

Flying to Quality: Cultural Influences on Online Reviews

Panagiotis Stamolampros

(p.stamolampros@uea.ac.uk)

Norwich Business School, University of East Anglia., Norwich, United Kingdom.

Nikolaos Korfiatis

(n.korfiatis@uea.ac.uk)

*Norwich Business School and Centre for Competition Policy (CCP), University of East Anglia
Norwich, United Kingdom.*

Panos Kourouthanassis

(pkour@ionio.gr)

Department of Informatics, Ionian University, Corfu, Greece.

Efthymia Symitsi

(E.Symitsi@uea.ac.uk)

Norwich Business School, University of East Anglia.

Forthcoming: Journal of Travel Research

This is an author accepted version scheduled to appear in the above journal. Please consult the final version at the publisher's website.

Flying to Quality: Cultural Influences on Online Reviews

Abstract

Customers increasingly consult opinions expressed online before making their final decisions. However, inherent factors such as culture may moderate the criteria and the weights individuals use to form their expectations and evaluations. Therefore, not all opinions expressed online match customers' personal preferences, neither can firms use this information to deduce general conclusions. Our study explores this issue in the context of airline services using Hofstede's framework as a theoretical anchor. We gauge the effect of each dimension as well as that of cultural distance between the passenger and the airline on the overall satisfaction with the flight as well as specific service factors. Using topic modeling, we also capture the effect of culture on review text and identify factors that are not captured by conventional rating scales. Our results provide significant insights for airline managers about service factors that affect more passengers from specific cultures leading to higher satisfaction/dissatisfaction.

Keywords: Electronic WOM, Online Reviews, Service Quality, Airlines, Cultural Differences, Structural Topic Model

Introduction

Do inherent cultural traits systematically affect customers' online rating behavior? This is the main question we address in this paper by employing a large dataset of online reviews from airline passengers. Online reviews form an important source of information for both consumers and firms (Godes and Mayzlin 2004; Hennig-Thurau et al. 2004; Dellarocas, Zhang, and Awad 2007; Dwyer 2007). With respect to travel and hospitality services, extant studies substantiate their importance as a way to capture service quality and satisfaction (Vermeulen and Seegers 2009; Sparks and Browning 2011). Online reviews help customers to mitigate their perceptions of risk and uncertainty before engaging in the service encounter (Sparks and Browning 2011; Sotiriadis and Van Zyl 2013), and as such, represent an important predictor of purchase decisions and service loyalty (Tanford and Montgomery 2015; Book, Tanford, and Chen 2016; Phillips et al. 2017).

To improve the understanding of the behavioral patterns driving individuals to adopt online reviews and formulate review judgments, scholars rely on a combination of performance and emotionally oriented stances that explain review behavior. This is done through the lenses of information quality of the review (Filiari and McLeay 2014), service quality features (Guo, Barnes, and Jia 2017), and review sentiment (Liang et al. 2015). Interestingly, the personal traits of the reviewer are hardly used as an explanatory factor with a notable exception the relation of the online review ratings with the reviewer personality (Jensen et al. 2013).

This study explores how reviewers' cultural values influence their provided online ratings and their textual justification. Personal culture is a predominant factor of service expectation and evaluation in the travel and hospitality context (among others, Cheok, Hede, and Watne 2015; Mazanec et al. 2015; Nath, Devlin, and Reid 2016) and its relation to online communication has been validated in the context of online social networks (Jackson and

Wang 2013; Krishnan and Lymm 2016; Sheldon et al. 2017). We perform a case study in the airline industry to illustrate the relationship between reviewer cultural traits and online rating scores. Based on a rich dataset from TripAdvisor (557,208 reviews), where passengers evaluate their experience with a particular airline, we explore passengers' rating behavior from different perspectives controlling for flight and passengers' characteristics. We follow Hofstede's (1984; 2010) framework as a theoretical anchor to capture passengers' cultural traits and relate them to the numerical score of the review (most commonly referred in the literature as review valence). Furthermore, we employ the cultural distance formalization of Kogut and Singh (1988) and introduce the cultural incongruence as a factor that affects passengers' negative online ratings towards airlines. Considering that TripAdvisor allows reviewers to rate specific service factors of their overall experience, we extend our analysis to the individual aspects of the rating. To make the results of our analysis more robust, we control for flight length, cabin class, and reviewer's level of contribution to TripAdvisor.

In addition to examining the effect of cultural traits and cultural distance on numerical ratings, we take advantage of recent applications of topic modeling in marketing research (Tirunillai and Tellis 2014; Guo, Barnes, and Jia 2017) to evaluate the impact of cultural traits on the textual content of reviews. To this end, we use the structural topic model methodology (Roberts, Stewart, and Airolidi 2016) that allows the inclusion of review metadata as covariates. In doing so, our study is novel from the perspective that it captures those service factors that are more important for each cultural dimension.

Our study contributes to the literature in several ways. To the best of our knowledge, this is the first comprehensive study that explores the effect of consumer cultural characteristics on online rating behavior manifested from both quantitative (overall rating, service aspect rating) and qualitative aspects (review text). Furthermore, we add to the literature that tackles cultural differences and how these unfold in service encounters by exploring the impact of the

cultural distance between the country of origin of the passenger and the service provider on online ratings. To illustrate the importance of our findings we measure the degree of informational content distortion caused by different response patterns, employing a within culture standardization of the overall satisfaction with a carrier and gauging its effect on the global airline ranking. In doing so, we reveal the loss of information that is attributed to cultural differences, providing important managerial implications for travel and hospitality stakeholders. Contrary to the extant literature that focuses on the effect of culture on specific countries, this study is based on an extensive sample which increases the statistical power of our analysis (passengers from 203 countries and territories). Our empirical analysis also offers methodological novelty with the introduction of the structural topic modeling, an extension of Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003), as a method to infer categories of interest for customers through the analysis of the review text and how these categories change relative to cultural dimensions.

To this end, the rest of the paper is organized as follows: Section 2 summarizes the theoretical grounding and hypotheses formulation of the study. The description of the data used, the methodology followed, and our results are presented in section 3. In section 4 we provide an analysis of the service aspect discourse on the textual content of online reviews and propose an alternative approach for the estimation of overall satisfaction, displaying the effect of cultural differences on the ranking of airline carriers. The paper concludes in section 5, discussing theoretical and managerial implications as well as the limitations of the present study.

Theoretical Background and Hypotheses Formulation

In highly competitive markets such as the airline industry, passenger satisfaction is a core element for corporate profitability and sustainability (Chen 2008). Several studies report the connection between customer service, customer satisfaction, and corporate profitability for airlines (Behn and Riley Jr 1999; Steven, Dong, and Dresner 2012). The literature tends to measure airline service quality through performance metrics, such as flight delays, customer complaints, mishandled baggage, consumer satisfaction indexes, or survey questionnaires mainly based on SERVQUAL (Parasuraman, Zeithaml, and Berry 1985; 1988), neglecting the importance of online reviews (See for example Suzuki, Tyworth, and Novack 2001; Chen 2008; Keiningham et al. 2014; Kuo and Jou 2014).

Compared to other informational cues, online reviews come with several attractive advantages. First, they directly capture individual passengers' perceptions of the service quality provided in contrast to aggregated operating performance measures. Second, they offer access to a large pool of passengers which would require significant effort and cost to collect through surveys. Third, they allow users to provide both quantitative and qualitative information, rating specific aspects and describing their overall experience. The latter could be used to extract factors of customer satisfaction that are not captured accurately through abstract numerical scales or they do not cover the whole spectrum of the multidimensionality of service quality (Tellis and Johnson 2007). Thus, this study aims to extract insights from both the numerical rating and the textual content of online reviews.

Cultural Effects on Customers' Evaluation

The importance of cultural differences with regards to the customers' expectations of service quality has long been established in the literature (Malhotra et al. 1994; Donthu and Yoo 1998; Furrer, Liu, and Sudharshan 2000). Winsted (1997) investigates service encounters of

American and Japanese consumers and reveals significant cross-national differences between factors that different nationalities value more. (Donthu and Yoo 1998) using Hofstede's (1984) cultural dimensions find that a customer's cultural orientation has a strong influence on her expectations about overall service quality. Several other studies report significant effects for different cultural dimensions (see for example, Crotts and Erdmann 2000; Liu, Furrer, and Sudharshan 2001; Voss et al. 2004; Kim, Lee, and Mattila 2014). A similar stream of the literature connects cultural characteristics with survey response patterns. For example, De Jong et al. (2008) report a positive relationship between extreme response values with individualism, uncertainty avoidance, and masculinity. Although the effect of culture on service evaluation has been investigated in the past, the impact of cultural traits on online reviews that serve as direct proxies of service quality remains unexplored. This is extremely important in the case of review aggregators that accumulate reviews from an international pool of reviewers, since the effect of cultural traits could be significant, distorting online reviews' informational content.

Hofstede Cultural Dimensions

Geert Hofstede's work, with the use of a world-wide survey of thousands of employees in IBM, proposes a cultural dimension framework of four dimensions that describe cross-cultural communication and the effect of societal values and culture on its members, namely: Power Distance, Individualism (vs. Collectivism), Uncertainty Avoidance and Masculinity (vs. Feminism) (Hofstede 1984). In subsequent studies, this framework is extended to include two further dimensions namely: Long-Term Orientation (vs. Short-Term Orientation) and Indulgence (vs. Restraint) (Hofstede, Hofstede, and Minkov 2010). Although alternative frameworks exist such as those derived from the GLOBE study (House et al. 2004), Hofstede's dimensions are the most widely used proxies for measuring cultural traits on a

national or individual scale. We discuss the hypothesized effect of each cultural dimension on review valence below.

