Please, talk about it! When hotel popularity boosts preferences

ABSTRACT

Many consumers post on-line reviews, affecting the average evaluation of products and services. Yet, little is known about the importance of the number of reviews for consumer decision making. We conducted an on-line experiment (n=168) to assess the joint impact of the average evaluation, a measure of quality, and the number of reviews, a measure of popularity, on hotel preference. The results show that consumers' preference increases with the number of reviews, independently of the average evaluation being high or low. This is not what one would expect from an informational point of view, and review websites fail to take this pattern into account. This novel result is mediated by demographics: young people, and in particular young males, are less affected by popularity, relying more on quality. We suggest the adoption of appropriate ranking mechanisms to fit consumer preferences.

1. INTRODUCTION

The way in which both software developers and consumers use the Internet is continuously changing towards an increasing management of user-generated content. This "collaborative" vision of the web, promoting a place where users can interact and share information, was coined about a decade ago with the term Web 2.0. Examples of Web 2.0 include social networks, video-sharing sites, fora, wikis, blogs, and other sites managing user-generated information. In this dynamic world of on-line marketing, the traditional influence of word-of-mouth has been fiercely amplified by the impressions from consumers posting their experience with products and services in social media websites.

Since Amazon.com Inc. started posting customer ratings and product reviews in 1995, most on-line businesses have realized that allowing customers to post reviews can increase sales and help suppliers

identify problems with their products and services. These information tools are being used by consumers who increasingly search and read comments and reviews from peers, facilitating choices and purchase decisions. In its last Trust Barometer 2013, the public-relations firm Edelman asked survey respondents across 20 countries how credible the information about a company was, depending on the informer. A total of 61% of respondents attributed high credibility to "*a person like yourself*", compared to only 49% to "*regular employees*" and 40% to "*the company's CEO*". A previous survey conducted in 2011 by the public-relations firm Weber Shandwick, found that traditional word-of-mouth (88%) and on-line reviews (83%) ranked as top factors, being "very" or "somewhat" influential on consumer perceptions about companies.

Within the service sector, travel is one of the fastest growing industries in e-commerce spending. ComScore Inc, a global research firm that tracks on-line traffic, reported that the travel category attracted 124 million visitors in January 2012, with an increase of 8% with respect to the previous year.

All the above phenomena combine in the form of travel review websites, revolutionizing the manner in which word-of-mouth opinions and recommendations on holiday destinations can be discussed and disseminated (Mauri and Minazzi, 2013; Litvin *et al.,* 2008).

Some review websites have become important obligatory points of passage (Yacouel and Fleischer, 2012). An example of the culmination of such on-line commentaries is the creation of ranking lists, such as the Trip Advisor Popularity Index. This offers a clear numbering system which instantly signals a hotel's level of quality and service to satisfy consumers (Jeacle and Carter, 2011). In this paper we focus our attention on popularity, in terms of the number of reviews written by people, to understand how it affects consumer decision making and how it interacts with consumer's on-line evaluation, a widely used measure of quality (Abrate *et al.*, 2011; Hu *et al.*, 2006; Koh *et al.*, 2010).

2. THEORY

2.1 The Relevance of Consumer Reviews

The importance of reviews is rather different in experience and search goods. Experience goods are products that require sampling or purchasing in order to evaluate the product quality. In this case there is a need to use one's senses to evaluate quality. Examples of experience goods include music (Bhattacharjee *et al.,* 2006; Nelson, 1970) and wine (Klein, 1998). Search goods, by contrast, are those where consumers can obtain relevant information on product quality prior to purchase. Examples here include cameras (Nelson, 1970) or medication (Weathers *et al.,* 2007). The dominant attributes of an experience good are compared or evaluated subjectively and with more difficulty (Huang *et al.,* 2009). However, the relevant characteristics of a search good are that it can be evaluated and compared easily in a more objective manner, without buying or sampling the product. Because the Internet enables consumers to learn from the experiences of others and to gather product information that is often hard to obtain in off-line settings (Klein, 1998; Lynch and Ariely, 2000), all attributes tend now to be searchable at low cost. This reduces the difference between search and experience goods. This "merging process" was initially highlighted by Alba et al. (1997), who suggested that all products involve a bundle of search and experience attributes. Hotel rooms fit perfectly in this framework. Although traditionally considered as experience goods due to the difficulties in gathering precise information, they are now moving toward search goods. Now, for example, travellers can judge if a room is suitable beforehand, and look for information on-line through rating sites (Tse, 2003).

Nonetheless, there are still some differences between search and experience goods. For search goods, the content and detail of the review itself is considered crucial (Jimenez and Mendoza, 2013; Mudambi and Schuff, 2010). The idea is that an in-depth review with search goods is highly diagnostic. Whilst, the social weight provided by the number of comments is also an important factor affecting consumer choice for experience goods.

When travellers book hotels on-line, they are typically provided with a list of relevant hotels. While presenting hotels in a list format seems appropriate in order to organize the information, it creates a new (spurious) attribute for them: their positioning. Spoerri (2008) showed that only information placed high in the list is considered relevant. Further, the relevance of the information decreases exponentially when presented in lower positions. Breese et al. (1998) confirmed this exponential decay of attention. One of the first efforts to model ranking by popularity was undertaken by Chen (2009), while the importance of ranking was again stressed more recently by Filieri and Mcleay (2014).

It has been shown that positive reviews have an effect in increasing the number of bookings and the economic results (Chavaler *et al.*, 2006; Godes and Mayzlin, 2009; Ye *et al.*, 2009). However, the actual number of reviews should also be taken into consideration when hotel travel websites present their rankings of hotels. Of primary importance, is also the role of negative reviews. While several studies have examined the content of negative consumer reviews on the web and their effects on perceived company reliability (Chatterjee, 2001; Noort and Willemsen, 2012; Sen and Lerman, 2007), only a few studies have to date discussed the controversial effect of negative popularity on preferences (Berger *et al.*, 2010; Khare *et al.*, 2011; Vermeulen and Seegers, 2009). The findings of these studies suggest that negative reviews decrease consumers' attitudes towards that alternative but increase consumers' awareness towards the same alternative, leaving for further research the overall effect on preference**.**

2.2 Reducing uncertainty or following the crowd

According to social comparison theory (Festinger, 1954), individuals possess a drive to compare themselves to other people. This can lead them to select popular alternatives in the belief that the majority is right (Denrell and Le Mens, 2012). Consumers look to other consumers for social clues as

their choices may be seen as a statement about individual values and taste (Mudambi and Schuff, 2010). In our analysis, we use the number of reviews of a hotel as a proxy for popularity, and we consider consumer's probability to post a review on the travel websites constant across hotels. These assumptions are considered reasonable since the majority of travel review websites only allow the posting of one review per transaction after the check-out. Other studies have applied a similar approach in the literature (Ye *et al.*, 2009)

