The Bright Side of Bad Choices: Evidence from Restaurant Exploration

Victor Martínez-de-Albéniz[1] • Manuel Sosa[2] • Clara Carrera[3]

Abstract

The value derived from hedonic goods is affected by reference effects at the time of consumption, usually in the form of quality standards. Consumption typically involves two steps: first, the consumer chooses a given good, among a pool of available choices; then, the consumer experiences the good and derives a satisfaction from it. We investigate the role of references in these two phases, taking into account potential selection bias in the satisfaction outcome. Using novel longitudinal data from restaurant reviews, we find evidence of quality loss aversion in the choice decision, in accordance with prospect theory. That is, people tend to disproportionally avoid choosing restaurants that are below their quality reference. At the same time, choosing a lower-quality good does not penalize satisfaction in the chosen restaurant as much as one would expect, which is consistent with the fact that consumers adjust their expectations downward when they choose a lower-quality option. This suggests that expectation adjustment protects consumers when they make apparent "bad" choices. Given these results, we show that it may be optimal to visit restaurants in a zigzag that alternates between high- and low-quality choices.

Keywords: Online user reviews, prospect theory, recommendation systems, Heckman selection correction

1 Introduction

When predicting what a person might like when choosing a hedonic good, e.g., a book, a movie or a restaurant (Hirschman and Holbrook 1982, Holbrook and Hirschman 1982), people often resort to their own past experiences as well as to the suggestions given by other people similar to them (Festinger 1954). Not surprisingly, with the advent of e-commerce and online recommender systems, people also consider the recommendations of online platforms when making purchasing decisions (Avery et al. 1999, Dellarocas 2003, 2006, Adomavicius and Tuzhilin 2005).

A fundamental assumption behind any recommendation (either given by members of an individual's social network or by an online recommender system) is that choice occurs after evaluating all the options available, and hence the selected option was the one that offered, on expectation, the highest net utility to that individual. In other words, we take for granted that there is consistency between the expected utility associated with a given choice, and the realized utility after the choice has been made, in the sense that maximizing expected utility also leads to maximal realized utility. This basic assumption has led to the proliferation of online recommender systems that only consider past choices of the individual or consumers with similar profile, but not the utility that these experienced. For example, collaborative filtering generates recommendations that offer items

[1]valbeniz@iese.edu. IESE Business School, University of Navarra, Av. Pearson 21, 08034 Barcelona, Spain.
[2]manuel.sosa@insead.edu. INSEAD, Technology and Operations Management, INSEAD, Singapore 138676.
[3]clara.carrera@insead.edu, INSEAD, Technology and Operations Management, Fontainebleau, France 77300.

similar to the ones selected in the past by the focal individual or consumers that share similar past purchasing behavior; content-based filtering also incorporates discrete features of products previously selected by the focal individual or similar others; see Ricci et al. (2011) for more details.

While this premise seems reasonable, hedonic goods by definition include an experience component that is typically subject to a number of behavioral effects. First, people build references that influence how they value the hedonic good, and specifically derive utility relative to such a reference (Tversky and Kahneman 1991). Such a reference effect is crucial in prospect theory, which predicts that people tend to react differently to expected gains and expected losses relative to the reference; people tend to disproportionally avoid expected losses, in comparison with similar expected gains (Tversky and Kahneman 1992). Second, the expectation formation process also plays an important role in determining the satisfaction level provided by the consumption of the hedonic good: experienced happiness is the difference between perceived reality and expectation (Anderson and Sullivan 1993, Baucells and Sarin 2012). Specifically, satisfaction of the consumer of a hedonic good is affected positively by the perception of the quality of such a good after consumption, as well as negatively by the expectation that was formed before starting experiencing the good itself, especially when it has a service or experience component (Maister 1984). The process that governs choice and outcome from hedonic good consumption is indeed complex precisely because choice and outcome are tightly coupled(Oliver 1980). It is hence no longer clear that forming a recommendation based on what the crowd chooses, even when the crowd is very similar to oneself, is a good idea.