Power Distance refers to the extent to which the less powerful members of organizations and institutions expect and accept that power is distributed unequally. High power distance cultures (e.g., Russia) tolerate inequalities and respect the social hierarchy; low power distance cultures follow a more egalitarian philosophy when evaluating service outcomes (e.g., Denmark). Differences in the service quality perceptions of individuals from high and low power distance cultures, are derived from their perceptions about the status and power of the service provider. For instance, Tam et al. (2016) highlight that individuals from high power distance cultures are predisposed to accept the status differences between a service provider and themselves because they view the service provider as more dominant compared to themselves. This attitude stems from service providers' possession of resources, experience, and skills. Donthu and Yoo (1998) refer explicitly to airlines as an example of that kind of power. In their study, they report that consumers ranked low on power distance have higher overall quality expectations compared to consumers from countries that are ranked high on that dimension. Furrer, Liu, and Sudharshan (2000) also find that in higher power distance countries, customers are more likely to tolerate failures from the more powerful service providers. Low power distant cultures, on the other hand, tend to underestimate asymmetries in the power balance between the service provider and themselves. Therefore, we expect passengers from high distance power countries to be less critical to airlines as they accept their authority and expertise, and we formulate the following research hypothesis:

H1: Power Distance has a positive effect on review valence.

Individualism specifies a social framework where humans take care of themselves and their families as opposed to collectivism where individuals promote tightly-knit frameworks and higher in-group integration in exchange for their loyalty. Many studies have identified differences in service quality perceptions between individualists and collectivists (e.g., Maiyaki 2013; Sabiote-Ortiz, Frías-Jamilena, and Castañeda-García 2016). In essence, individualism is associated with higher service quality expectations. Customers from countries with high level of individualism (e.g., United States) are more likely to complain about disconfirmations in the perceived service quality (Liu, Furrer, and Sudharshan 2001; Kim, Lee, and Mattila 2014). We expect that this behavior is also reflected when individuals evaluate the service quality of airline companies. Thus, we formulate the following research hypothesis:

H2: Individualism has a negative effect on review valence.

The *uncertainty avoidance* dimension measures individuals' tolerance and comfort with ambiguity. High uncertainty avoidance cultures (e.g., Belgium) tend to have more stress and anxiety compared to low uncertainty cultures (e.g., Sweden). Moreover, they take fewer risks and are more reluctant with new technologies compared to their counterparts with low uncertainty avoidance values. Extant studies find that online reviews may serve as a mitigation instrument to reduce uncertainty in service encounters (Filieri 2015; Z. Liu and Park 2015). With regards to cultural values, there is evidence that in order to alleviate the emotion of uncertainty and reduce post-purchase cognitive dissonance, individuals from high uncertainty avoidance countries are more likely to praise good service quality but provide more critical feedback in cases of poor service quality encounters compared to individuals from low uncertainty avoidance countries (Groschl and Doherty 2006; Tseng 2017). Donthu and Yoo (1998) posit that customers with higher uncertainty avoidance, because of their risk

averse nature, search more the attributes of the product and service, and therefore have higher expectations. Voss et al. (2004) also report a negative relationship among customer evaluations and uncertainty avoidance. Additionally, Reimann, Lünemann, and Chase (2008) find that clients from countries with a higher degree of uncertainty avoidance are less satisfied than clients from lower uncertainty avoidance countries when their service expectations are not met as a result of service defects. This is explained by the narrow zone of tolerance of customers of countries with higher degree of uncertainty avoidance. Consequently, we formulate the following hypothesis:

H3: Uncertainty Avoidance has a negative effect on review valence.

Service quality perceptions are reported to differ between *masculine and feminine oriented cultures*. Masculine oriented cultures (e.g., Japan) value achievement, success, and materialism while feminine oriented cultures adhere to a lifestyle that favors quality of life and interpersonal relations (e.g., Norway). With respect to service evaluations, individuals from high masculine cultures have a stronger motivation to provide feedback than those in more feminine cultures because they want to express their experience with the service to others (Fang et al. 2013). Such individuals are more likely to complain about poor service quality than individuals from more feminine cultures because they are less tolerant of service failures, and they perceive themselves to have the power to confront service providers for the unsatisfactory experience or even terminate their future interactions with them (Torres, Fu, and Lehto 2014; Van Vaerenbergh et al. 2014). In the specific context of airline passengers, Crotts and Erdmann (2000) report that passengers from masculine societies are more likely to report defector attitudes while passengers from feminine societies are more loyal to specific airlines. As such, we expect that travelers from masculine oriented cultures approach the

service evaluation process with a more critical perspective. Thus, we examine the following hypothesis:

H4: Masculinity has a negative effect on review valence.

Time orientation captures humans' consideration of their future. Hofstede distinguishes between individuals that are willing to make sacrifices now for their long-term benefit (a life strategy coined as "*long-term orientation*") and individuals who focus on achieving immediate gratifications than waiting for long-term fulfillment (coined as "*short-term orientation*"). Studies suggest that individuals from long-term oriented cultures (e.g., South Korea) are less likely to provide negative feedback pertaining to the service experience compared to short-term oriented individuals (e.g., those from Argentina), because they are not willing to uphold the risk of compromising their long-term relationships with the service provider (Liu, Furrer, and Sudharshan 2001; Ryu and Moon 2009). To the contrary, short-term oriented individuals have higher expectations from service providers, and, as such, are expected to be more critical (Mazaheri, Richard, and Laroche 2011; Meng and Mummalaneni 2011). In effect, long-term oriented individuals value loyalty with the service provider (Bartikowski, Walsh, and Beatty 2011; X. R. Li et al. 2011) and we expect this behavior to be reflected in their online ratings. Thus, we hypothesize:

H5: Long-Term Orientation has a positive effect on review valence.

A final inclusion to Hofstede's cultural dimensions is *indulgence*. Indulgence is interpreted as the degree to which individuals can control their impulses. Customers from high scoring indulgence countries (e.g., Mexico) actively follow their needs and desires whilst customers from low scoring indulgence countries tend to value restraint (e.g., Estonia). This behavior is also reflected in the use of online tools that enable social interaction.

Restraint-oriented cultures exhibit a reluctance in using online social networks (Krishnan and Lymm 2016; Stump and Gong 2017), which may be attributed to their averseness to self-disclosure. Likewise, indulgence and restraint are associated with emotional valence. Scholars report that indulgent cultures are happier than restraint ones (e.g., Park, Baek, and Cha 2014), with a more positive attitude as they are more optimistic and more likely to remember positive emotions. Individuals from restrained societies on the other hand, are less happy, less likely to remember positive emotions and more pessimists (Hofstede, Hofstede, and Minkov 2010). Consequently, either because of a more positive (negative) stance in life, or a higher possibility to recall the positive (negative) emotions from their experience or even a more “open” (“closed”) attitude to a service provider we expect that this difference is also mirrored to the emotional valence of their reviews. Thus, we formulate the following hypothesis:

H6: Indulgence has a positive effect on review valence.

Cultural norms influence both individuals’ expectations and their perceptions of received service quality (Weiermair 2000). The previous sections argue that cultural differences between customers of service providers influence the degree of accumulated satisfaction with service encounters and have an impact on evaluation ratings. Nevertheless, the literature suggests that cultural differences between individuals and service providers may also be the cause of service conflicts, which may be attributed to variations in their culturally biased standards. Such conflicts are likely to be weaker when two cultures are similar than when they are diverse (M. Li 2014). Scholars measure the degree of dissimilarities between cultures with “*cultural distance*” (Ye, Zhang, and Yuen 2013; Cheok, Hede, and Watne 2015). In service encounters where cultural distance is high, individuals may perceive mismatches in their service expectations and actual service performance, that are attributed to

the deficiency of the service providers to account for various cultural standards (Laroche et al. 2004; Paswan and Ganesh 2005). An essential proxy of cultural difference is the difference between individuals' and service providers' countries of origin. Evidence in the literature suggests that customers formulate stronger loyalty ties towards service providers from the same country of origin (Javalgi, Cutler, and Winans 2001; Thelen and Shapiro 2012). This is attributed to increased comfort perceptions during the service encounter (Paswan and Ganesh 2005). Therefore, we expect that airline passengers will favor airlines from countries with similar cultural characteristics and we propose the following research hypothesis:

H7: The cultural distance between passenger and airline's country of origin has a negative effect on review valence.

Data, Methods and Results

Sample

We collected reviews from TripAdvisor, the most popular review aggregator which provides booking services to all travel related activities. TripAdvisor adopts a mixed model that allows it to function as both an online travel intermediary and review aggregator, with its ratings used by hotels and restaurants worldwide as an indication of service quality. Due to the lenient availability of data content to researchers, TripAdvisor has been heavily used in the literature of electronic word-of-mouth (e.g., see Crotts, Mason, and Davis 2009; Pearce and Wu 2016). While TripAdvisor's primary offering to consumers comprises aggregating ratings of hotels and restaurants, the company has recently launched a section where passengers share and evaluate their flight experiences with a specific carrier. Therefrom, we gathered all publicly available reviews until 08/2017 comprising a total of $N= 557,208$ reviews. In addition to the review text, metadata containing information about passenger/reviewers'

country of residence, flight date, name of air carrier, route, cabin class (Economy, Economy Premium, Business Class, and First Class), and an overall rating for the flight experience (in an ordinal scale from 1 to 5), were also collected. Each rating is also accompanied by an optional rating (aspect rating) for eight specific aspects of the flight experience namely: (i) *Legroom*, (ii) *Seat Comfort*, (ii) *Customer Service*, (iii) *Value for Money*, (iv) *Cleanliness*, (v) *Check-in and Boarding*, (vi) *Food and Beverage* and (vii) *Inflight Entertainment / Wi-Fi Connectivity*. Unlike other review aggregators (e.g., Booking.com), TripAdvisor does not aggregate the ratings given to individual aspects to form the overall score. This allows us to evaluate our theoretical model not only on the overall score but also on the individual ratings given for various aspects of the flight experience.