There are at least two possible explanations as to why people might prefer to see a number of reviews. First, a large number of reviews might lead consumers to feel more sure of their purchase decision. When more reviews are present, consumers increase their behavioral intention because they perceive them to be more informative (Park *et al.*, 2007; Petty and Cacioppo, 1984). This then reduces the uncertainty and the perceived risk (Klein, 1998). Another possible explanation, when people tend to go with the crowd, is that the hotel experience is a "conspicuously" consumed service, hence social norms could also be at play. Intuitively, an individual who believes a popular alternative to be poor might still choose that alternative anyway because it is popular. Sociologists would distinguish between the normative and informational facets of social influence. The former compels a person to do as others do so as to conform to their expectations. The latter leads individuals to accept the information obtained from others as evidence about reality (Deutsch and Gerard, 1955). In this paper, we use the term "normative" according to the above definition. Review websites, such as Trip Advisor, generally rank their hotels based on informational criteria. This considers the number of reviews as a measure of the reliability of the different evaluations. In principle, one can determine whether people endorse popular choices just to go with the others or whether to reduce informational uncertainty. In the hotel scenario for example, when a high number of reviews is present but the on-line reputation is low, popularity would still boost preferences if people simply want to go with the crowd. If, on the contrary, people believe that the number of reviews has only an informative effect, then for hotels with a poor

reputation a high number of reviews would have a negative effect on preference. In this instance, the high number of reviews being a guarantee that the hotel is bad. These considerations lead to a first general hypothesis for this paper:

H1. Popularity, measured through the number of reviews, affects people's preferences

In addition, based on the theoretical background presented above, we can also derive two alternative hypotheses. On the one hand, according to informational social theory and what on-line review websites generally do, we can expect that:

H2a. Because of informational social influence, the impact of popularity is expected to be positive when quality, measured as the average online rating, is high and negative when quality is low

On the other hand, the normative facets of social influence would lead us to predict that:

H2b. Because of normative social influence, the impact of popularity should be positive, regardless the level of quality

If the two effects presented above moderate each other we would expect that:

H2c. A high number of reviews increases the ranking of well reputed alternatives more than of less reputed ones

2.3 The cost and value of seeking unbiased reviews and their effects on individual differences

Traditional media normally use professional experts to review products or services. While prior evidence suggests that expert reviews are more persuasive than non-expert ones (Petty and Cacioppo, 1984), information delivered by a non-marketer source has been shown to be credible (Herr *et al.*, 1991; Sparks *et al.*, 2013). On the Internet, many reviews are consumer generated and fellow opinions are considered more trustable, favouring the increasing demand for unbiased information (Mauri and Minazzi, 2013).

Reading reviews easily accessible to all Internet users may help consumers in choosing (Dellarocas, 2003). However, seeking information is both costly and time consuming and there are trade-offs between the perceived cost and the benefit of search (Stigler, 1961) and between the invested effort and the accuracy of the decision (Johnson and Payne, 1985). Consumers tend to search for information until the marginal cost equals the marginal benefit (Huang *et al.,* 2009). For these reasons, they are willing to use numerical content ratings to save cognitive resources and to reduce energy expenditure. More formally, Sirakaya and Woodside (2005) contend that since the purchase of a tourism service is extensive, complex, and risky, the decision process occurs in stages. This decision is described as a funnel-like process in which consumers eliminate options from the "total set" (comprised of all available options) to construct a "consideration set", and then form a smaller "choice set" from which they ultimately choose (Jones & Chen, 2011; Sirakaya & Woodside, 2005). Hotel search starts by entering the desired destination and dates of travel which results in a list of available hotels (i.e., the total set). In rating websites like TripAdvisor.com, the search continues by further eliminating irrelevant hotels according to their on-line rating and the number of reviews, resulting in a smaller "consideration set". Some of the hotels in the consideration set are subjected to further examination by clicking on their icon and receiving additional information (such as exact location, staff, cleanness, price and detailed reviews of consumers). The outcome of this stage is a smaller "choice set" from which the final decision of which hotel to book is then made. This theory has

received support from a recent empirical study (Jones & Chen, 2011), which found that hotel consideration sets are comprised of 10 hotels on average, whereas choice sets average about 4. This funnel-like process helps consumers choose by forming a smaller and more comparable set of alternatives at each stage.

The star rating has been shown to serve as a numeric clue for the review content, especially for users familiar with the Internet (Kulkarni *et al.*, 2012; Poston and Speirer, 2005), The reported average star ratings are especially useful for experience goods, where review depth has been shown to be less relevant (Mudambi and Schuff, 2010). In addition, the use of just one important attribute was shown to be positively correlated with the value of saving time and cognitive resources (Payne *et al.,* 1996). In consumer research, people were shown to have different invested times when evaluating products or services (Okada, 2005). Meyers-Levy (1988) argued that males had a tendency not to process all available information as a basis for judgment, taking decisions more quickly than females. In contrast, females rely on a broad variety of information and they usually attempt an elaboration of all available information unless they are restricted by memory constraints. In the hotel and food industry, decision-making processes were shown to be more time consuming among females and high age groups, while income seems not to play an effect (Barber *et al.*, 2006; Han *et al.*, 2009). Based on the above arguments and on the evolutionary psychology literature on impulsiveness (Cross *et al.*, 2011; Wilska, 2003), we argue that there is a demographic effect in the trade-off between using multiple pieces of information and saving time and cognitive effort. More formally, to highlight how heterogeneous people are in these dimensions, we hypothesize that:

H3: People from different ages and gender will use different strategies according to their willingness to invest time. This may influence the way people are affected by popularity

3. METHOD

An on-line experiment was conducted to test the above hypotheses. We did not resemble the whole consumer decision making process but we just replicated the key step from the total set to the consideration set, as discussed in Jones and Chen (2011) and Sirakaya and Woodside (2005). Nonetheless, measuring the effect of popularity (number of reviews) for good and bad quality scores is important, as the ranking of review websites such as TripAdvisor.com is based on these two variables.

3.1 Participants

We targeted on-line consumers drawn from an Italian web community with an email invitation. The e-mail stated that two researchers were conducting a study on on-line reviews, and that participants had a chance of winning a 25 euros mobile phone voucher. The overall response rate was 60%, above the usual threshold found in tourism and hospitality studies (Hung and Law, 2011). As in Sparks et al. (2013), we elicited the familiarity with hotel bookings and with online reviews. Most participants (91%) had experience with booking hotels online, most (87%) also indicated that they relied on reviews when making a hotel booking, and a majority (65%) said that they often or always used the Internet to actually book their hotel. We reasonably assumed that this sample, well balanced between males and females, was suitable to represent the online users of travel websites.

Participants accessed and took the on-line experiment through a link that was sent to them from December 2012 to January 2013. The financial incentive was sent to three randomly extracted respondents.