The literature in psychology and marketing has studied the effect of references in repeated service interactions (Rust and Chung 2006), and suggested policies to manage retention at a given firm (Rust et al. 1999, Aflaki and Popescu 2014, DeCroix et al. 2019). In these works, customers make a choice to return, or leave the service provider permanently. For hedonic goods where the number of choices is large, however, the literature is silent about the consumer's exploration process through the available options. In this context, reference effects are built from past experiences, while expectations are formed from the available information about each option. We propose in this paper a framework to describe how consumers choose among many possible hedonic goods, and how satisfied they are with their choices. We are mostly interested in clearly separating choices, from outcomes associated with such choices. This allows us to better understand how expectations, which are statistically related to the choice that has actually been made, influence the outcome, i.e., the satisfaction with the use or consumption of the hedonic good chosen. Once this link is established, we are able to simulate counterfactual alternative choices and make a recommendation that optimizes consumer satisfaction levels. Of course, this entails significant empirical difficulties, and requires obtaining a longitudinal data set with individuals making several choices (more than one at a minimum) over time and reporting their associated satisfaction.

We thus formulate a two-stage selection-outcome model, that incorporates the main behavioral effects related to the dynamic states the consumer goes through, proposed by existing theories. Specifically, our model considers reference effects formed from previous experiences, as well as information collected by other consumers that provide a quality signal about potential choices. We consider the two relevant elements of exploration, namely choice (for which expectation is relevant) and outcome (for which both perception of the actual quality of the good during consumption and expectation about the expected quality of the good formed during the choice process matter).

We test our model with a novel longitudinal dataset about restaurant choices and dining experience evaluations, made by a large collection of consumers over several years, obtained from different European cities and review platforms. The dataset is constructed from consumers' restaurant reviews posted online, and for which we can create consumer "trajectories", i.e., the sequence of visits to restaurants that each consumer has posted. We estimate a two-stage selection-outcome Heckman correction model, where the probability that a consumer selects a given restaurant depends on consumer and restaurant characteristics as well as consumer-specific reference effects (Heckman 1979). Then, the probability that the consumer has an excellent experience in the selected restaurant is also a function of the same reference effects as well as proper selection-correction and consumer- and restaurant-specific controls.

We find that indeed reference effects are present in both selection and outcome. Interestingly, quality reference (i.e., the consumer-specific signal that captures whether a restaurant is good or not) plays a different role in selection and outcome stages of our model. Consistent with prospect theory, people tend to disproportionally avoid selecting restaurants that are below their quality reference. However, while controlling for such a "quality loss aversion", when a consumer visits a restaurant that is below her quality reference, the lower than expected quality of the selected restaurant does not translate in lower than expected probability of having an excellent restaurant experience; on the contrary, the probability of having an excellent dining experience when choosing a lower-than-the-reference quality restaurant is comparatively higher than when the choice has an expected quality that exceeds the quality reference, controlling for the quality level of the selected restaurant. This suggests that making a "bad" choice, i.e., going below the individual's quality reference, does not hurt as much as one would anticipate: the consumer indeed experiences a satisfaction bonus, or equivalently disproportionally adjusts expectations down so that actual satisfaction is boosted up. We call this the value of low expectations, or the bright side of (a priori) "bad" choices.

Our paper makes three main contributions to the literature. First, our study extends the study of reference effects to a context of multiple firms, without repetition, in contrast with most of existing works. This requires defining the latent construct of expectation for options that have never been tried before by the consumer, for which we use the information provided by other users that have provided feedback on that choice. Moreover, while the extant literature provides evidence through laboratory experiments (Tse and Wilton 1988) or surveys (Parasuraman et al.

1988, Anderson and Sullivan 1993), we use retrospective data from reviews, which has the advantage of providing longer time windows allowing and multiple sequential choices by the same individual.