<INSERT TABLE 1 HERE>

Table 1 provides the description of the characteristics of our sample. The 557,208 reviews in our dataset are written by 376,519 passengers originating from 203 countries and territories, providing ratings for 489 airlines registered in 147 countries. Approximately half of the reviews in our sample are in English (254,424) with an average text length of 560 characters.

Table 2 provides a breakdown of the review valence in our sample by service aspect and cabin class, as well as, additional metadata available for each review. The average overall rating for all reviews in our sample was relatively good ($M=3.68$, $SD = 1.29$) and not substantially different from the ratings given to the other aspects of the flight experience.

<INSERT TABLE 2 HERE>

Dependent Variables and Controls

The dependent variable used in our model (review score) is an ordinal Likert-scale with values between 1 and 5 and captures the overall satisfaction of the passenger with the service he/she received by an airline during a flight. The individual ratings for the various aspects of

the flight were also employed as dependent variables to evaluate our theoretical model. We obtained the values for the Hofstede dimensions using passengers' self-reported country of origin. We also controlled for additional variables that could have an influence on the overall rating or on the rating of a specific aspect, such as cabin class, flight distance, and reviewers' (passengers') level of contribution to TripAdvisor. Cabin class was coded as a categorical variable with four levels (Economy, Economy Premium, Business Class, and First Class). Flight distance was measured as the geographical distance (in kilometers) between the departure and the destination airport and was estimated via the Haversine method using the coordinates (latitude and longitude) obtained from Google's geolocation API. Finally, reviewers' level of contribution was sourced by the review metadata, which are displayed on each review.

Based on the work of Hofstede, Hofstede, and Minkov (2010), we examined the effect of the six cultural dimensions on passengers' ratings. To this end, we used ordered logistic regression analysis with review ratings as dependent variables controlling for the variables discussed previously. Consequently, our econometric specification for the i -th review r_i^c that belongs to the rating category c had the following form:

$$r_i^c = \beta_0 + \beta_1 PowerDistance + \beta_2 Individualism + \beta_3 UncertaintyAvoidance \\ + \beta_4 Masculinity + \beta_5 LongTermOrientation + \beta_6 Indulgence \\ + \beta_7 \sum_{m=1}^4 Cabin\ Class + \beta_8 FlightDistance + \beta_9 ReviewerExpertise + \epsilon$$

The Effect of Cultural Dimensions on Passengers' Rating

Table 3 reports the results of each rating category. Multicollinearity was evaluated for all models using the variance inflation factor and was not found to cause any concern in any of our econometric specifications. The results reveal a significant positive effect of Power

Distance ($\beta_1=0.003, p<0.001$), supporting H1. The direction and significance of this effect is found to be similar across all the service categories/aspects that are evaluated in our model. Our results supported hypotheses H2 to H4, revealing a significant negative effect for Individualism ($\beta_2=-0.003, p<0.001$), Uncertainty Avoidance ($\beta_3=-0.001, p<0.001$), and Masculinity ($\beta_4=-0.001, p<0.001$). Similar effects are also reported for the individual service factors. Long-Term Orientation displays an opposite effect from the examined hypothesis (H5) though at a lower significance level ($\beta_5=-0.001, p<0.05$). We find a positive association for most tangible aspects such as seating and legroom the relationship while for more intangible aspects, such as customer service and check-in, the direction is opposite. This is in line with the findings of Furrer, Liu, and Sudharshan (2000) who describe that in long-term oriented cultures, reliability, responsiveness, and empathy are extremely important while tangibles are not so necessary. With regards to H6, results show that the effect of Indulgence is not significant to the overall rating ($\beta_6=0.000, p>0.05$), however it has a positive and statistically significant effect on most of the other service factors (apart from *food* and *value for money* categories where the effect is not significant). Thus, H6 was partially supported.

With regards to our control variables, both flight distance and reviewer expertise had a positive effect. The effect of the former can be attributed to the fact that long distance flights are usually performed by bigger aircrafts and provide more services to passengers. On the other hand, reviewers that contribute less to TripAdvisor are likely to be those that post for retaliation to service failures, in contrast to those that are more active contributors to the platform. Lastly, we intuitively found cabin class upgrades to result in more positive rating due to the upgrade in level of service.

<INSERT TABLE 3 HERE>

<INSERT TABLE 4 HERE>

Does the Cultural Distance between Passenger and Airline Influence the Rating Behavior?

Hitherto, our results assessed the effect of each Hofstede dimension on the overall rating as well as on specific service factors. Considering that Hofstede dimensions can also be used to explain not only individual cultural traits but also cross-national differences, we extended our analysis on how cross-national differences impact the overall score and the operational aspects captured by the ratings. To test this effect we computed the cultural distance using the Kogut and Singh (1988) formula as follows:

$$Distance_{pa} = \frac{1}{6} \sum_{n=1}^6 \frac{(D_{pi} - D_{ai})^2}{V_i}$$

, where D_{pi} is the i -th Hofstede dimension for the country of the passenger p , D_{ai} is the value for the same dimension for the country of the airline a , and V_i denotes the variance of that dimension. The econometric model is estimated as before, replacing the individual Hofstede dimensions with the computed cultural distance.

The results reported in Table 4 display a strong negative association with the overall score ($\beta = -0.027$, $p < 0.001$), supporting H7. However, the impact varies with the individual service aspects. More specifically, the relationship is positive for legroom, seating, and value for money. However, service aspects that are more subject to cultural influences from the country of origin of the carrier, such as the interaction with the personnel (customer service and check-in/boarding), and inflight entertainment receive a lower rating on average when the cultural distance between the passenger and the carrier increases. The same, though insignificant, direction could be observed for the food category.

The Effect of Cultural Dimensions on the Informational Content of Online Reviews

Passengers' ratings to a set of predefined service factors provide useful information for airline managers but at the same time carry the limitation that the preselection of these categories constrain their informational content. Other factors that may please or irritate the passengers that are not explicitly defined on the rating scales cannot be captured. Textual analysis allows us to overcome this limitation by exploring the informational content of the review text. Using recent advances in topic models, we explored how the textual content of a review varies with passengers' cultural dimensions.

Topic modeling has gained attention in marketing, tourism, and hospitality research (Tirunillai and Tellis 2014; Guo, Barnes, and Jia 2017) as an important methodology for exploring customer provided textual information. In principle, topic modeling is a set of unsupervised machine learning techniques which self-organize textual corpora in groups of topics evaluating how specific groups of words appear together using both volume and context as inputs. In our analysis, we consider recent advances in topic modeling and specifically structural topic models (Roberts et al. 2014; Roberts, Stewart, and Airoldi 2016). Structural Topic Modeling (STM) is a probabilistic topic modeling method where topic coverage and word distribution are inferred using Bayesian techniques. It builds on established probabilistic topic models such as Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) or Correlated Topic Model (Blei and Lafferty 2006) where documents (which in our case are the review texts) represent a mixture of latent topics and each of these topics is described by a word distribution.

An essential difference of the STM method compared to other topic models, such as LDA or Correlated Topic Model, is that it allows the inclusion of document metadata (or covariates). This novelty enables us to connect additional characteristics about a document

with a documents' degree of association with a topic (topical prevalence) and the degree of association of a word with a topic (topical content), thus relaxing the highly restrictive assumption of exchangeability that is found in LDA. Exchangeability assumes that all authors are equally likely to write a document, while in STM the probability of topic prevalence relates to other covariates. In our case, this allowed us to connect each Hofstede dimension with the topics derived from our analysis and reach useful conclusions about the topics that are discussed more based on passengers' cultural traits.

We followed a three-step process to perform our analysis through STM. First, the text of each review was pre-processed to create an appropriate corpus for analysis. Second, we fit an STM model to identify the number of topics that describe better the variability of the corpus and labeled them accordingly with the help of experts. Finally, we analyzed the effect of Hofstede dimensions on the prevalence of the topics obtained from our STM solution. We describe these steps in more details in the sections that follow.

Text Preparation for Analysis

We constrained our analysis only to reviews written in English since our topic model approach works best with text corpora in this language. From the total sample of $N=557,208$ reviews, $N_{\text{eng}}=254,424$ reviews are in English and form the initial corpus that was used for our analysis. We followed the text pre-processing workflow used in previous studies in the literature (Tirunillai and Tellis 2014; Guo, Barnes, and Jia 2017) to prepare the text for our analysis. This included (a) word text tokenization, (b) elimination of numbers, punctuation marks, (c) removal of language stop words (using the SMART stop-word list), as well as context-specific stop words such as names of airlines and routes, and words with length under a specific threshold (number of characters <3), and (d) filtering of the remaining words using part-of-speech (POS) tagging to keep only nouns as well as adverbs and adjectives (in order to capture sentiment). For step (d) we used the Stanford NLP parser. After pre-processing, the

remaining words were lemmatized to group words with the same root form and we filtered the terms, keeping those that appeared in at least 1% of the total reviews in our initial corpus (N_{eng}). This produced a set of $N_{stm}=184,502$ reviews which comprised our final corpus.