Of the 168 individuals who participated in the on-line experiment, 161 respondents passed data-quality checks. The excluded respondents consisted of those who failed to pass the Kendall's tau test for consistency in the conjoint exercise (see section *Analysis*). Table 1 reports the demographics of the respondents included in the analysis. Regarding the variable age, we categorized people in two groups,

young and non-young, based on a threshold at age 25. This threshold is in line with what is generally used for the adoption of new technologies in tourism marketing by consumers (Pagani, 2004; Wood, 2004).

[Table 1 here]

3.2 Apparatus

Visual stimuli, as suggested by Holbrook and Moore (1981), were used instead of the more traditional method of verbal descriptions of products and their relative attributes. Thus, we created illustrations adapted from the website TripAdvisor.com. Figure 1 shows an example of the cards presented to a respondent.

[Figure 1 here]

3.3 Procedure

Once participants accessed the website, they were informed that they were going to rank their preferences regarding hotels based on two dimensions: the number of reviews and the online reputation score which, as mentioned before, represented the constructs of popularity and quality respectively. Participants were told that they would evaluate 15 cards with different situations. Before assessing the 15 cards, the form elicited information on participants' demographical characteristics (gender and age) and on the familiarity with hotel bookings, as explained in detail above. After the whole process, but before the 'thank you' page, they were invited to provide comments, via an open-ended question, specifically on the clarity of the instructions and the strategies used to come up with a decision.

3.4 Design

A rank conjoint experiment, which measures the importance of the features of a good or service by asking respondents to rank different scenarios within the choice set (Green and Rao, 1971; Gustaffson

et al., 2001), was adopted for the framework of this study. The goal was to elicit the weight respondents give to quality (score/rating) and popularity (number of reviews) when browsing hotel comparison websites with the intention to book a hotel. Through careful consideration and examination of hotel comparison websites, we decided to operationalize quality and popularity at 7 levels (Table 2). For quality, we included all ratings from 2 to 5 in 0.5 unit increments, as half ratings are quite common on hotel comparison websites and the typical range of reviews goes from 2 to 5 (Mazzarol *et al.*, 2007). For popularity, we included a very large range of reviews, from 1 to 4096 through a geometric progression with common ration 4, to conform to empirical previous results where the effect of popularity on sales was shown not to be linear and of a similar range (Chevalier and Mayzlin, 2006; Forman and Batia Wiesenfeld, 2008).

[Table 2 here]

Using combinations of one level from each of these two attributes to characterize a scenario, there are 49 (7 x 7) different hotel profiles that could be generated. However, in order to ensure feasibility of the conjoint exercise, to maximize participation, and to avoid respondents' fatigue, we decided to expose respondents to subsets of 15 scenarios only. Therefore based on Dean and Voss (1999), we created a D-optimal block design for a full profile experiment: 15 different versions (blocks) of 15 different hotel profiles were generated from the 49 possible scenarios. Each respondent was randomly assigned to one version and was exposed to all 15 hotel profiles within the selected version. We considered this design to be appropriate, in relation to the sample size, to allow robust estimation of linear, quadratic, and cubic effects for quality, popularity, and their interaction term. Each respondent was exposed to all 15 scenarios at the same time, so that he or she could make appropriate comparisons and assign the rank to each profile, from 1 (most preferred scenario) to 15 (least preferred scenario). To

confirm the feasibility and simplicity of the task, none of the respondents complained about the length, complexity, or composition of the questionnaire in the final open-ended question.

3.5 Analysis

In the conjoint framework adopted for this project, we assume that a respondent's ranking of each scenario can be decomposed into the sum of contributions from the various attributes. For each attribute, the contribution is the part-worth associated to the level describing the scenario. In other words, the part-worth is the marginal utility of the attribute in the individual's ranking of the conjoint scenario. Main outcome of conjoint analysis is the estimation of the part-worth associated to each level of each attribute considered in the conjoint design.

We used hierarchical Bayes (HB) regression (*for review, see* Rossi *et al.,* 2005) to estimate the part-worths associated to the levels of the two attributes included in the study. This approach avoids potential estimation bias from unobserved preference heterogeneity in rank-conjoint by estimating a distribution of preferences for each parameter in the model. HB part-worths are estimated for each respondent (individual-level analysis). The coefficients from this regression model were also the basis for the estimation of the relative importance of the attributes of the study. HB coefficient estimation was conducted using the function *estimate.Rank.VaryingTasks* available in the package R-sw Conjoint (Demia Studio Associato, 2012), the only package we are aware of especially dedicated to the analysis of rank conjoint data.

To select the most appropriate model for the data, we tested a number of alternative formulations. While in most conjoint studies interaction levels effects are often disregarded as they add complexity to both the design and the analysis, we thought considering and testing them to be essential, as we believed that the preference associated to quality depends on popularity. In addition, for each partial utility we extended the standard linear specification, proposing a nested structure that allows for quadratic and cubic terms. In fact, we believed that preference might not necessarily be linearly

associated with quality and/or popularity. Model selection was based on the Akaike information criterion (AIC), a popular measure of the relative goodness of fit of a model (for review, see Bozdogan, 1987). At the core of the AIC there is the sum of squared residuals, where a residual is the difference between the observed rank of a conjoint scenario and its estimated rank. The model selection process led us to identify the following model as the best one: quality (quadratic), popularity (quadratic), interaction quality by popularity (linear).[1]

Therefore, conjoint preferences were modeled using a utility function that is assumed to be linear in the parameters of the form:

$$
\begin{aligned}
U &= \beta_0 + \beta_1 * quality + \beta_2 * quality^2 + \beta_3 * popularity + \beta_4 * popularity^2 + \\
&+ \beta_5 * quality * popularity + \varepsilon \\
&= V + \varepsilon
\end{aligned} \tag{1}
$$

where *quality* and *popularity* are the levels of the same attributes, *quality*popularity* is the interaction term, *V* is considered to be the systematic component of utility and is a function of the levels of the two attributes and of their interaction term. The βs are the part-worth utilities associated to the various function terms and ε represents the unobserved error term. Table 3 presents the estimates for the part-worth utilities, all statistically significant.

[Table 3 here]

Analysis of the observed and the estimated ranks was also used to identify inconsistent respondents to be excluded from the subsequent analysis. More precisely, we computed Kendall's tau rank correlation coefficient, a statistic commonly used to measure the similarity of the orderings of two measured quantities, between the observed and estimated rank of each of the scenarios respondents have been exposed to. By computing a Kendall's tau coefficient for each participant, we could identify

---

[1] Detailed information on the model selection exercise is available on request

7 respondents with a low score, an evident sign of poor data quality. We decided to exclude them from the original sample of 168. Reasons for a low Kendall's tau coefficient could be lack of understanding of the exercise, low motivation or lack of engagement. Fatigue, which is a typical cause of poor data quality in conjoint experiments, can be safely excluded from the likely sources of inconsistency for this study as the experiment required at most 7 minutes.