Second, even though selection bias has been well studied in many economic settings (e.g., labor markets, see the seminal paper of Heckman 1979), we are not aware of applying the concept to the analysis of consumption satisfaction, and in particular to that of hedonic goods. The existence of such a selection bias, which we document in this paper, has strong implications for recommendation systems. Indeed, such systems can work towards two different, yet contrasting, goals. The first possibility is to recommend items that have a high choice probability. Such option is preferred when the recommender is rewarded from the choices made, e.g., in online advertising such as Google Ads, where the advertising platform is paid per click to the ad, or in online marketplaces such as Tmall or Amazon, where the marketplace collects a commission when an item is purchased. This seems to be the norm in most recommendation systems that use collaborative or content-based filtering (Ricci et al. 2011). Unfortunately, these recommendations do not necessarily suggest the most satisfactory outcomes for the user, because (i) they do not take into account the impact of expectation formation by the choices made, which are consumer-dependent; and (ii) they do not consider the potential influence of such recommendation on the formation of new reference effects, which could become relevant in future consumption occasions, beyond the current one. The second possibility for recommendation systems is to recommend items that maximize (ex-post) consumer satisfaction, which, as we find in our setting, does not necessarily coincide with maximizing (ex-ante) choice probability. In this case, recommendation should take into account consumer quality evaluations, which may be measurable (e.g., a good review on Amazon) or invisible to the recommender (e.g., whether the ad led to a subsequent fruitful interaction between user and advertiser). To the best of our knowledge, this insight has been overlooked by the literature and might have important implications for the design of better recommendation systems. At best, this tension has been explored analytically through modelling, e.g., the analytical trade-off between revenue and quality explored in L'Ecuyer et al. (2017), where an online platform optimizes recommendations taking into consideration the long-term implication of quality on the size of the market.

Third, our work has prescriptive implications for the consumer, in the sense that our estimation results may pave alternative ways to suggest more agreeable experience trajectories to consumers. Previous research suggests that when facing a sequence of hedonic choices, it is optimal to select a trajectory of increasing experience quality (Das Gupta et al. 2015). However, our findings suggest that if one takes into account dynamic reference effects (some of which might be of lower quality) one may obtain higher utility overall, and are reminiscent of the alternating service level regime of Aflaki and Popescu (2014) in a context of consumer choices.

The rest of the paper is organized as follows. Section 2 presents theoretical framework and predictions, based on the related literature. Section 3 describes the context and data. We report and discuss the estimation of our econometric models in §4. Section 5 concludes the paper.

2 Theoretical Framework

We develop our model by first formulating a *naïve* model of consumer satisfaction without behavioral effects. Then we add reference effects to it. Finally, we distinguish choice and outcome stages embedded in the model while accounting for their interdependence. This leads to theoretical predictions that we then test empirically.

2.1 Rational choices

The simplest model for modelling consumer satisfaction requires defining the (random) utility experienced by a given individual i when choosing option j in J, the set of available choices. We can define the experienced utility of choosing j as

$$U_{ij} = \beta X_j + \varepsilon_{ij} \tag{1}$$

where X_j includes the characteristics of option j, β is a coefficient to be estimated, and ε_{ij} a random shock experienced by the consumer, with a distribution independent and common across individuals i and choices j.

We assume that ε_{ij} is observed at the time of choice. In this situation, the consumer would make different choices when the same alternatives are presented to her, depending on the realization of the random shock. This process would generate a choice probability, e.g., to the Multinomial Logit (MNL) when ε_{ij} is Gumbel-distributed. Under model (1), one can estimate the coefficient β by having access to choice characteristics and choice probabilities. In this setting, a recommendation system would simply recommend the items with higher expected utility, given that it does not have access to the consumer-specific shock ε_{ij} (see Heese and Martínez-de Albéniz 2018).

2.2 Reference effects

Hedonic goods are evaluated by a series of intangible dimensions, especially when they include subjective and intimate taste aspects. For this reason, we should expect consumer preferences to exhibit strong inter-temporal associations. In particular, because past experiences are known to influence taste (Oliver 1980, Anderson and Sullivan 1993, Baucells and Sarin 2012), future satisfaction is statistically associated with past consumption choices; they contribute to the formation of a reference against which choices are evaluated.

