
[64] J. Park, Y. M. Baek, and M. Cha, "Cross-cultural comparison of nonverbal cues in emoticons on twitter: Evidence from big data analysis," Journal of Communication, Special Issue on Big Data in Communication Research, vol. 64, no. 2, pp. 333–354, 2014.
[65] R. L. Stump and W. Gong, "Social networking sites: An exploration of the effect of national cultural dimensions on country adoption rates and usage patterns," International Journal of Electronic Business, vol. 13, no. 2-3, pp. 117–42, 2017.
[66] E. International. (2017) Top 100 city destinations ranking. (last accessed: 23.07.2018). [Online]. Available: http://go.euromonitor.com/wtm2017-top-100-cities-destinations.html
[67] T. Erl, W. Khattak, and P. Buhler, "Big data fundamentals Ű concepts, drivers and techniques," Journal of Service Research, 2015.
[68] J. E. Liebowitz, Big data and business analytics. CRC Press, Boca Raton, FL, 2013.
[69] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kauffman, Cambridge (MA), 2016.
[70] H. A. Lee, R. Law, and J. Murphy, "Helpful reviewers in tripadvisor, an online travel community," Journal of Travel and Tourism Marketing, vol. 28, pp. 675–688, 2011.
[71] R. Filieri, "What makes an online consumer review trustworthy?" Annals of Tourism Research, vol. 58, pp. 46–64, 2016.
[72] P. Duverger, "Curvilinear effects of user-generated content on hotels' market share: a dynamic panel-data analysis," Journal of Travel Research, vol. 52, no. 4, pp. 465–478, 2013.
[73] M. M. Mariani, M. Di Felice, , and M. Mura, "Facebook as a destination marketing tool: Evidence from italian regional destination management organizations," Tourism Management, vol. 54, pp. 321–343, 2016.
[74] M. De Mooij and G. Hofstede, "Cross-cultural consumer behavior: A review of research findings," Journal of International Consumer Marketing, vol. 23, pp. 181–192, 2011.
[75] H. Ekbia, M. Mattioli, A. G., A. Ghazinejad, T. Bowman, V. Suri, A. Tsou, S. Weingart, and C. Sugimoto, "Big data, bigger dilemmas: A critical review," Journal of the Association of Information Science and Technology, vol. 66, no. 8, pp. 1523–1545, 2015.
[76] S. Coleman, R. Göb, G. Manco, A. Pievatolo, X. Tort-Martorell, and M. Seabra Reis, "How can smes benefit from big data? challenes and path forward," Quality and Reliabiltiy Engineering Interantional, vol. 32, pp. 2151–2164, 2016.
[77] S. Fan, R. Y. Lau, and J. L. Zhao, "Demystifying big data analytics for business intelligence through the lens of marketing mix," Big Data Research, vol. 2, no. 1, pp. 28 – 32, 2015, special Issue on Computation, Business, and Health Science.
[78] Z. Liu and S. Park, "What makes a useful online review? implication for travel product websites," Tourism Management, vol. 47, pp. 140–151, 2015.
[79] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," International Journal of Information Management, vol. 35, no. 2, pp. 137–144, 2015.

MARCELLO MARIANI is Professor of Entrepreneurship and Management at the Henley Business School, University of Reading and member of the Henley Center for Entrepreneurship, the Academy of Management and the European Institute for Advanced Studies in Management as well as Senior Fellow of the Higher Education Academy (UK). After earning a PhD in Business Administration at the Department of Management (University of Bologna, Italy) and a Post-Doc at the School of Business of the University of Technology (Sydney, Australia), he has been an academic faculty member of the University of Bologna and the Bologna Business School and acted as visiting professor at the Stern School of Business, New York University (New York, USA) and a number of other universities. He authored more than 100 international journal and conference publications in the areas of management, entrepreneurship, innovation and marketing, with a focus on inter-organisational relationships, the impact of ICTs and data analytics on managerial decisions and policy making, performance analysis and measurement, service innovation, make or buy decisions. His current research interests include big data analytics, eWOM, digital business models, digital transformation and coopetition strategies.

GIUSEPPE DI FATTA (M '03) is Associate Professor of Computer Science and the Head of the Department of Computer Science at the University of Reading, UK. In 1999, he was a research fellow at the International Computer Science Institute (ICSI), Berkeley, CA, USA. From 2000 to 2004, he was with the High-Performance Computing and Networking Institute of the National Research Council, Italy. From 2004 to 2006, he was with the University of Konstanz, Germany. His research interests include data mining algorithms, distributed and parallel computing, big data in sciences and data-driven multidisciplinary applications. He has published over 100 articles in peer-reviewed conferences and journals. He serves in the editorial board of the Elsevier Journal of Network and Computer Applications. He is the co-founder of the IEEE ICDM Workshop on Data Mining in Networks and has chaired several international events, such as the 2015 International Conference on Internet and Distributed Computing Systems. He is a Member (M) of IEEE since 2002 and a fellow of the Higher Education Academy, UK, since 2009.

MARCO DI FELICE received his Bachelor Degree (summa cum Laude) and Master Degree (summa cum Laude) in Computer Science respectively in 2002 and 2004, both from the University of Bologna, Italy. From August 2007 to December 2007 and April 2008 to August 2008 he was a visiting researcher at the Broadband Wireless Networking Laboratory, School of Electrical and Computer Engeneering, Georgia Institute of Technology, Atlanta, USA. From September to December 2009, he was a visiting researcher at the Northeastern University, Boston, USA. Since October 2015, he is Associate Professor at the Department of Computer Science and Engineering of the University of Bologna. He authored more than 100 international journal and conference publications on mobile systems, protocols, modeling and simulation. His current research activities include: the Internet of Things (IoT) and Industry 4.0, mobile networks and applications, data mining and analytics.