4. RESULTS

After obtaining the individual-level part-worths for the two attributes and their interaction term, we retrieved the total preference in terms of rank for each possible scenario for each respondent and then we averaged the results over the whole sample.

4.1 Quality versus popularity

We used conjoint analysis to estimate the relative importance of the studied attributes. In particular, relative importance of an attribute is a function of the variation in preference associated with a variation in the attribute's levels.   In other words, the bigger the difference between the highest and the lowest utility levels of an attribute, the more significant is its partial contribution to preference. This can help us to test H1, showing the relative importance of popularity.

 [Figure 2 here]

Figure 2 shows the average impact for the two stimuli: quality and popularity. It is apparent that quality is the most important stimulus, with an average impact score of 67.4%. However, popularity counts quite a lot, 32.6%.

Figure 3 shows the preferences of the stimulus levels for quality and popularity, letting the reader to see the effect of the parameters presented in Table 3. The direction of the arrows represents the transition of the underlying dimension from low to high. It is worth recalling here that popularity is measured in geometric progression (i.e., from 1 to 4096 reviews), then changes of one level in

14

popularity measure the impact on preference of two consecutive values of that scale. Quality is confirmed to be the main driver for preference. As one would expect, low quality is associated with low preference and high quality with high preference. This association is clearly non-linear: impact of quality on preference is high for low quality levels and it decreases for higher quality levels, showing decreasing returns. There is also a clear interaction with popularity. In fact, impact of popularity on preference is higher when quality is high, while the impact is considerably lower, but still positive, when quality is low. If we take the point where quality is at the minimum level, 1, and popularity at its maximum, 7, we can see that this condition is significantly preferred to the one where both quality and popularity are at their minimum, 1 ($t(160) = 17.18$, $p < .001$).

[Figure 3 here]

These results support and go beyond H2c: not only the effect of popularity is not symmetric, being different in absolute value when quality is high or low, but it is always positive, giving credit to the theory that people tend to go with the crowd. Therefore, H2a, the negative role of popularity when quality is low is not supported by the analysis. The role of normative social influence, H2b, is indeed more relevant.

4.2 The impact of heterogeneity of respondents' preference on the overall ranking of hotels

In order to investigate to what extent quality and popularity impacted differently depending on the type of respondent, participants were categorized in relation to their demographics. Results for the demographic segments are presented in Table 3.

[Table 4 here]

Young male respondents tend to consider mostly quality when making a ranking decision. Their preferences differ significantly from older males ($t(84) = 7.63$, $p < .001$), young females ($t(74)=9.46$, p

<.001) and older females ($t(87)=11.24$, p <.001). Young females perform similarly to older males, while older females differ significantly in their preferences from younger females ($t(73)=3.19$, p <.005), taking more into account the number of reviews when booking a hotel.

To strengthen the heterogeneity analysis, we identified homogeneous groups through segmentation analysis. Based on the individual-level part-worths for the two attributes and their interaction term, we run a segmentation analysis to identify groups of respondents with similar preference functions. In particular, we used latent class analysis (LCA) which is a statistical method for finding groups or subtypes of related cases (latent classes) in multivariate data (*for review, see* Bartholomew *et al.*, 2011; Hagenaars and McCutcheon, 2005). Input variables for the LCA were the part-worth utilities which characterize the preference in choosing a hotel based on the number of reviews and average of consumers' reputation. We chose a solution based on the average weight of evidence (AWE) because this metric combines information on model fit and information on classification errors (Banfield and Raftery, 1993; Celeux et. al., 1997). The chosen solution, the one with the lowest AWE, is characterized by three groups. Table 4 shows the size of the groups and their profiling in terms of importance scores and demographics. Figure 4 shows the preferences of the stimulus levels for quality and popularity for each of the three groups.

[Table 5 here]

[Figure 4 here]

Group 1 is predominant and is characterized by people who take into account both quality and popularity when making a choice, with quality however being around twice as important as popularity. Group 2 considers quality as the main driver of choice while popularity clearly plays only a little role. For group 3, popularity plays a major role, being almost as important as quality in the decision process.

There is a clear interaction effect: popularity gains importance for high levels of quality in group 1 and group 3, while in group 2 the effect of popularity remains low across all the levels of quality. We can clearly see from Table 4 that the second group is mainly composed by young males while older females are predominant in group three. The first group, which is the biggest one, is more heterogeneous and primarily composed by young females and older males. These results confirm that the demographical profile is a clear driver of choice (*H3*).

4.3 Time and decision making

The third hypothesis suggested that demographics might play a role in the strategy taken to arrive at a decision and this could affect the extent to which people are affected by popularity. In order to test this, we first measured if some groups were focusing only on one attribute in order to arrive at a decision. To do so, we artificially created a ranking based on lexicographic preferences. Under this ordering structure, consumers strictly prefer a higher quality level, considering popularity to have an effect *only* when quality is equal among hotels. This ranking was then correlated with the actual ranking stated by participants. The most used measure of correlation for rank data, Kendall's tau coefficient, was higher in the young male group (M = .89, SD = .0733698) than in the other groups (M = .79, SD = .12). This difference was significant by means of t test, t (159) = 5.22, $p < .001$.

A two-way ANOVA examined the amount of time spent to complete the task across our demographic segments. Invested time differed significantly across gender ($p < 0.001$) and age ($p < 0.001$)*,* while there was no significant interaction between gender and age. Tukey post-hoc comparisons indicate that the young male group (*M* = 3.4, 95% CI [3.10 3.69]) invested less time than all the other groups. On the opposite, older females invested more time than all the other groups. Figure 5 presents graphically the whole set of results.

[Figure 5 here]

4.4 Open-ended Remarks and Considerations

The open-ended remarks were further examined by the investigators in order to understand the basis of the respondents' decisions. Here we present some of the open-ended remarks of the respondents by topic.

A great majority is arguing that a high stated quality without enough reviews may be a signal of unreliability. Interestingly, some respondents identified a threshold of popularity to be trustable.

Respondent # 114: "4 reviews are not enough to be trustable. A larger number of reviews reduce the uncertainty"

Respondent #130: "with more than 100 reviews I can trust it"

A considerable number of participants elaborated that price is another key variable, even more important among the youngest group, constrained by limited budget. Further, the content of the review itself is relevant for some participants, even more than the number of reviews.

Respondent #55: "I also consider the content of the review"

Respondent #112: "I would like to see the detailed review"

Respondent #119: "I also take into consideration the price, and the location of the hotel"

These comments conform to the theory that reviews are indeed informative (Li and Hitt, 2008) and the content of the reviews needs to be properly investigated.

Concordant to Hypothesis 1, participants' comments confirmed that hotel consideration significantly increases when the number of reviews is high. The finding, supported by the quantitative analysis, of a positive effect of popularity even when the on-line reputation is negative was not present in respondents' comments. Based on some respondent comments the opposite would be true.