A well-known behavioral framework to incorporate such effects is Prospect theory (Tversky and Kahneman 1979), which suggests that expected utility is considered in relation with a reference level. Deviations from the reference are not symmetric: negative deviations are penalized more than positive deviations are appreciated, which is known as loss aversion. This phenomenon has been documented in a variety of settings (Kalyanaram and Winer 1995, Mazumdar et al. 2005) and its existence should change firm decisions such as pricing (Popescu and Wu 2007, Nasiry and Popescu 2011) or service level (Aflaki and Popescu 2014).

In our context of hedonic good sequential consumption, the reference level forms from past consumptions and is likely to get updated with each new experience. Since our empirical context is restaurant exploration, histories tend to be short (i.e., consumer tend to report few online reviews), and more vivid and salient memories, that is, those restaurant visits that generated an online review, are likely to serve as a reference in future decisions. As a consequence, we adopt last choice as a valid proxy of the reference level. However, our model is flexible to incorporate alternative specifications for the reference formation process.

Specifically, we assume that consumers use the last choice k as a reference in evaluating the experience utility of the next consumption j. The utility thus becomes

$$U_{ij}(k) = \beta X_j + f(X_j - X_k) + \varepsilon_{ij} \tag{2}$$

where $f(Y)$ accounts for the differential effect of choice j when the reference is k. As we see later in the empirical study, an example of this could be a function of quality rating differences between j and k, as well as price differences, difference in dining features, or physical distance between restaurants j and k. Given such reference effects, loss aversion would imply that, when a consumer visits a "better" choice, e.g., a restaurant with a higher quality signal compared to her quality reference level, i.e., $Y \geq 0$, then she derives an increasing, positive utility. In contrast, when the choice is "worse" relatively to her reference level, i.e., $Y \leq 0$, then the utility decreases rapidly and takes a disproportionate drop. In our study, as in most of the prospect theory literature, we use a piecewise-linear function $f(Y)$ defined as

$$f(Y) = \begin{cases} \gamma_l Y & \text{if } Y \leq 0, \\ \gamma_h Y & \text{if } Y \geq 0. \end{cases}$$

With this structure, loss aversion is equivalent to having $\gamma_l > \gamma_h$.

As before, with reference effects a recommendation system would simply recommend the items with the highest expected utility. In particular, if reference effects do not change the order of choice's expected utility, e.g., when $\gamma_l > \gamma_h \geq 0$, then it remains optimal to recommend the higher-rated choices above the lower-rated ones.

2.3 Differentiating choice and outcome

The consumption of many hedonic goods (such as going dining to a restaurant or visiting a museum) typically does not occur instantaneously and the time dimension introduces certain complexity in evaluating consumer satisfaction (Kahneman 2011). For example, the process of going dining to a given restaurant lasts for at least a few hours. It starts with the decision of a consumer to choose a particular restaurant, taking into account the existing information about it, and about relevant competing choices. Once such a choice is made, the consumer builds an expectation for the dining experience, which combines choice information, as well as individual reference effects, presumably based on its latest most salient dining experience. The visit then takes place, and according to the literature in services, see e.g., Rust and Chung (2006), consumer satisfaction is the difference between perception of the actual dining experience, and expectation formed during the choice of the restaurant to visit. This implies that, even though perceptions are usually positively associated with expectations, a positive shock in expectations might result in a decrease in consumer satisfaction. As a result, there is a conceptual difference between choices, which are made to maximize the ex-ante experience expectation, and ex-post satisfaction, the outcome of using or consuming the good chosen (Oliver 1980). This difference is at the core of our research question, which is important because it challenges the implicit assumption made by most recommendation systems that the expectation building process does not change satisfaction.

More formally, differentiating expectation and satisfaction requires us to create two utility variables. We denote $V_{ij}(k)$ the expectation formed after choosing j and having reference k, which we define as:

$$V_{ij}(k) = \beta X_j + f(X_j - X_k) + \varepsilon_{ij}. \tag{3}$$

An individual will choose j when $V_{ij}(k)$ is higher than alternatives $V_{ij'}(k)$ for $j' \neq j$. We also define $W_{ij}(k, e)$ as the actual satisfaction associated with choice j, where e is the expectation formed about j prior to consumption. Assuming that actual satisfaction is directly proportional with the perceived quality q of the choice j and negatively related with the expectation e formed about choice j, then