Estimating the Number of Topics

Our topic solution was estimated in R using the stm package. Following Roberts, Stewart, and Tingley (2017), we ran an iterative process to select and evaluate the number of topics using three criteria: (a) Heldout likelihood (a measure on how the number of topics explains the overall variability in our corpus), (b) Exclusivity of topic words to the topic and (c) Semantic coherence of the topic structure. We used the recommended approach of initializing our estimation with spectral decomposition in addition to a seed vector (K) of the candidate number of topics rather than using Gibbs sampling on the Dirichlet distribution (Lee and Mimno 2014). Considering that the primary metadata associated with review text is its numerical rating, we used the overall score as the primary covariate to estimate the topic solution that contained both positive and negative aspects of the same topic in our final model. We began with an initial number of $K_{min}=8$ topics as a seed value since this is the number of rating aspects that are provided by TripAdvisor on its review interface, and evaluated the heldout likelihood for a maximum of $K_{max}=40$ topics in our sample.

The candidate topic solutions with the highest heldout likelihood was then evaluated against the ratio of their semantic coherence and exclusivity. Semantic coherence is a criterion developed by Mimno et al. (2011) that increases based on the frequency of co-occurrence of the most probable words in each topic of the topic solution. On the other hand, exclusivity considers the mutual appearance of the most probable words in more than one topic and can be used to evaluate overall topic quality for each candidate model. A combination of these criteria can be captured through the FREX criterion (Roberts, Stewart,

and Airoldi 2016), which considers a weighted harmonic mean of a word's rank in terms of exclusivity and frequency in a k-topic solution as follows:

$$FREX_{k,u} = \left(\frac{\omega}{ECDF(\beta_{k,u} / \sum_{j=1}^k \beta_{j,u})} + \frac{1 - \omega}{ECDF(\beta_{k,u})} \right)^{-1}$$

, where $k \in K$ is the k -th topic, u is the word under consideration, β is the word distribution for this topic, and ω a prior used to impose exclusivity (in our case it was equal to 0.7).

After considering the above criteria, we selected a K=20 topic solution to describe the variability of our corpus given the relationship between heldout likelihood, semantic coherence, and exclusivity. The final output comprised of 184,502 reviews and a 413-word dictionary. For the labeling of the topics, two experts with experience in dealing with airline customer service were recruited to evaluate each topic of the optimal topic solution and assign a label. Both experts agreed that the selected topic solution had a high degree of coherence in terms of the top loading reviews and assigned mutually acceptable labels. Table 5 provides the estimated topic solution along with the words with the highest FREX score and the assigned labels.

<INSERT TABLE 5 HERE>

For each topic, we estimated the expected proportion by averaging the loading of each document on the topic solution over the total documents in the final corpus. As can be observed in the third column of Table 5, delays and staff praise are the topics with the highest prevalence. Other significant topics refer to service failure recoveries such as refund after a flight is cancelled, customer service complaints, and critiques towards the staff.

Estimating the Effect of Cultural Dimensions on Review Text

Having estimated our topic model solution, we evaluated the effect of each cultural dimension on the prevalence of the topics in our corpus. To achieve that, we regressed the

topic proportions for the estimated topic solution with each of the Hofstede cultural dimensions controlling at the same time for the review score and all the controls used in our previous specification. This allowed us to draw proportional odds from the conditional expectation of topic prevalence given the metadata associated with this review. In our case, this corresponded to the loading of a particular review on a topic using the differential effect of its associated metadata (Hofstede's dimensions).

<INSERT FIGURE 1 HERE>

Figure 1 displays the expected change of topic proportions for low and high values of each of the Hofstede dimensions, providing interesting insights for the cultural effects on the review text. In the continuum of the Power Distance dimension, passengers from more power distance societies are more critical to staff and more prone to complain about baggage fees, delays and service recovery failures. As we approach the other extreme, passengers are more willing to praise the staff and they are very appreciative of staff assistance and in-flight services such as food/beverage and entertainment. A very similar picture is observed at the continuum of the Individualism cultural dimension. On the other hand, Uncertainty Avoidance, Indulgence, and Masculinity display smaller marginal effects on topic prevalence as they lie closer to the dotted line which represents zero effect. More specifically, ticket cost has the highest change (increase) in topic prevalence for Uncertainty Avoidance, value for money exhibits the highest change (decrease) for Indulgence, while staff praise and staff assistance have the strongest effects for Masculinity. Finally, the effect of Long-Term Orientation is quite strong and when moving from short-term to long-term oriented cultures, passengers are more sensitive to check-in /boarding, price, and value for money. Moving on to the opposite extreme for this dimension, passengers are more sensitive to extra fees and the general baggage policy of the carriers.

Cultural Bias Correction and its Effect on Airline Ranking

Our results reveal a robust influence of the cultural characteristics on both the numerical and textual part of the reviews. Considering that online reviews stem from an international pool of reviewers with different cultural backgrounds, cross-cultural differences may distort the informational content derived from online reviews. Not accounting for this distortion may lead to misinterpretations especially for deducing general conclusions about the quality of a service provider. In this section, we study the magnitude of that distortion on the quality overall ranking of airlines based on the provided review ratings. In particular, we gauged the difference between the overall ranking of airlines between the raw and standardized ratings that consider the country-of-origin of the reviewer. The overall standardized satisfaction of the reviewer (y) was estimated as follows:

$$y_r = \frac{r_i - \text{mean}(r_c)}{\text{dispersion}(r_c)}$$

, where r_i was the observed rating (*overall score*) of the reviewer i which was standardized by the overall mean and dispersion of all reviews (r_c) submitted to TripAdvisor by reviewers from the same country (within-country standardization). This approach is widely used when dealing with cross-cultural response biases (Fischer 2004). To ensure the same level of variation for the within-country used sample, we used a threshold considering only reviews submitted by passengers coming from countries that account for at least 0.5% of our total sample. A second filtering rule required that airlines have at least 0.5% of total reviews. This resulted in a sample of 37 international airlines which are used to rank them based on the overall score and the standardized overall score. To evaluate the loss of information in this ranking due to cultural influence, we employed Kendall's tau rank correlation coefficient to check the concordance among the ranking pairs. Our results indicated a strong ($\tau=87\%$) and significant ($T = 946, p < 0.000$) coefficient of concordance. This reveals a level of

disagreement suggesting that when passengers' cultural traits are not taken into consideration, this can lead to a distortion of the ranking information extracted from the passenger reviews. Table 6 shows the top twenty airlines ranked by the overall score (*Column A*) and the standardized overall score (*Column B*). The differential effect of the dispersion in the ranking of each airline (Δ Rank) is shown in the final column. We should note here that the results are similar when we apply different sampling filters or use the whole sample.

<INSERT TABLE 6 HERE>

Conclusion, Implications, and Limitations

Summary of Contribution

Our study contributes to the ever-growing stream of literature on online reviews and electronic word-of-mouth (Purnawirawan et al. 2015; Book, Tanford, and Chen 2016; Choi et al. 2016; Phillips et al. 2017; Symitsi, Stamolampros, and Daskalakis 2018) by providing new insights into how cultural traits might affect evaluations in the context of online ratings for service encounters. While the majority of studies in electronic word-of-mouth consider the case of product evaluations from a single country, our study of airline ratings provides an analysis of reviews using a pool from international passengers examining ratings across multiple countries. We also examine the existence of such effects on post-purchase evaluations and not on the selection process, which is the primary focus of the extant literature. Although such dynamics may exhibit a significant effect on the informational content of online reviews, they have remained rather unexplored.

Our findings show that there are variations in airline passengers' perceived service quality satisfaction based on the differences in their inherent cultural values. These differences are reflected not only on the overall perceived satisfaction regarding the service

quality of an airline company, but also on all individual aspects (e.g., perceived satisfaction from staff, food, seat comfort, cleanliness). Interestingly, our findings also document a negative association between the cultural distance of the passenger and the airline company and the accumulated perceived satisfaction from the service quality. This implies that passengers are more satisfied from airline companies that are more closely associated with their cultural values. As such, this study contributes to the literature that relates culture and service satisfaction in the broader travel and hospitality context (Laroche et al. 2004; Reisinger and Crofts 2010; M. Li 2014) and provides deeper insights on the specific behavioral patterns exhibited by individuals from different cultures paving the way for the development of atypical service quality profiles based on passengers' cultural orientations. We summarize the theoretical and practical contributions of our study in the following sections.

Theoretical Implications

Our study provides further evidence to the body of literature tackling the impact of cultural traits on service evaluations in multiple service contexts (e.g., Groschl and Doherty 2006; Torres, Fu, and Lehto 2014; Sabiote-Ortiz, Frías-Jamilena, and Castañeda-García 2016; Stamolampros and Korfiatis 2018; and many others). However, we empirically demonstrate that these associations may exhibit fluctuations when considering the discrete constituent service quality features and the cultural distance between the passenger and the service provider. Specifically, although extant studies have documented that at an inter-personal level, cultural dispositions may directly affect the extent to which individuals attribute service achievements (or failures) to the provider (Weiermair 2000; Laroche et al. 2004), cultural-similarities (or divergences) between individuals and service providers can alleviate (or reinforce) these perceptions of service satisfaction (dissatisfaction). Notably, our analysis showcases that cultural distance plays an important role in determining dissatisfaction not

only to the overall service experience, but also to all individual service quality dimensions that relate to interpersonal interactions (i.e., check-in, boarding, and customer service).

Moreover, this study provides a more in-depth understanding in explaining the effect of cultural values on service evaluation by considering not only the numerical part (review score) but also the textual content of online reviews. Such an approach has been overlooked by existing studies, which focus primarily on the statistical dependence between cultural values and online ratings (Fang et al. 2013; Purnawirawan et al. 2015). Our methodological stance utilizes a novel text mining approach to determine which service quality features are evaluated more favorably by each cultural dimension thus, enriching the cross-cultural research on services evaluation with an alternative approach.