Respondent #129: "Many negative reviews are a signal that the hotel is really bad"

Respondent #140: "If the hotel has many reviews and low quality average it means it is really bad"

These comments provided by few respondents and to be discussed in the next section, would favor H2a, a hypothesis that was not supported by our quantitative analysis.

5. DISCUSSION OF RESULTS AND LIMITATIONS

The findings of this study show that, whereas average reputation is the most influential stimulus when assessing a hotel, popularity, measured as the number of reviews, also has a significant impact on preferences. In particular, here we highlight the exact relevance of the number of reviews for different levels of on-line reputation, showing that popularity has, on average, a positive effect which becomes larger for high levels of on-line reputation. This can help to explain that, although the informative effect of something being popular may play a role, the normative social effect, and then popularity per se, is relevant. Interestingly, while few participants argued that negative popularity, i.e. many bad reviews, has a negative effect on their preferences, there is a large silent majority that does not comment on the reasons of their behavior. These show a tendency to consider even negative popularity as a signal of quality. This behavior goes against the value of information: more reviews should be more informative, therefore more bad reviews should strengthen the idea that the hotel is really bad. Also, based on real world evidence, review websites, such as Trip Advisor, are ranking according to informational social influence. In this analysis we found that this is *not* the way people process information.

Heterogeneity across respondents plays a role. Demographics are indeed a moderator of the impact of popularity. While young males tend to save time considering just the most relevant attribute, quality,

confirming previous literature on impulsiveness, all the other demographic groups and in particular older females are weighting popularity much more when they have to state their preference.

Our study is not without limitations. First, even though we tried to create a sample which represents the population of on-line users of travel websites, the composition of our sample is not a perfect replication of on-line users of travel websites. Second, participants were responding to a scenario and not actually booking and spending their own money. However, the use of a simulated consumer experience does not necessarily weaken internal and external validity as previous studies have reported (Louviere and Woodworth, 1983; Lynch, J. G., 1982). In addition, by using controlled conditions we show how changes in rankings are actually caused by variations in popularity. Third, additional consumer information such as the frequency of booking on-line and variables such as hotel location and price could have been added to the study. This would have added more detail to the general picture, but would have gone beyond the scope of this research. We chose instead to replicate only one step of the consumer decision process. Finally, the scenarios do not show the distribution of the hotels ratings. Two similar medium quality hotels in terms of on-line reputation can be the result of midpoint ratings or the result of balancing ratings and this could potentially have an impact on consumer choice (Eisend, 2006; Purnawirawan *et al.*, 2012).

6. CONCLUSIONS AND IMPLICATIONS

The wealth of available information on the Internet has increased the need to focus attention across the abundance of information sources. This research contributes to knowledge development in consumption and in the hospitality industry by examining the impact of popularity (measured as the number of reviews), and quality (measured as on-line reputation) provided by former consumers. We were particularly interested in testing if consumers tend to prefer popular alternatives even if the crowd they are following defines those alternatives as poor and if these effects vary across certain

demographics. Our results reveal that the presence of many reviews, despite the on-line evaluation of the review itself, creates a positive shift on preference and that something being popular affects preferences more among females and older people. A vast majority choose based on the normative social influence, while some people go with the informational social influence. We find a silent majority that prefers alternatives rated as bad by many other consumers than alternatives which are also rated bad but by fewer consumers. This behavior goes against the value of information because the judgment of multiple others should provide more certainty about the real quality of a service (Huang and Chen, 2006) , as a "talkative minority" argues explaining how it came out with a preference. Other studies on silent majorities in consumption often deal with no complainers who experience service failure (Chebat *et al.*, 2005; Sheth *et al.*, 2000; Voorhees *et al.*, 2006).

In order to take into account the preferences of the vast majority, effective ranking mechanism has to be applied. If you manage a ranking website showing different hotels, on the one hand our results seem to indicate that quality, relative to popularity, is the most efficient stimulus across all the segments. On the other hand, ranking mechanisms have to reflect the popularity effect adequately. Popularity has a great impact on female and older consumer segments, although still relatively less important than quality. So, if you are creating rankings dealing mostly with consumers groups that consider popularity as a key variable, focusing on both quality and popularity would be necessary. Some psychological extensions of this study can be assessed in future studies. In particular, while males and females were already shown to differ in the way they are influenced by others and how this effect is moderated by time constraints (Mitchell and Walsh, 2004), it would be important to better understand the boundary conditions of such behavior and if other socio-demographic variables, such as income and education level, might also play a role.

Addressing a gap in research recently highlighted by Serra, Cantallops and Salvi (2014), this study has useful implications for hospitality and review website managers. Considering that hotels are facing intense competition due to a stagnant economy, managers need to improve their marketing strategies to enhance the popularity of their hotels. As in Bennet and Rundle (2004), we find that the rating of on-line reputation is important but it is not enough. We clearly show that the review being good or bad is not the only relevant element; what is also important is having a large volume of reviews. One possible explanation is that popularity is more likely to attract the interest of on-line consumers (Zhang *et al.*, 2010), giving credit to the theory that volume counts more than valence (Liu, 2006). Another explanation is that people may infer a high number of reviews as a signal of something valuable to try and not as a sum of negative experiences. Perhaps assuming that there are missing attributes not detected in the rating. Future research should help to disentangle these alternative explanations and extend the study to examine further steps in the consumer decision process. For instance, it would be interesting to assess that part of the process where consumers move from the consideration set to the choice set, evaluating hotel attributes in more detail.

REFERENCES

Abrate, G., Capriello, A., Fraquelli, G., 2011. When quality signals talk: Evidence from the Turin hotel industry. Tourism Management 32 (4), 912-921.

Addelman, S., 1962a. Orthogonal main-effect plans for asymmetrical factorial experiments. Technometrics 4 (1), 21-46.

Addelman, S., 1962b. Symmetrical and asymmetrical fractional factorial plans. Technometrics 4(1), 47-58.

Alba, J., Lynch, J., Weitz, B., Janiszewski, O., Lutz, R., Sawyer, A., Wood, S., 1997. Interactive home shopping: consumer, retailer and manufacturer incentives to participate in electronic marketplaces. Journal of Marketing 61, 38-53.

Banfield, J. D., Raftery, A. E., 1993. Model-based Gaussian and non-Gaussian clustering. Biometrics 49, 803-821.

Barber, N., Almanza, B. A., Donovan, J. R., 2006. Motivational factors of gender, income and age on selecting a bottle of wine. International Journal of Wine Marketing 18 (3), 218-232.

Bartholomew, D. J., Knott, M. Moustaki, I., 2011. Latent variable models and factor analysis: a unified approach, 3rd ed. John Wiley & Sons, London, UK.

Bennett, R., Rundle-Thiele, S., 2004. Consumer satisfaction should not be the only goal. The Journal of Services Marketing 18, 514-523.

