

## Reviewer #5 Response

Thank you for your review observations. I will quote each observation and respond to it while indicating my corrections in the text. Aside from these corrections, we also performed a general grammar check and proofread to the new manuscript, including unchanged sections. Because of comments by reviewers, we redid the experiment with more balanced data, so the results shown are different in their quantitative values. However, our conclusions and hypothesis testing remained largely unchanged by this. We also added a new research rationale and focus, which meant new hypotheses were added to the manuscript. The new manuscript focuses on how these populations interact with the high-standard spirit of Japanese hospitality and service called Omotenashi. In this environment, will the cultural differences be strong enough that their satisfaction stems from different factors?, or will both populations be largely satisfied by service and staff related terms?

This study used Ctrip as the data base. However, majority users of Ctrip is Mainland Chinese, but Chinese-speaking tourists can also be Taiwanese, Malaysian Chinese, Singaporean Chinese and Hong Konger too. Therefore I suggest the authors change the term Chinese-speaking tourists to Mainland Chinese tourists.

Thanks for pointing this information out. We pointed this out in the text.

In literature review, the authors listed three hypothesis relating to Chinese and English speaking tourists in 3.1 and then a separate section 3.2 to explain the difference between Chinese and Western tourists. I would suggest the authors combine these two sections so as to give a more comprehensive view on the setting of the three hypothesis.

We respectfully disagree with this suggestion since the content of section 3.1 and 3.2 are very different. While 3.1 sets the premise of differences in customer satisfaction and the reasoning behind them existing, 3.2 discusses previous literature on specific behaviors and cultures of each kind of tourist. Since we added a new section 3.1, these two sections are now 3.2 and 3.3 respectively.

In formular (1), what are alpha, x and y?

We added a description of the formula to the manuscript, as well as a short description of the SVM algorithm to make it more understandable.

In the Data Collection, can the authors explain what kinds of hotels (category, Hotel types, or star rating) were extracted? Are the same hotels extracted from both Ctrip and Tripadvisor? How many hotels are there? I believe different types of hotels would have generated different satisfaction level and perspectives. Please

provide more details about these hotels.

In regards to this point, we apologize and would like to explain. In the original manuscript, the experiments were done while the data was still being collected further. After finishing the manuscript, the original data had changed. Not only this, but we realized our mistake from this and another comment from a different reviewer. Indeed the data should reflect the same kinds of hotels, the same kinds of categories. Since we did not have these data, we decided to match both English and Chinese comments to come from exactly the same hotels, matching by name. This reduced the number of comments on each language, but it made the data fairer. We don't have data on what type of hotel or star rating they had since this data wasn't available to us, but we can be sure they are the same 580 hotels for both kinds of reviews.

Because of this change in datasets, we also redid the entire experiment, calculating the frequencies and RBO again accordingly.

The time period resulted to be from July 2014 to July 2017. The price range was from 2000 yen per night, to 188,000 yen a night.

We added this paragraph to the Data collection section:

"However, in order to make the data and comparisons we draw from each of these datasets fair, we filtered both databases to only contain reviews from hotels that were in both datasets, using their English names to do a search match. We also filtered them to be in the same date range and cut off reviews outside of each other's date ranges. After filtering, we found that the number of hotels in common in the data collected was 580, and that the common date range for reviews was from July 2014 to July 2017. We also found that the price for a night in these hotels ranges from low priced capsule hotels at 2000 yen per night, to high end hotels 188,000 yen a night as the far ends of the bell curve. Within this range, from *Ctrip* there was 49,561 reviews comprised of 105,076 sentences, and from *TripAdvisor* there was 41,611 reviews comprised of 352,022 sentences.

We found that after filtering the data, the number of reviews was similar for both English and Chinese reviews, but that English reviews tend to be longer in general."

In Table 2, I am quite surprise to see that the last item "Chinese person" can be the top keywords with only 16 occurrences. Can the authors explain why there are so few negative Chinese keywords?

Since we repeated the experiment, the frequencies have changed, however this keyword still only has 16 occurrences. The lack of negative keywords comes from a peculiarity of the Chinese reviews from Ctrip. Most of the reviews point out the positive extensively, while only a few sentences are negative. It's a limitation of our paper that the negative keywords depend on the sample taken

at the beginning, since we could not label more manually at the time.

Regarding the keyword "Big", this is a very general word and can be describe as "Big" in size, "Big in feelings, or could be negative too "big disappointment". Can the authors explain further how this can be avoided in your data analysis?

One thing to note is that using SVM, we labeled sentences as positive or negative, so negative usage of the word when counting the frequencies is negligible. In addition to this, we also had the same question when we saw our results the first time, and took a small sample to observe the usage of the word. However we didn't deem it appropriate in the text at first, but with this comment and thinking it over once again, we have added the next paragraph to the manuscript:

“From these two first ranking keywords (“big” and “clean”), we can assert that room quality is the most critical satisfaction factor for Chinese customers. However, the word “big” is widespread in the language and could be used in different contexts to point out positive and negative aspects. Since we are interested in how the word is a keyword of positive sentences, we took a small sample of the positively classified sentences that contained the word “big” in Chinese. A team of Chinese speaking lab members then pointed out the usage of the word. Generally speaking, it was used for the spaciousness of rooms, bathtubs, and communal bath facilities.”

For the managerial and environmental keywords, what criteria were used for classifying these keywords?

We explain the difference between managerial and environmental factors in different sections of the manuscript. However, since it might still be unclear, we specified the method for classifying the keywords in the Results section, adding the following text:

“We manually labeled the top 625 keywords of each language into either managerial or environmental by considering how the word would be used when writing a review. If the word implies an issue that could be solved or managed by the hotel staff or management, we consider it managerial. If the word is describing factors that are unchangeable by the staff or management, we consider them environmental. The interpretation of these keywords is shown in the Tables 8 and 9.”

We also added the tables that were previously in the appendix clearing this issue into the main text for readability.

Editorial issues The Appendix is actually part of the manuscript. This manuscript contains too many Tables. First some was mentioned in the manuscript but appeared in Appendix. For example line 306, Table 7 and Section B was mentioned. What is Section

B? Please move all relevant tables to the manuscript but not in Appendix. I would suggest the authors to consider to either combine some of them or not to mention them in the manuscript. Besides, the numbering of the tables was not in sequence in the manuscript. Line 360, after Table 3 and 4, it suddenly jumps to Table 10.

We moved some tables that we deemed important out of the appendix, but other tables got removed as per this request. We also joined the tables with the top keyword results as per this and another reviewer's request.