Managerial Implications

Our study has significant implications for practitioners in the aviation and travel and hospitality sector. Online reviews provide a valuable tool for managers to efficiently explore customer preferences as well as firms' strengths and weaknesses in their service encounters. From that perspective, using online reviews as a performance measurement tool compared to standard methods of measuring customer-perceived service quality such as SERVQUAL (Parasuraman, Zeithaml, and Berry 1985; 1988), airlines are able to extract more context-specific and detailed information from much larger samples. Even though traditional survey-based methods (relying on questionnaires) provide a valid source of information, they come at a cost as they require time, careful sample selection procedures, and resources while at the same time are usually constrained to a limited (though representative) number of respondents. At the same time review text could be used to extract factors of customer satisfaction that may not be able to be measured through survey scales.

Our analysis also exposes several dynamics that affect passengers' perceived quality related to airline services and more importantly the influence of cultural factors. We demonstrate that the cultural background of passengers has a significant impact on their perceived expectations and overall satisfaction from the service encounter. In the context of an airline, operation managers need to collect upfront information about the cultural values of their customers, pinpoint whether there are differences or commonalities with the inherent cultural values of the service provider, and determine when adaptation may be necessary for regional or global interpersonal service approaches. This passenger-driven approach may dictate the design of a plethora of service interventions, such as tailored interpersonal interactions and associated before, in-, and after-flight service offerings, adapted communication/ marketing strategies, and personalized features in the website of the airline, including its central reservation system, based on passenger cultural values/ differences.

Along with this line, our content analysis of prevalent service quality features per cultural dimension may inform the development of cultural passenger clusters. Indeed, our analysis suggests that passengers from high power and individualism countries give emphasis to intangible service quality traits, such as baggage policies and delays, while passengers from masculine and long-term orientation countries give emphasis to service quality traits that relate to interpersonal interactions and overall value. Since the cultural trait of a passenger stems from the amalgamation of all individual cultural dimensions, identifying the prevalent service quality features per cultural dimension may be used by airline managers for further improving the respective service offerings per passengers' cultural cluster. This kind of information could also be used for airlines to explore new markets. A first step before any expansion to new routes could be to understand what passengers in those markets value more and evaluate the fitness of carrier's marketing mix based on these insights.

Our findings also reveal a distortion in the informational content of online reviews as a result of cross-national differences. Firms, customers, and policymakers should be cautious to the interpretation of raw data deriving from online reviews as the mean overall rating (or its dispersion) can be explained partially by different rating patterns. Therefore, in case of an international pool of reviewers, the informational content should be weighted by appropriate measures to eliminate such influences. Finally, review aggregators that accumulate opinions from an international pool of reviewers should employ alternative measures that consider cross-national differences in response patterns for revealing the true quality of a product or service.

Limitations

Nonetheless, our study is not without limitations, which are directly derived from the nature of online reviews. Several biases have been established in the literature of e-WOM such as self-selection (X. Li and Hitt 2008), and response biases (Hu, Zhang, and Pavlou 2009). Online reviews could be subject to manipulation (Choi et al. 2016). Nonetheless, primary data offers the opportunity to elicit and control for customers' personal characteristics, which are sparsely available on online reviews. Therefore, we are not able to control for several demographics factors (such as sex, age, level of education, etc.). Second, our analysis is not performed in the highest micro-level, (e.g., comparing the evaluation of different types of customers within the same flight or route), as we do not have enough observations for this type of analysis. However, by controlling for cabin class and duration of the service encounter (in the form of flight distance), we alleviate such concerns. Third, Hofstede's dimensions, that capture cultural traits and the cultural distance with the service provider, are aggregate measures. Within country variation exists, and individual responses could be more representative of the actual culture of a customer (see for example, Donthu and Yoo 1998). Nonetheless, our large sample size allows us to infer unbiased results as it can be assumed to

be representative of the whole population and not the result of outliers from the within-country variation. Last, our estimated topic model is restricted only to reviews that are in English and do not consider reviews written in other languages, thus excluding possible influences from non-English speakers.

Overall, our study by simultaneously assessing the impact of culture on both the numeric and the textual part of online reviews provides evidence about the influence inherent factors such as culture may moderate the criteria and the weights individuals use to form their expectations and evaluations. Therefore, customers that rely on online reviews to make their decisions should be cautious about the case that not all opinions expressed online match their personal preferences.

References

- Bartikowski, Boris, Gianfranco Walsh, and Sharon E Beatty. 2011. "Culture and Age as Moderators in the Corporate Reputation and Loyalty Relationship." *Journal of Business Research* 64 (9):966–972.
- Behn, Bruce K, and Richard A Riley Jr. 1999. "Using Nonfinancial Information to Predict Financial Performance: The Case of the US Airline Industry." *Journal of Accounting, Auditing & Finance* 14 (1):29–56.
- Blei, David M., and John D. Lafferty. 2006. "Correlated Topic Models." In *In Proceedings of the 23rd International Conference on Machine Learning*, 113–120. MIT Press.
- Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (Jan):993–1022.
- Book, Laura A, Sarah Tanford, and Yang-Su Chen. 2016. "Understanding the Impact of Negative and Positive Traveler Reviews: Social Influence and Price Anchoring Effects." *Journal of Travel Research* 55 (8):993–1007.
- Chen, Ching-Fu. 2008. "Investigating Structural Relationships between Service Quality, Perceived Value, Satisfaction, and Behavioral Intentions for Air Passengers: Evidence from Taiwan." *Transportation Research Part A: Policy and Practice* 42 (4):709–717.
- Cheok, Jason, Anne-Marie Hede, and Torgeir Aleti Watne. 2015. "Explaining Cross-Cultural Service Interactions in Tourism with Shenkar's Cultural Friction." *Current Issues in Tourism* 18 (6):539–560.
- Choi, Sungwoo, Anna S Mattila, Hubert B Van Hoof, and Donna Quadri-Felitti. 2016. "The Role of Power and Incentives in Inducing Fake Reviews in the Tourism Industry." *Journal of Travel Research* 56 (8):975–87.
- Crotts, John C, and Ron Erdmann. 2000. "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* 10 (6):410–419.
- Crotts, John C, Peyton R Mason, and Boyd Davis. 2009. "Measuring Guest Satisfaction and Competitive Position in the Hospitality and Tourism Industry: An Application of Stance-Shift Analysis to Travel Blog Narratives." *Journal of Travel Research* 48 (2):139–151.
- De Jong, Martijn G, Jan-Benedict EM Steenkamp, Jean-Paul Fox, and Hans Baumgartner. 2008. "Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation." *Journal of Marketing Research* 45 (1):104–115.
- Dellarocas, Chrysanthos, Xiaoquan Michael Zhang, and Neveen F Awad. 2007. "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures." *Journal of Interactive Marketing* 21 (4):23–45.
- Donthu, Naveen, and Boonghee Yoo. 1998. "Cultural Influences on Service Quality Expectations." *Journal of Service Research* 1 (2):178–186.
- Dwyer, Paul. 2007. "Measuring the Value of Electronic Word of Mouth and Its Impact in Consumer Communities." *Journal of Interactive Marketing* 21 (2):63–79.
- Fang, Hui, Jie Zhang, Yang Bao, and Qinghua Zhu. 2013. "Towards Effective Online Review Systems in the Chinese Context: A Cross-Cultural Empirical Study." *Electronic Commerce Research and Applications* 12 (3):208–220.
- Filieri, Raffaele. 2015. "What Makes Online Reviews Helpful? A Diagnosticity-Adoption Framework to Explain Informational and Normative Influences in e-WOM." *Journal of Business Research* 68 (6):1261–1270.
- Filieri, Raffaele, and Fraser McLeay. 2014. "E-WOM and Accommodation: An Analysis of the Factors That Influence Travelers' Adoption of Information from Online Reviews." *Journal of Travel Research* 53 (1):44–57.