Berger, J., Sorensen, A. T., Rasmussen, S. J., 2010. Positive effects of negative publicity: when negative reviews increase sales. Marketing Science 29 (5), 815-827.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., 2006. Consumer search and retailer strategies in the presence of online music sharing. Journal of Management Information System, 23 (1), 449-467.

Bickart, B., Schindler, R., 2001. Internet forums as influential sources of consumer information. Journal of Interactive Marketing 15 (3), 31-40.

Bozdogan H., 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika 52, 345-370.

Breese, J. S., Heckerman, D., Kadle, C., 1998. Empirical analysis of predictive algorithms for collaborative filtering. In S. M. Gregory and F. Cooper (eds.). Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, School of Business, University of Wisconsin,

Madison, WI, 43-52.

Celeux, G., Biernacki, C., Govaert, G., 1997. Choosing models in model-based clustering and discriminant analysis. Technical Report. Rhone-Alpes: INRIA.

Chatterjee, P., 2001. Online reviews: do consumers use them? Advances in Consumer Research 28, 129–133.

Chebat, J. C., Davidow, M. Codjovi, I., 2005. Why some dissatisfied consumers fail to complain. Journal of Service Research 7 (4), 328-342.

Chen, Y. C., 2009. Essays on mobile advertising and commerce. Doctoral dissertation. Science Technology and Management, Harvard University, Boston, MA.

Chevalier, J., Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. Journal of Marketing Research 43 (3), 345-354.

Cross, C. P., Copping, L. T., Campbell, A., 2011. Sex differences in impulsivity: A meta- analysis. Psychological Bulletin 137 (1), 97-130.

Dean A., Voss D., 1999. Design and Analysis of Experiments. New York: Springer-Verlag.

Dellarocas, C., 2003. The digitalization of word of mouth: promise and challenges of online feedbacks mechanisms. Management Science 49 (10), 1407-1424.

Demia Studio Associato, 2012. R-sw Conjoint: a Package for Advanced Conjoint Analysis, available from http://www.demia.it.

Denrell, J., Le Mens, G., 2012. Following the crowd can lead people who learn only from their own experience to agree with a mistaken majority. Working Paper. Retrieved from http://idei.fr/doc/conf/iast/paper_lemens.pdf

Deutsch, M., Gerard, H. B., 1955. A study of normative and informational social influences upon individual judgment. Journal of Abnormal Social Psychology 51, 629-36.

Eisend, M., 2006. Two-sided Advertising: A Meta-analysis. International Journal of Research in Marketing 23, 187-198.

Fagerstrom A., Ghinea, G., 2011. On the motivating impact of price and online recommendations at the point of online purchase. International Journal of Information Management 31, 103-110.

Festinger, L., 1954. A theory of social comparison processes. Human Relations 7, 117-140.

Filieri, R., Mcleay, F., 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.

Forman, C., Batia Wiesenfeld, A. G., 2008. The relationship between reviews and sales: the role of

reviewer identity disclosure in electronic markets. Information Systems Research 19 (3), 291-313.

Godes, D., Mayzlin, D., 2009. Firm-created word-of-mouth communication: evidence from a field test. Marketing Science 28 (4), 721–739.

Green P. E., Rao V. R., 1971. Conjoint measurement for quantifying judgemental data. Journal of Marketing Research 8, 355-63.

Gustaffson A., Herrmann A., Huber F., 2001. Conjoint analysis as an instrument of market research practice. In: Conjoint Measurement: Methods and Applications, Gustaffson A., Herrmann A., & Huber F. (Eds.), Berlin: Springer, 5-46.

Hagenaars, J.A & McCutcheon, A., 2002. Applied latent class analysis. Cambridge: Cambridge University Press.

Han, H., Hsu, J., Lee, J., 2009. Empirical investigation of the roles of attitudes toward green behaviors, overall image, gender, and age in hotel customers' eco-friendly decision-making process. International Journal of Hospitality Management 28 (4), 519-528.

Herr, P.M., Kardes, F.R., Kim, J., 1991. Effects of word-of-mouth and product-attribute information on persuasion: an accessibility-diagnosticity perspective. Journal of Consumer Research 17 (4), 454-62.

Holbrook, M. B., Moore, W. L. 1981. Feature interactions in consumer judgments of verbal versus pictorial presentations. Journal of Consumer Research 8, 103-113.

Hu, N., Pavlou, P. A., Zhang, J., 2006. Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication. Proceedings 7th ACM Conference on Electronic Commerce, Ann Arbor, Michigan, USA, June 11-15, 324-330.

Huang, J. H., Chen, Y. F. 2006. Herding in online product choice. Psychology and Marketing 23, 413-428.

Huang, P., Lurie, N. H., Mitra, S., 2009. Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods. Journal of Marketing 73 (2), 55-69.

Hung, K., Law, R., 2011. An overview of Internet-based surveys in hospitality and tourism journals. Tourism Management 32 (4), August 2011, 717-724.

Jeacle, I., Carter, C. 2011. In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. Accounting, Organizations and Society 36, 293-309.

Jiménez, F. R., Mendoza, N. A., 2013. Too Popular to Ignore: The Influence of Online Reviews on Purchase Intentions of Search and Experience Products. Journal of Interactive Marketing,

Available online 10 July 2013.

Jones, P., Chen, M. M., 2011. Factors determining hotel selection: Online behaviour by leisure travellers. Tourism and Hospitality Research 11(1), 83–95.

Johnson, E. J., Payne, J. W., 1985. Effort and accuracy in choice. Management Science 31 (4), 395-414.

Khare, A., Labrecque, L. I., Asare, A. K., 2011. The assimilative and contrastive effects of word-pf-mouth volume: an experimental examination of online consumer ratings. Journal of Retailing 87 (1), 111-126.

Klein, L.R., 1998. Evaluating the potential of interactive media through a new lens: search versus experience goods. Journal of Business Research 41 (3), 195-203.

Koh, N. S., Hu. N., Clemons, E. K., 2010. Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. Electronic Commerce Research and Applications 9 (5), 374-385.

Kulkarni, G., Ratchford, B. T., Kannan P.K., 2012. The Impact of Online and Offline Information Sources on Automobile Choice Behavior. Journal of Interactive Marketing 26 (3), 167–175.

Li, X., Hitt L. M., 2008. Self-Selection and Information Role of Online Product Reviews. Information Systems Research 19 (4), 456-474.

Litvin, S., Goldsmith, R., Pan., B., 2008. Electronic word-of-mouth in hospitality and tourism management. Tourism Management 29, 458-468.

Liu, Y., 2006. Word of mouth for movies: its dynamics and impact on box office revenue. Journal of Marketing 70, 74-89.

Louviere, J. J., Woodworth, G., 1983. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. Journal of Marketing Research 20 (4), 350-367

Lynch, J. G., 1982. On the external validity of experiments in consumer research. Journal of Consumer Research 9(3), 225-239.