$$W_{ij}(k, e) = a_1 q + a_2 e + \upsilon_{ij}.$$

Assuming that the perceived quality of j is evaluated against the same reference used to make the choice and also that the expectation about choice j follows the same specification of (2), then, we can rewrite the equation above as

$$W_{ij}(k) = bX_j + g(X_j - X_k) + \upsilon_{ij}, \tag{4}$$

where b and g capture respectively choice characteristics and reference effects, net of expectation associated with choice j, which is in fact what we are looking for. In other words, the estimation of (4) will measure the total impact of choice characteristics and reference on satisfaction, already taking into account that it is the outcome of a complex decision process. We can again adopt a piecewise-linear function for g:

$$g(Y) = \begin{cases} \delta_l Y & \text{if } Y \leq 0, \\ \delta_h Y & \text{if } Y \geq 0. \end{cases}$$

Given the reference effects, our satisfaction model would imply that, when a consumer visits a "better" choice, e.g., a restaurant with a higher quality signal compared to her quality reference level, i.e., $Y \geq 0$, then she is likely to receive a high quality service which would get discounted by the high expectation associated by her choice. In contrast, our model implies that when the consumer visits a "worse" choice relatively to her reference level, i.e., $Y \leq 0$, then she is likely to receive a low-quality service which would get compensated by the negative expectation formed when deciding to visit a worse choice. We argue (and will test empirically) that because consumers disproportionally formed more negative expectations when making worse choices then, the net effect after compensating for the disproportional negative expectation is positive. In other words, because of $\gamma_l > \gamma_h$, we should expect that $\delta_l < \delta_h$.

Note that choice and satisfaction may not be independent, i.e., υ_{ij} might be correlated with ε_{ij}. This fact would introduce selection bias which must be corrected during the empirical testing of our predictions (Heckman 1979).

2.4 Theoretical predictions

Our theoretical models have been fully specified by Equations (3) and (4). These models capture, based on established theoretical grounds on individual decision making, how reference effects influence not only choice of a good (Equation 3) but also the satisfaction of such a choice (Equation 4).

On the one hand, loss aversion has been widely documented in making choices (e.g., Kalyanaram and Winer 1995, Mazumdar et al. 2005). Hence, we also expect that individuals choose hedonic goods to use or consume by avoiding options that are expected to be of lower quality, compared to their reference quality level. As a result, in Equation (3), theory would predict that $f(Y)$ exhibits loss aversion, i.e., $\gamma_l > \gamma_h$. The left panel of Figure 1 captures such a prediction by plotting how the probability of choosing item j increases with the observable quality signal of the item in the selection set. Such a plot shows a significant discontinuity at the quality reference level so that the effect of the quality signal is more pronounced in the regime in which items exhibit a quality signal lower than the quality reference level. Such a steeper slope in the "worse quality" region captures the loss aversion effect.

On the other hand, because expectation negatively affects experienced utility, the satisfaction of using or consuming a chosen hedonic good should be affected by two opposing effects. First, the perception of the quality of the good should be proportional to the actual quality of the good. Even though individuals may evaluate such a quality with respect to their quality reference, there is not theoretical reason to expect a disproportionally different effect (from the ex-ante expectation process) of the quality reference on the actual perception of the quality of the good. The second effect is the negative effect that the expectation level brought to the use or consumption of the good has on the experienced utility. Because "worse than the reference" goods involve a disproportionally lower expectation about the potential satisfaction they could get from it, then one would expect a higher discount due to the effect of negative expectations leading to higher than expected experienced utility for hedonic goods in the worse-quality regime. As a result, we expect that experienced utility exhibits gain seeking, i.e., $\delta_h > \delta_l$ in Equation (4). The right panel of Figure 1 captures such a prediction by plotting the net effect of the quality of a given good on the probability of experiencing outstanding satisfaction by using or consuming the chosen good j. As expected, as the observable quality level of the chosen good increases, then the probability of having an outstanding satisfaction with using the good increases. However, such a positive effect decreases for goods that have worse (than the quality reference) quality. Such a flatter slope captures the positive effect that using a good for which disproportionally low expectations have been built during its selection process.