- Fischer, Ronald. 2004. "Standardization to Account for Cross-Cultural Response Bias: A Classification of Score Adjustment Procedures and Review of Research in JCCP." *Journal of Cross-Cultural Psychology* 35 (3):263–282.
- Furrer, Olivier, Ben Shaw-Ching Liu, and Devanathan Sudharshan. 2000. "The Relationships between Culture and Service Quality Perceptions Basis for Cross-Cultural Market Segmentation and Resource Allocation." *Journal of Service Research* 2 (4):355–371.
- Godes, David, and Dina Mayzlin. 2004. "Using Online Conversations to Study Word-of-Mouth Communication." *Marketing Science* 23 (4):545–560.
- Groschl, Stefan, and Liz Doherty. 2006. "The Complexity of Culture: Using the Appraisal Process to Compare French and British Managers in a UK-Based International Hotel Organisation." *International Journal of Hospitality Management* 25 (2):313–334.
- Guo, Yue, Stuart J Barnes, and Qiong Jia. 2017. "Mining Meaning from Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent Dirichlet Allocation." *Tourism Management* 59:467–483.
- Hennig-Thurau, Thorsten, Kevin P Gwinner, Gianfranco Walsh, and Dwayne D Gremler. 2004. "Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?" *Journal of Interactive Marketing* 18 (1):38–52.
- Hofstede, Geert. 1984. *Culture's Consequences: International Differences in Work-Related Values*. Vol. 5. Sage Publications.
- Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov. 2010. *Cultures and Organizations: Software of the Mind: Intercultural Cooperation and Its Importance for Survival*. 3rd ed. McGraw-Hill.
- House, Robert J., Paul J. Hanges, Mansour Javidan, Peter W. Dorfman, and Vipin Gupta. 2004. *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Sage publications.
- Hu, Nan, Jie Zhang, and Paul A. Pavlou. 2009. "Overcoming the J-Shaped Distribution of Product Reviews." *Communications of the ACM* 52 (10):144–47.
- Jackson, Linda A, and Jin-Liang Wang. 2013. "Cultural Differences in Social Networking Site Use: A Comparative Study of China and the United States." *Computers in Human Behavior* 29 (3):910–921.
- Javalgi, Rajshekhar G, Bob D Cutler, and William A Winans. 2001. "At Your Service! Does Country of Origin Research Apply to Services?" *Journal of Services Marketing* 15 (7):565–582.
- Jensen, Matthew L, Joshua M Averbeck, Zhu Zhang, and Kevin B Wright. 2013. "Credibility of Anonymous Online Product Reviews: A Language Expectancy Perspective." *Journal of Management Information Systems* 30 (1):293–324.
- Keiningham, Timothy L, Forrest V Morgeson, Lerzan Aksoy, and Luke Williams. 2014. "Service Failure Severity, Customer Satisfaction, and Market Share An Examination of the Airline Industry." *Journal of Service Research* 17 (4):415–31.
- Kim, Min Gyung, Chung Hun Lee, and Anna S Mattila. 2014. "Determinants of Customer Complaint Behavior in a Restaurant Context: The Role of Culture, Price Level, and Customer Loyalty." *Journal of Hospitality Marketing & Management* 23 (8):885–906.
- Kogut, Bruce, and Harbir Singh. 1988. "The Effect of National Culture on the Choice of Entry Mode." *Journal of International Business Studies* 19 (3):411–432.
- Krishnan, Satish, and John Lymm. 2016. "Determinants of Virtual Social Networks Diffusion: Insights from Cross-Country Data." *Computers in Human Behavior* 54:691–700.
- Kuo, Chung-Wei, and Rong-Chang Jou. 2014. "Asymmetric Response Model for Evaluating Airline Service Quality: An Empirical Study in Cross-Strait Direct Flights." *Transportation Research Part A: Policy and Practice* 62:63–70.

- Laroche, Michel, Linda C Ueltschy, Shuzo Abe, Mark Cleveland, and Peter P Yannopoulos. 2004. "Service Quality Perceptions and Customer Satisfaction: Evaluating the Role of Culture." *Journal of International Marketing* 12 (3):58–85.
- Lee, Moontae, and David Mimno. 2014. "Low-Dimensional Embeddings for Interpretable Anchor-Based Topic Inference." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1319–28. Doha, Qatar: Association for Computational Linguistics.
- Li, Mimi. 2014. "Cross-Cultural Tourist Research: A Meta-Analysis." *Journal of Hospitality & Tourism Research* 38 (1):40–77.
- Li, Xiang Robert, Chengting Lai, Rich Harrill, Sheryl Kline, and Liangyan Wang. 2011. "When East Meets West: An Exploratory Study on Chinese Outbound Tourists' Travel Expectations." *Tourism Management* 32 (4):741–749.
- Li, Xinxin, and Lorin M. Hitt. 2008. "Self-Selection and Information Role of Online Product Reviews." *Information Systems Research* 19 (4):456–74.
- Liang, Ting-Peng, Xin Li, Chin-Tsung Yang, and Mengyue Wang. 2015. "What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach." *International Journal of Electronic Commerce* 20 (2):236–260.
- Liu, Ben Shaw-Ching, Olivier Furrer, and Devanathan Sudharshan. 2001. "The Relationships between Culture and Behavioral Intentions toward Services." *Journal of Service Research* 4 (2):118–129.
- Liu, Zhiwei, and Sangwon Park. 2015. "What Makes a Useful Online Review? Implication for Travel Product Websites." *Tourism Management* 47:140–151.
- Maiyaki, Ahmed Audu. 2013. "Moderating Effect of Individualism/Collectivism on the Association between Service Quality, Corporate Reputation, Perceived Value and Consumer Behavioural Intention." *Journal of Marketing & Management* 4 (1):1–20.
- Malhotra, Naresh K, Francis M Ulgado, James Agarwal, and Imad B Baalbaki. 1994. "International Services Marketing: A Comparative Evaluation of the Dimensions of Service Quality between Developed and Developing Countries." *International Marketing Review* 11 (2):5–15.
- Mazaheri, Ebrahim, Marie-Odile Richard, and Michel Laroche. 2011. "Online Consumer Behavior: Comparing Canadian and Chinese Website Visitors." *Journal of Business Research* 64 (9):958–965.
- Mazanec, Josef A, John C Crotts, Dogan Gursoy, and Lu Lu. 2015. "Homogeneity versus Heterogeneity of Cultural Values: An Item-Response Theoretical Approach Applying Hofstede's Cultural Dimensions in a Single Nation." *Tourism Management* 48:299–304.
- Meng, Juan, and Venkatapparao Mummalaneni. 2011. "Cultural Influences on Web Service Quality Perceptions of E-Retailing Consumers." *Journal of Marketing Channels* 18 (4):303–326.
- Mimno, David, Hanna M Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. "Optimizing Semantic Coherence in Topic Models." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272. Edinburgh, United Kingdom: Association for Computational Linguistics.
- Nath, Prithwiraj, James Devlin, and Veronica Reid. 2016. "Expectation Formation in Case of Newer Hotels: The Role of Advertising, Price, and Culture." *Journal of Travel Research* 55 (2):261–275.
- Parasuraman, Ananthanarayanan, Valarie A Zeithaml, and Leonard L Berry. 1988. "Servqual: A Multiple-Item Scale for Measuring Consumer Perc." *Journal of Retailing* 64 (1):12.

- Parasuraman, Anantharathan, Valarie A Zeithaml, and Leonard L Berry. 1985. "A Conceptual Model of Service Quality and Its Implications for Future Research." *Journal of Marketing* 49 (4):41–50.
- Park, Jaram, Young Min Baek, and Meeyoung Cha. 2014. "Cross-Cultural Comparison of Nonverbal Cues in Emoticons on Twitter: Evidence from Big Data Analysis." *Journal of Communication* 64 (2):333–354.
- Paswan, Audhesh K, and Gopala Ganesh. 2005. "Cross-Cultural Interaction Comfort and Service Evaluation." *Journal of International Consumer Marketing* 18 (1–2):93–115.
- Pearce, Philip L, and Mao-Ying Wu. 2016. "Tourists' Evaluation of a Romantic Themed Attraction: Expressive and Instrumental Issues." *Journal of Travel Research* 55 (2):220–232.
- Phillips, Paul, Stuart Barnes, Krystin Zigan, and Roland Schegg. 2017. "Understanding the Impact of Online Reviews on Hotel Performance: An Empirical Analysis." *Journal of Travel Research* 56 (2):235–249.
- Purnawirawan, Nathalia, Martin Eisend, Patrick De Pelsmacker, and Nathalie Dens. 2015. "A Meta-Analytic Investigation of the Role of Valence in Online Reviews." *Journal of Interactive Marketing* 31:17–27.
- Reimann, Martin, Ulrich F Lünemann, and Richard B Chase. 2008. "Uncertainty Avoidance as a Moderator of the Relationship between Perceived Service Quality and Customer Satisfaction." *Journal of Service Research* 11 (1):63–73.
- Reisinger, Yvette, and John C Crotts. 2010. "Applying Hofstede's National Culture Measures in Tourism Research: Illuminating Issues of Divergence and Convergence." *Journal of Travel Research* 49 (2):153–164.
- Roberts, Margaret E, Brandon M Stewart, and Edoardo M Airoidi. 2016. "A Model of Text for Experimentation in the Social Sciences." *Journal of the American Statistical Association* 111 (515):988–1003.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2017. "Stm: R Package for Structural Topic Models." <http://www.structuraltopicmodel.com>.
- Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4):1064–1082.
- Ryu, Sungmin, and Chul Woo Moon. 2009. "Long-Term Orientation As A Determinant Of Relationship Quality Between Channel Members." *The International Business & Economics Research Journal* 8 (11):1–9.
- Sabiote-Ortiz, Carmen M, Dolores M Frías-Jamilena, and J Alberto Castañeda-García. 2016. "Overall Perceived Value of a Tourism Service Delivered via Different Media: A Cross-Cultural Perspective." *Journal of Travel Research* 55 (1):34–51.
- Sheldon, Pavica, Philipp A Rauschnabel, Mary Grace Antony, and Sandra Car. 2017. "A Cross-Cultural Comparison of Croatian and American Social Network Sites: Exploring Cultural Differences in Motives for Instagram Use." *Computers in Human Behavior* 75:643–651.
- Sotiriadis, Marios D, and Cinà Van Zyl. 2013. "Electronic Word-of-Mouth and Online Reviews in Tourism Services: The Use of Twitter by Tourists." *Electronic Commerce Research* 13 (1):103–124.
- Sparks, Beverley A, and Victoria Browning. 2011. "The Impact of Online Reviews on Hotel Booking Intentions and Perception of Trust." *Tourism Management* 32 (6):1310–1323.
- Stamolampros, Panagiotis, and Nikolaos Korfiatis. 2018. "Exploring the Behavioral Drivers of Review Valence: The Direct and Indirect Effects of Multiple Psychological Distances." *International Journal of Contemporary Hospitality Management* 30 (8):Forthcoming.