Lynch, J. G., Ariely, D., 2000. Wine online: Search costs affect competition on price, quality, and distribution. Marketing Science 19, 83-103.

Mauri, A., Minazzi, R., 2013. Web reviews influence on expectations and purchasing intentions of hotel potential customers. International Journal of Hospitality Management 34, 99-107.

Mazzarol, T., Sweeney, J., Soutar, G., 2007. Conceptualizing word-of-mouth activity, triggers and conditions: an exploratory study. European Journal of Marketing 41 (11), 1475-1494.

Meyers-Levy, J., 1988. Influence of sex roles on judgment. Journal of Consumer Research 14, 522-

530.

Mitchell, V., Walsh, G., 2004. Gender differences in German consumer decision-making styles. Journal of Consumer Behaviour 3 (4), 331-346.

Mudambi, S. M., Schuff, D., 2010. What makes a helpful online review? A study of consumer reviews on amazon.com. MIS Quarterly 34 (1), 185-200.

Nelson, P., 1970. Information and consumer behavior. Journal of Political Economy 81 (4), 78 (20), 311-329.

Noort, G, Willemsen, L. M., 2012. Online Damage Control: The Effects of Proactive Versus Reactive Webcare Interventions in Consumer-generated and Brand-generated Platforms. Journal of Interactive Marketing 6 (3), 131–140.

Okada, E. M., 2005. Justification effects on consumer choice of hedonic and utilitarian goods. Journal of Marketing Research 41 (1), 43-53.

Pagani, M., 2004. Determinants of adoption of third generation mobile multimedia services. Journal of Interactive Marketing, 18 (3), 46-59.

Payne, J., Bettman, J. R., 1996. When time is money: decision behavior under opportunity-cost time pressure. Organizational Behavior and Human Decision Processes 66 (2), 131-152.

Park, D. H., Lee, J., Han, J., 2007. The effect of online consumer reviews on consumer purchasing intention: the moderating role of involvement. International Journal of Electronic Commerce 11 (4), 125-148.

Petty, R.E., Cacioppo, J.T., 1984. The effects of involvement on response to argument quantity and quality: central and peripheral routes to persuasion. Journal of Personality and Social Psychology 46 (1), 69-81.

Poston, R., Speier, C., 2005. Effective use of knowledge management systems: a process model of content ratings and credibility indicators. MIS Quarterly 29 (2), 221-244.

Purnawirawan, N., De Pelsmacker, P., Dens, N. 2012. Balance and Sequence in Online Reviews: How Perceived Usefulness Affects Attitudes and Intentions. Journal of Interactive Marketing 26 (4), 244–255.

Rossi P. E., Allenby G. M., McCulloch R., 2005. Bayesian Statistics and Marketing. West Sussex, England: John Wiley & Sons, Ltd.

Sen, S., Lerman, D., 2007. Why are you telling me this? An examination into negative consumer reviews on the Web. Journal of Interactive Marketing 21 (4), 76-94.

Serra Cantallops, A., Salvi, F., 2014. New consumer behavior: A review of research on eWOM and

hotels. International Journal of Hospitality Management 36, 41-51.

Sheth, J. N., Sisodia, R. S., Sharma, A., 2000. The Antecedents and consequences of customer-centric marketing. Journal of the Academy of Marketing Science 28 (1), 55-66.

Sirakaya, E., Woodside, A. G., 2005. Building and testing theories of decision making by travellers. Tourism Management 26 (6), 815–832.

Sparks, B. A., Perkins, H. E. and Buckley, R., 2013. Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior, Tourism Management 39, 1-9.

Spoerri, A., 2008. Authority and ranking effects in data fusion. Journal of the American Society for Information Science and Technology 59 (3), 450-460.

Stigler, G. J., 1961. The Economics of Information. Journal of Political Economy 69 (3), 213-225.

Tse, A. C., 2003. Disintermediation of travel agents in the hotel industry. International Journal of Hospitality Management 22, 453-460.

Vermeulen, I. E., Seegers, D., 2009. Tried and tested: The impact of online hotel reviews on consumer consideration. Tourism Management 30, 123-127.

Voorhees, C. M., Brady, M. K., Horowitz, D. M., 2006. A voice from the silent masses: an exploratory and comparative analysis of Noncomplainers. Journal of the Academy of Marketing Science 34 (4), 514-527.

Weathers, D., Sharma, S., Wood, S. L., 2007. Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. Journal of Retailing 83 (4), 393-401.

Wilska, T. A., 2003. Mobile Phone Use as Part of Young People's Consumption Styles. Journal of Consumer Policy 26 (4), 441-463.

Wood, L., 2004. Dimensions of brand purchasing behavior: consumers in the 18–24 age group. Journal of Consumer Behavior 4 (1), 9-24.

Yacouel, N., Fleischer, A., 2012. The role of cybermediaries in reputation building and price premiums in the online hotel market. Journal of Travel Research, 51(2), 219–226.

Ye, Q., Law, R., Gu, B., 2009. The impact of online user reviews on hotel room sales. International Journal of Hospitality Management 28 (1), 180-182.

Zhang, Z. Qiang, Y., Law, R., Li, Y., 2010. The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. International Journal of Hospitality Management 29, 694-700.

TABLES AND FIGURES

Table 1 - Demographics of the respondents included in the analysis

Age \ Gender	Male	Female	Total
<25 yo	45	31	76
25+ yo	41	44	85
Total	86	75	161

Table 2 – Attributes and levels included in the conjoint exercise

Attribute	Level 1 *Low*	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7 *High*
Rating	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Number of reviews	1	4	16	64	256	1024	4096

Table 3 – Parameter Estimates of the HB model

Coeff.	Attribute	Effect	Value	Standard Error	p-value
$\hat{\beta}_0$	intercept	-	0.715	0.032	<0.001
$\hat{\beta}_1$	quality	Linear	1.997	0.046	<0.001
$\hat{\beta}_2$	quality	Quadratic	-0.334	0.027	<0.001
$\hat{\beta}_3$	popularity	Linear	0.262	0.031	<0.001
$\hat{\beta}_4$	popularity	Quadratic	-0.068	0.031	0.028
$\hat{\beta}_5$	quality*popularity	Linear	1.713	0.063	<0.001

Table 4 – Relative importance scores

Gender	Total	Male		Female	
Age		<25 yo	25+ yo	<25 yo	25+ yo
Sample size	161	45	41	31	44
Quality	67.4%	77.7%	64.3%	66.2%	60.5%
Popularity	32.6%	22.3%	35.7%	33.8%	39.5%
St. dev.	10.4%	5.1%	10.5%	5.4%	8.9%

Table 5 –LCA groups with profiling

Group	*n*	Imp. Quality	Imp. Popul.	Male <25 yo	Male 25+ yo	Female <25 yo	Female 25+ yo
1	61	66.0%	34.0%	4.4%	53.7%	93.5%	18.2%
2	51	77.3%	22.7%	95.6%	14.6%	6.5%	0.0%
3	49	58.8%	41.2%	0.0%	31.7%	0.0%	81.8%