In our empirical study, we thus expect worse restaurants to be chosen disproportionally less, but at the same time the satisfaction derived from them to be less sensitive to quality. In other words, bad choices, of worse restaurants, result in a not-so-bad outcome, by means of expectation adjustment: the consumer already discounts the (to be expected) bad experience and hence does not perceive it as so bad. This is what we call the *value of low expectations*.

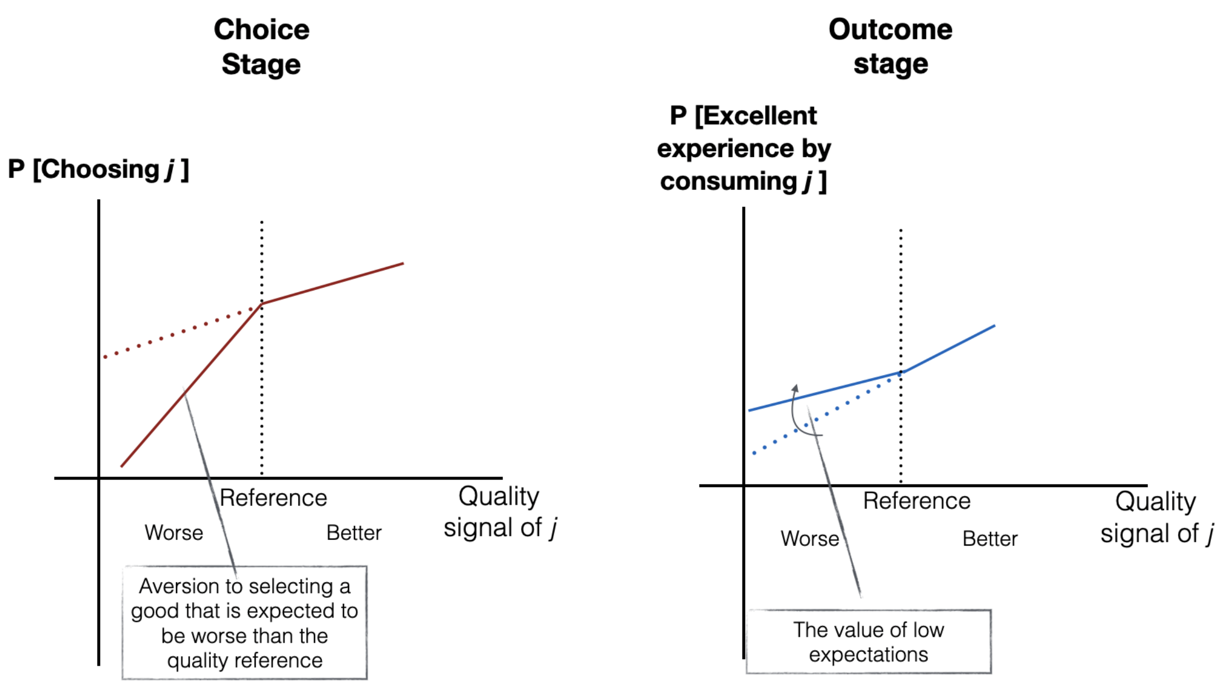

Figure 1: Theoretical predictions concerning choice and outcome of hedonic goods

3 Empirical Set-up

3.1 Data

We have obtained large samples of consumer generated reviews about all the restaurants in Barcelona (Spain), which is considered, along with Paris (France), among the most influential cities in the gastronomical scene (Love Exploring 2019). We collected data from the two most popular review platforms in this industry: TripAdvisor.com and TheFork.com. The data collected is open and public information available on these platforms to any consumer, for free and without doing any registration on these online platforms.

Past research has established that reviews might contain different type of information depending on whether they are spontaneously provided by the consumer, or are solicited by a given establishment – typically hotels– with an email request (Magno et al. 2018). In case of restaurants, it is highly unlikely the existence of solicited reviews because restaurant managers usually do not have TripAdvisor's user's email information and hence it is unlikely for them to request consumers to provide written feedback on their experience after a restaurant visit. In addition, TripAdvisor reviews are not posted on the website immediately after being written; they are subjected to a verification process that aims to detect suspicious posting patterns. Such a process takes into account the IP and email address of the reviewer, and sometimes requires a proof of purchase in the form of a payment bill. Although we cannot entirely rule out the absence of fake reviews in the TripAdvisor data, previous research using TripAdvisor data suggests that the existent of such illicit reviews is insignificant (Hollenbeck et al. 2019).