- Steven, Adams B, Yan Dong, and Martin Dresner. 2012. "Linkages between Customer Service, Customer Satisfaction and Performance in the Airline Industry: Investigation of Non-Linearities and Moderating Effects." *Transportation Research Part E: Logistics and Transportation Review* 48 (4):743–754.
- Stump, Rodney L, and Wen Gong. 2017. "Social Networking Sites: An Exploration of the Effect of National Cultural Dimensions on Country Adoption Rates and Usage Patterns." *International Journal of Electronic Business* 13 (2–3):117–142.
- Suzuki, Yoshinori, John E Tyworth, and Robert A Novack. 2001. "Airline Market Share and Customer Service Quality: A Reference-Dependent Model." *Transportation Research Part A: Policy and Practice* 35 (9):773–788.
- Symitsi, Efthymia, Panagiotis Stamolampros, and George Daskalakis. 2018. "Employees' Online Reviews and Equity Prices." *Economics Letters* 162:53–55.
- Tam, Jackie LM, Jackie LM Tam, Piyush Sharma, Piyush Sharma, Namwoon Kim, and Namwoon Kim. 2016. "Attribution of Success and Failure in Intercultural Service Encounters: The Moderating Role of Personal Cultural Orientations." *Journal of Services Marketing* 30 (6):643–658.
- Tanford, Sarah, and Rhonda Montgomery. 2015. "The Effects of Social Influence and Cognitive Dissonance on Travel Purchase Decisions." *Journal of Travel Research* 54 (5):596–610.
- Tellis, Gerard J, and Joseph Johnson. 2007. "The Value of Quality." *Marketing Science* 26 (6):758–773.
- Thelen, Shawn T, and Terri Shapiro. 2012. "Predicting Negative Consumer Reactions to Services Offshoring." *Journal of Services Marketing* 26 (3):181–193.
- Tirunillai, Seshadri, and Gerard J Tellis. 2014. "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation." *Journal of Marketing Research* 51 (4):463–479.
- Torres, Edwin N, Xiaoxiao Fu, and Xinran Lehto. 2014. "Examining Key Drivers of Customer Delight in a Hotel Experience: A Cross-Cultural Perspective." *International Journal of Hospitality Management* 36:255–262.
- Tseng, Aihua. 2017. "Why Do Online Tourists Need Sellers' Ratings? Exploration of the Factors Affecting Regretful Tourist e-Satisfaction." *Tourism Management* 59:413–424.
- Van Vaerenbergh, Yves, Chiara Orsingher, Iris Vermeir, and Bart Larivière. 2014. "A Meta-Analysis of Relationships Linking Service Failure Attributions to Customer Outcomes." *Journal of Service Research* 17 (4):381–398.
- Vermeulen, Ivar E, and Daphne Seegers. 2009. "Tried and Tested: The Impact of Online Hotel Reviews on Consumer Consideration." *Tourism Management* 30 (1):123–127.
- Voss, Christopher A, Aleda V Roth, Eve D Rosenzweig, Kate Blackmon, and Richard B Chase. 2004. "A Tale of Two Countries' Conservatism, Service Quality, and Feedback on Customer Satisfaction." *Journal of Service Research* 6 (3):212–230.
- Weiermair, Klaus. 2000. "Tourists' Perceptions towards and Satisfaction with Service Quality in the Cross-Cultural Service Encounter: Implications for Hospitality and Tourism Management." *Managing Service Quality: An International Journal* 10 (6):397–409.
- Winsted, Kathryn Frazer. 1997. "The Service Experience in Two Cultures: A Behavioral Perspective." *Journal of Retailing* 73 (3):337–360.
- Ye, Ben Haobin, Hanqin Qiu Zhang, and Peter P Yuen. 2013. "Cultural Conflicts or Cultural Cushion?" *Annals of Tourism Research* 43:321–349.

Index of Tables and Figures

Table 1: Sample characteristics

Variable	
Total Number of Airlines	489
Total Number of Reviews	557,208
Total Number of Users	376,519
Total Number of Countries (Reviewers)	203
Total Number of Countries (Airlines)	147
Number of reviews in English	254,424
Average Length of English Review Text (characters)	560.27

Table 2: Descriptive statistics

<i>Ratings per Service Category</i>	<i>M</i>	<i>SD</i>
Overall Score	3.68	1.29
Seat Comfort	3.46	1.11
Customer Service	3.75	1.34
Cleanliness	3.94	1.03
Food and Beverages	3.32	1.27
Legroom	3.45	1.14
Inflight Entertainment/ Wi-Fi	3.01	1.47
Value for Money	3.66	1.23
Check-in and Boarding	3.81	1.25
<i>Additional Variables</i>	<i>M</i>	<i>SD</i>
Flight Distance (Kilometers)	4215.10	3784.09
Reviewer's Level of Contribution	3.83	1.88
Overall Score for Economy Class	3.64	1.29
Overall Score for Premium Economy	3.92	1.22
Overall Score for Business Class	3.98	1.21
Overall Score for First Class	4.00	1.26

Note: Rating scale is for 1(minimum) to 5(maximum) stars level of satisfaction.

Table 3: Results of ordered logistic regression for each aspect of the rating score with the Hofstede dimensions controlling for flight distance, reviewer expertise and cabin class upgrades.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Power Distance	0.003*** (0.000)	0.008*** (0.000)	0.004*** (0.000)	0.007*** (0.000)	0.005*** (0.000)	0.008*** (0.000)	0.003*** (0.000)	0.003*** (0.000)	0.004*** (0.000)
Individualism	-0.003*** (0.000)	-0.001*** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.003*** (0.000)	-0.000 (0.000)	-0.007*** (0.000)	-0.001*** (0.000)	0.000 (0.000)
Uncertainty Avoidance	-0.001*** (0.000)	-0.005*** (0.000)	-0.000 (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.008*** (0.000)	-0.001*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)
Masculinity	-0.001*** (0.000)	-0.003*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	0.001** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)
Long Term Orientation	-0.001* (0.000)	0.003*** (0.000)	-0.001*** (0.000)	0.001** (0.000)	-0.001** (0.000)	0.004*** (0.000)	-0.005*** (0.000)	0.001*** (0.000)	-0.001*** (0.000)
Indulgence	0.000 (0.000)	0.001*** (0.000)	0.002*** (0.000)	0.004*** (0.000)	-0.000 (0.000)	0.002*** (0.000)	0.005*** (0.000)	-0.000 (0.000)	0.002*** (0.000)
Flight Distance	0.104*** (0.003)	0.114*** (0.003)	0.112*** (0.003)	0.123*** (0.004)	0.316*** (0.004)	0.137*** (0.003)	0.603*** (0.003)	0.053*** (0.003)	0.099*** (0.004)
Reviewer Expertise	0.061*** (0.002)	0.026*** (0.002)	0.033*** (0.002)	0.002 (0.002)	0.009*** (0.002)	0.035*** (0.002)	0.049*** (0.002)	0.036*** (0.002)	0.016*** (0.002)
Premium Economy	0.416*** (0.018)	0.881*** (0.019)	0.424*** (0.019)	0.437*** (0.021)	0.476*** (0.021)	1.175*** (0.019)	0.607*** (0.019)	0.069*** (0.019)	0.381*** (0.021)
Business Class	0.458*** (0.011)	1.460*** (0.012)	0.635*** (0.012)	0.644*** (0.014)	0.957*** (0.013)	1.939*** (0.013)	0.695*** (0.012)	0.160*** (0.012)	0.559*** (0.014)
First Class	0.644*** (0.024)	1.811*** (0.026)	0.874*** (0.027)	0.690*** (0.029)	1.105*** (0.029)	2.174*** (0.027)	1.141*** (0.026)	0.182*** (0.025)	0.824*** (0.030)
<i>McFaden R²</i>	0.36	0.38	0.36	0.35	0.36	0.39	0.39	0.36	0.34
<i>AIC</i>	1042181	938963	952187	696674	763882	932399	907991	966501	748988
<i>LL</i>	-521076	-469467	-476078	-348322	-381926	-466184	-453981	-483236	-374479
<i>Observations</i>	359,424	334,263	334,668	265,998	254,234	333,994	301,051	331,334	266,754

Note: Standard errors in parenthesis. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, Model specifications for dependent variable: (1) Overall Score, (2) Seat Comfort, (3) Customer Service, (4) Cleanliness, (5) Food and Beverage, (6) Legroom, (7) Inflight entertainment / Wi-Fi, (8) Value for Money, and (9) Check-in and Boarding

Table 4: Results of ordered logistic regression for each aspect of the rating score with the cultural distance between the country of the reviewer and the country of the airline controlling for flight distance, reviewer expertise and cabin class upgrades.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cultural Distance	-0.027*** (0.003)	0.013*** (0.003)	-0.046*** (0.003)	-0.001 (0.003)	-0.001 (0.003)	0.016*** (0.003)	-0.069*** (0.003)	0.081*** (0.003)	-0.056*** (0.003)
Flight Distance	0.044*** (0.003)	0.035*** (0.004)	0.080*** (0.004)	0.066*** (0.004)	0.261*** (0.004)	0.067*** (0.004)	0.579*** (0.004)	-0.022*** (0.004)	0.072*** (0.004)
Reviewer Expertise	0.059*** (0.002)	0.023*** (0.002)	0.029*** (0.002)	-0.003 (0.002)	0.007** (0.002)	0.031*** (0.002)	0.048*** (0.002)	0.033*** (0.002)	0.012*** (0.002)
Premium Economy	0.467*** (0.018)	0.980*** (0.019)	0.452*** (0.020)	0.507*** (0.022)	0.539*** (0.021)	1.286*** (0.019)	0.642*** (0.019)	0.152*** (0.019)	0.418*** (0.021)
Business Class	0.400*** (0.012)	1.409*** (0.013)	0.594*** (0.013)	0.597*** (0.015)	0.938*** (0.015)	1.900*** (0.014)	0.687*** (0.013)	0.102*** (0.013)	0.519*** (0.015)
First Class	0.508*** (0.025)	1.703*** (0.027)	0.792*** (0.028)	0.618*** (0.030)	1.008*** (0.029)	2.112*** (0.028)	1.089*** (0.027)	0.147*** (0.026)	0.797*** (0.031)
<i>McFaden R²</i>	0.42	0.43	0.41	0.40	0.41	0.43	0.44	0.41	0.40
<i>AIC</i>	961026	864056	878645	644736	703765	860972	833048	888949	692672
<i>LL</i>	-480503	-432018	-439313	-322358	-351872	-430476	-416514	-444465	-346326
<i>Observations</i>	328,281	305,263	305,618	243,539	232,060	305,026	272,793	302,648	244,223

Note: Standard errors in parenthesis. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, Model specifications for dependent variable: (1) Overall Score, (2) Seat Comfort, (3) Customer Service, (4) Cleanliness, (5) Food and Beverage, (6) Legroom, (7) Inflight entertainment / Wi-Fi, (8) Value for Money, and (9) Check-in and Boarding

Table 5: Labels, distribution and FREX score for the top 7 keywords in the topic solution.