Figure 1 – Screenshot of the cards presented to a respondent
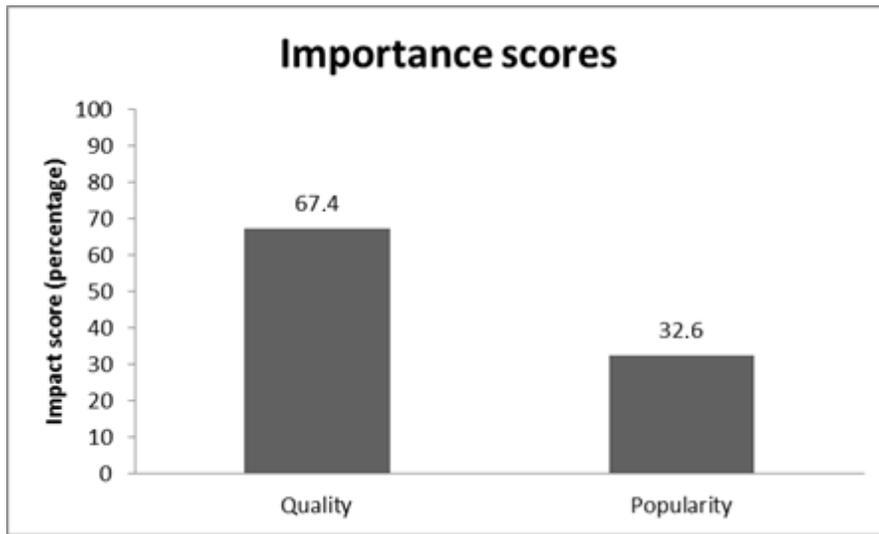
Figure 2 – Relative importance scores

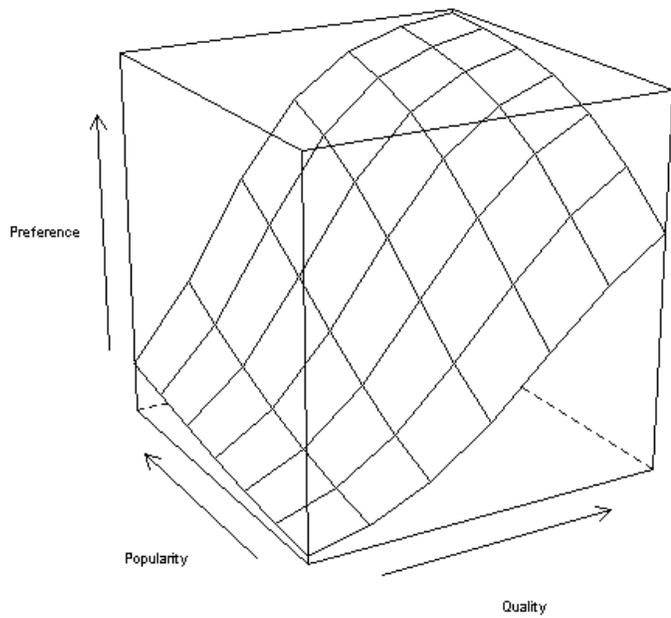**Figure 3 – Rank preference for any combination *quality – popularity* (*n* = 161)**

Figure 4 – Rank preference for any combination *quality – popularity* for each group

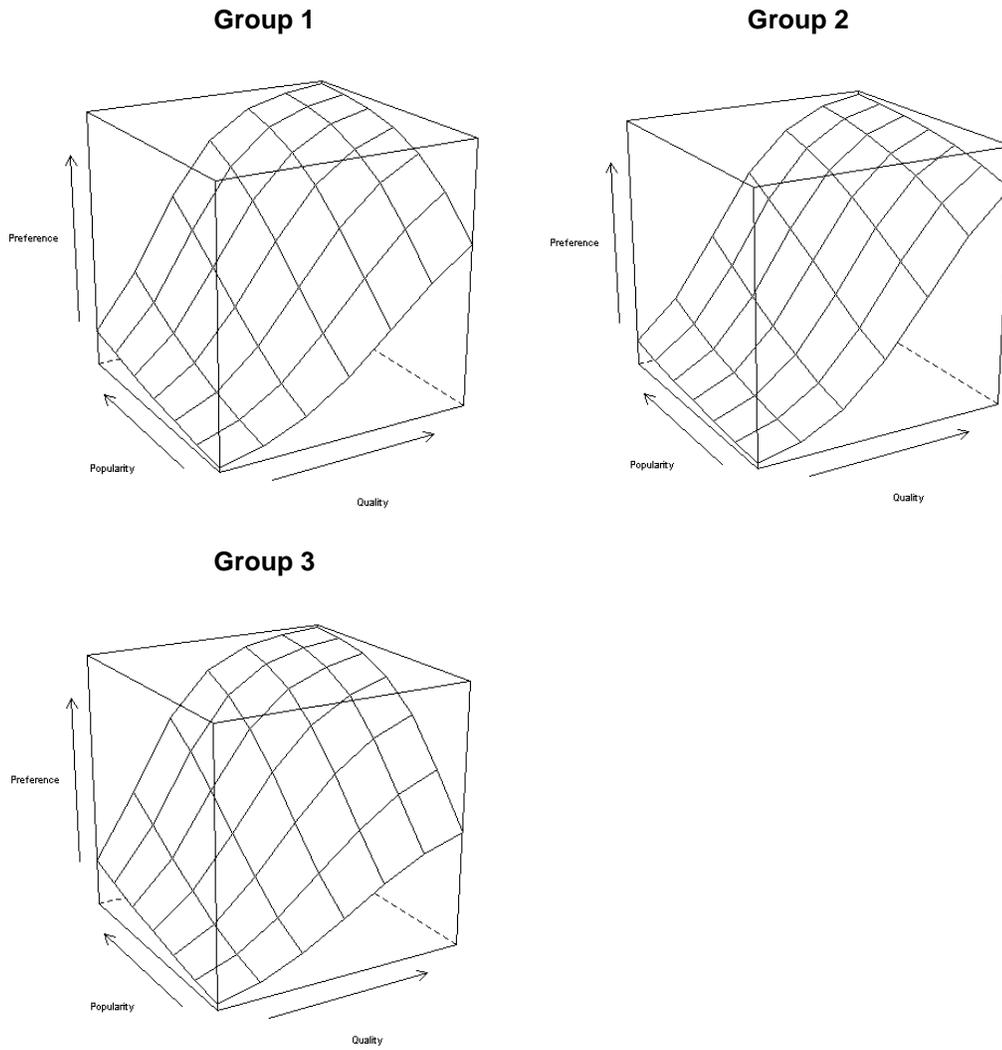**Group 1**   **Group 2**

Group 3

Figure 5 - Invested time across demographic groups