In order to test for the robustness of our results to the way reviews are posted in TripAdvisor, we also run our analysis on the dataset obtained from TheFork. TheFork provides the ideal dataset for robustness check because only only consumers who have booked the table through TheFork and effectively visited the restaurant can post a review, where the reservation date is given by the system. Furthermore, because every posting is connected to a reservation for which the restaurant must pay a commission to TheFork, positive or negative fake reviews have a high cost and hence are less likely to occur. However, due to this commercial relationship between TheFork and the restaurants, the number of restaurants listed on this website is smaller than that in TripAdvisor, and certainly not exhaustive: many restaurants cannot be booked via TheFork, and thus are not available on this platform.

Our main analysis is performed on a large sample of restaurant reviews from Barcelona (Spain). TripAdvisor's complete data set in Barcelona includes 1,015,590 reviews from 485,844 consumers for 9,509 restaurants between 2004 and 2017. However, many of the restaurants barely have any reviews, and their evolution cannot be tracked over time due to lack of observations. Hence, we select restaurants with at least 20 reviews per year from consumers with more than one review per year (Allon and Zhang 2017).

We create consumer "trajectories" by linking reviews that appear in consecutive order for the same consumer. For example, consider a reviewer that wrote 4 reviews in total; such a trajectory would result in three valid reviews for our analysis (reviews number 2, 3 and 4), where the consumer had a previous restaurant in mind. In addition, we remove consumers that posted several reviews on the same day (so that we cannot determine the order in which restaurants were visited) and exclude first-time reviews (so that all observations have a predecessor). This leaves us with a final sample of 83,274 reviews by 49,139 consumers in 2,534 restaurants in Barcelona.

We replicate the same analysis with two additional samples. First, and as mentioned, we test the robustness of our findings to the characteristics of the consumer review platform by analyzing a sample of restaurants from Barcelona featured in TheFork.com. As expected, there are fewer restaurants observed, and users tend to have longer histories; after applying a identical filtering criteria, the sample comprises 202,383 reviews from 54,447 consumers in 1,885 restaurants during the same time period. We also test the robustness of our findings to a different location and culture by using a TripAdvisor's sample from restaurant reviews in Paris during the same time period. Using the same filtering criteria as described above the TripAdvisor's Paris dataset provides a sample of 124,943 reviews from 66,646 consumers in 3,499 restaurants.

Our samples are built based on the following information displayed on each restaurant's webpage hosted by the two review platforms studied:

- **Reviews**. Each review consists on an integer mark, on a scale from 1 to 5 in TripAdvisor (1 to 10 in The Fork), accompanied by a posting date, a title and a comment. It is connected to a restaurant and a reviewer (with unique IDs). Note that because we have access to all the reviews made for the restaurants of interest in a given city, we can reconstruct the review history of each reviewer that gets included in our sample.

- **Restaurants**. For each restaurant, we obtain three types of data: location, price category, and restaurant features. With the restaurants' displayed addresses on TripAdvisor (and TheFork) we obtained, using Google's Geocoding API, the geographical coordinates of each of the restaurants in our dataset. With respect to price, each restaurant's page in these platform displays price information either based on a price bracket (e.g., EUR 15-30) or a price tag (e.g, Cheap, Medium or Expensive), or both. As for restaurant features, restaurants' page display primarily the types of cuisine it serves (e.g., Mediterranean, Italian, Japanese) and also the types of meals it offers (e.g., breakfast, lunch and/or dinner).