#	Topic Label	Prop. (%)	Top 7 FREX words
1	Delays	11.56	delay, minute, late, hour, due, connection, departure
2	Staff (Praise)	9.24	friendly, helpful, professional, efficient, clean, courteous, staff
3	Seating (Critique)	6.68	row, front, seat, uncomfortable, aisle, exit, window
4	Value for Money	6.59	good, overall, value, food, perfect, money, bit
5	Traveler Experience	5.97	many, best, domestic, past, job, need, possible
6	Staff (Critique)	5.94	attendant, water, stewardess, old, steward, terrible, passenger
7	Check-in/Boarding	5.91	check, luggage, hand, security, easy, queue, allowance
8	Mode of Travel	5.82	trip, return, direct, stop, family, home, non
9	Inflight Entertainment	5.42	entertainment, movie, screen, inflight, selection, average, quality
10	Service Failure/Response	5.22	phone, email, credit, call, agent, card, change
11	Staff Assistance	5.16	crew, cabin, holiday, special, child, nothing, much
12	Ticket Cost	5.01	low, price, budget, cheap, cost, cheaper, fare
13	Food/Beverage	4.86	coffee, snack, meal, drink, sandwich, free, tea
14	Baggage Policy	3.46	bag, carry, charge, line, checked, fee, item
15	Business Class	3.03	flat, lounge, business, class, bed, sleep, access
16	Wi-Fi Connectivity	2.64	wifi, free, board, duty, app, hassle, access
17	Economy Premium	2.16	economy, premium, comfort, upgrade, difference, section, worth
18	Legroom (Praise)	2.06	leg, plenty, extra, lot, comfortable, seating, bit
19	Legroom(Critique)	1.86	room, space, enough, tall, foot, amount, bit
20	Frequent Flyer	1.41	flyer, mile, world, member, traveller, point, part

Table 6: The effect of cultural differences on the ranking of airlines based on online reviews.

Airline	Raw Rank(A)	Airline	Standardized Rank(B)	Δ Rank(A-B)
Azul	1	Emirates	1	-3 ↓
Singapore Airlines	2	Southwest Airlines	2	-1 ↓
Emirates	3	Singapore Airlines	3	+2 ↑
Southwest Airlines	4	Azul	4	+2 ↑
Jet2.com	5	Jet2.com	5	0
Aeroflot	6	JetBlue	6	-6 ↓
JetBlue	7	Qatar Airways	7	+1 ↑
Qatar Airways	8	Virgin Atlantic Airways	8	+1 ↑
Virgin Atlantic Airways	9	Turkish Airlines	9	+1 ↑
Turkish Airlines	10	Lufthansa	10	+1 ↑
Avianca	11	KLM	11	+3 ↓
KLM	12	Aeroflot	12	+1 ↑
Lufthansa	13	Cathay Pacific	13	+3 ↑
Jet Airways	14	Avianca	14	-1 ↓
Cathay Pacific	15	Jet Airways	15	+2 ↑
LATAM Airlines	16	Qantas	16	-3 ↓
Qantas	17	Delta Air Lines	17	+1 ↑
Delta Air Lines	18	Norwegian	18	+1 ↑
Norwegian	19	LATAM Airlines	19	+1 ↑
Aerolineas Argentinas	20	Transavia	20	>-1↓

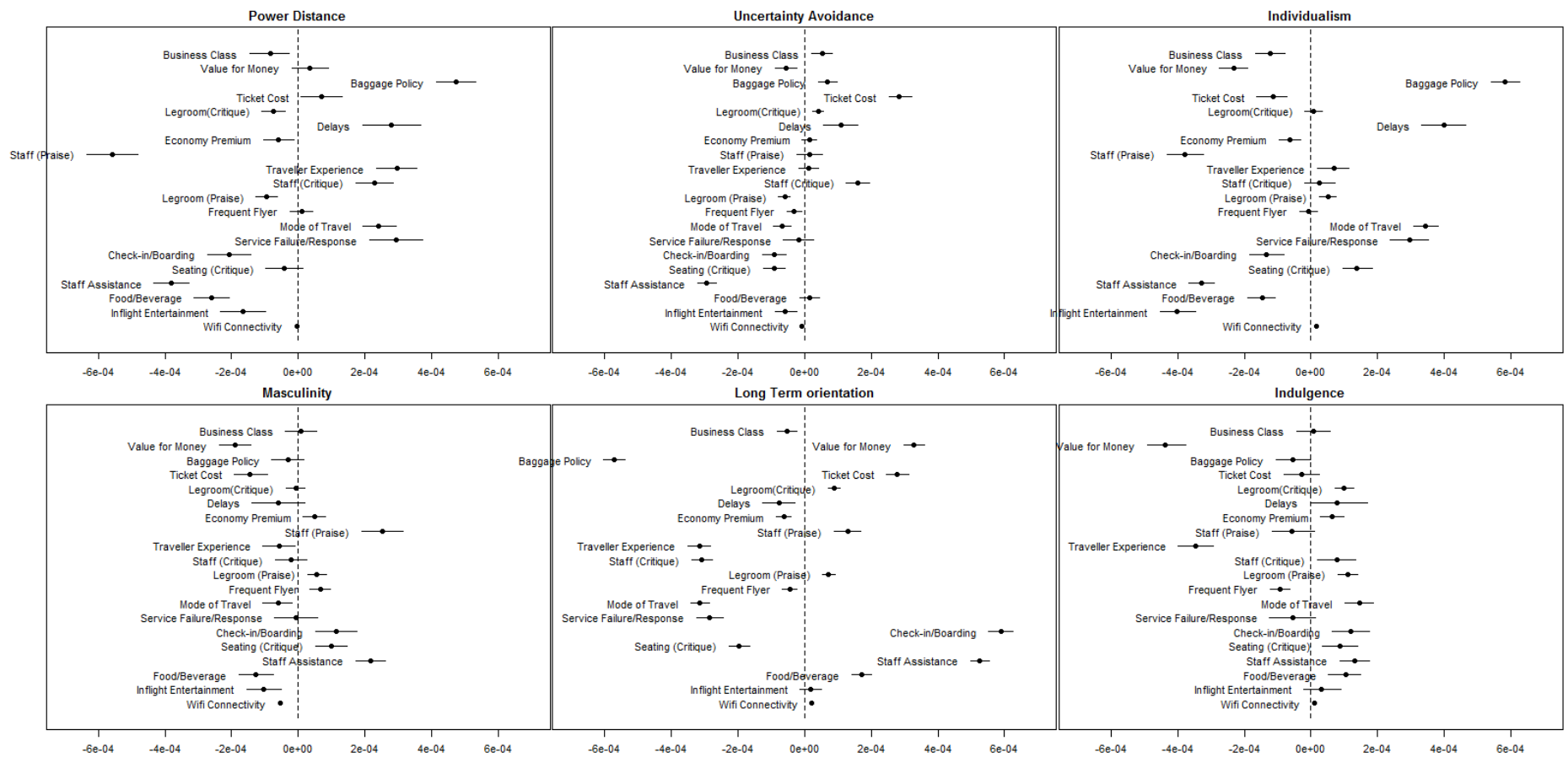


Figure 1: Proportional odds on topic prevalence for each of the Hofstede dimensions. Zero effects are marked with a dotted line. For each figure, topics are plotted across the continuum (low to high) of the values of the respective Hofstede dimension. Horizontal axis shows the increase (decrease) in topic prevalence for the plotted topic per unit of each Hofstede dimension.

Appendix: Top 10 countries in our dataset and their corresponding Hofstede values

	Total Number of Reviews	Power Distance	Uncertainty Avoidance	Individualism	Masculinity	Long-Term Orientation	Indulgence
UNITED STATES	70,054	40	46	91	62	26	68
UNITED KINGDOM	50,578	35	35	89	66	51	69
ITALY	22,069	50	75	76	70	61	30
FRANCE	20,543	68	86	71	43	63	48
AUSTRALIA	19,890	36	51	90	61	21	71
BRAZIL	18,680	69	76	38	49	44	59
ARGENTINA	16,278	49	86	46	56	20	62
CANADA	14,903	39	48	80	52	36	68
GERMANY	12,222	35	65	67	66	83	40
SPAIN	11,399	57	86	51	42	48	44