3.2 Model specification

Because we want to study the determinants of satisfaction associated with dining in given restaurant conditional on having selected such a restaurant (instead of any other restaurant in the selection set), we use a two-stage Heckman selection correction model (Heckman 1979). Such a model corrects any selection bias from a non-randomly selected sample. In our case we have a non-randomly selected sample of restaurants that have been selected by the consumers in our (nearly exhaustive) large sample of restaurant reviews. The Heckman correction is implemented in a two-stage regression. In the first stage, we model the selection process by which consumers select a restaurant to visit next. In the second stage, we model the satisfaction achieved after having visited the selected restaurant conditional on having made such a selection. This two-stage statistical approach allows us to properly estimate the probability of having an outstanding restaurant dining experience using a non-randomly selected sample of restaurant reviews.

Formally, the first stage corresponds to a selection model in which, based on our theoretical arguments, we estimate the probability that a given restaurant j is selected by consumer i. Because the Heckman correction method involves a normality assumption, the first stage selection model must be a probit regression of the form

$$Prob(D_{ij} = 1|X_{ij}) = \Phi(\gamma X_{ij}) \tag{5}$$

where D_{ij} indicates whether the option was selected ($D_{ij} = 1$ if consumer i selects restaurant j and $D_{ij} = 0$ otherwise), X_{ij} is the vector of explanatory variables, γ is a vector of unknown parameters, and Φ is the cumulative distribution of the standard normal distribution. Estimating such a selection model results in predictions of the probability of restaurant j being selected to be visited by consumer i.

Note that this specification does not preclude the possibility of a consumer choosing more than one restaurant in a given sample path. For this reason, we introduce, within the covariate X_{ij}, information about how much choice was available to consumer i to control for the amount of competition for restaurant j. We nevertheless compare this approach to the multinomial logistic (MNL) choice model (Anderson et al. 1992) in our robustness study, and find that our chosen approach does not introduce any qualitative difference in our conclusions.

In the second (outcome) stage, we correct for the non-random selection of restaurants by including a transformation of the restaurant selection probabilities estimated in the first stage as an additional explanatory variable (Heckman 1979). First, we estimate the probability of having an excellent dining experience in restaurant j as follows

$$Prob(W_{ij}^* = 1|X_{ij}) = \Phi(\delta X_{ij}) \tag{6}$$

where W_{ij}^* indicates whether a given visit to restaurant j resulted in an outstanding dining experience ($W_{ij}^* = 1$ if consumer i had an excellent dining experience and $W_{ij}^* = 0$ otherwise). Since W_{ij}^* is a binary variable, we use a probit formulation to estimate such a probability (Van de Ven and Van Praag 1981), so that W_{ij} follows a normal distribution $W_{ij} = \delta X_{ij} + \varsigma_{ij}$. Because such an outcome is only observed when consumer i chooses to visit restaurant j instead of any other restaurant, the conditional probability of an excellent dining experience given that consumer i does choose to visit restaurant j can be expressed as

$$Prob(W_{ij} = 1|X_{ij}, D_{ij} = 1) = \Phi(\delta X_{ij} + E[\varsigma_{ij}|X_{ij}, D_{ij} = 1]) \tag{7}$$

Then, assuming that the error terms of the first and second stage are jointly normal, the outcome model is specified as follows

$$Prob(W^*_{ij} = 1|X_{ij}) = \Phi\Big(\delta X_{ij} + \rho\sigma_\varsigma\lambda(\gamma X_{ij})\Big) \tag{8}$$

where ρ is the correlation between the unobserved determinants of selecting restaurant j and the unobserved determinants of consumer's satisfaction after dining at restaurant j; σ_ς is the standard deviation of ς_{ij}, and λ is the inverse Mills ratio evaluated at γX_{ij}. This regression can be estimated by constructing the λ term by replacing γ with Probit estimates from the first stage and then including it as an additional explanatory variable in the outcome regression. Since the second stage is also a probit, then the two-stage model is called a heckprobit. (For an introduction and discussion of the heckprobit as an extension of the Heckman correction model refer to Van de Ven and Van Praag 1981.) Note that since $\sigma_\varsigma > 0$, the coefficient on λ can only be zero if $\rho = 0$, so testing the null that the coefficient on λ is zero is equivalent to testing for any sample selection bias. Moreover, a positive coefficient on λ would indicate that people who select a restaurant following the prediction of our selection regression are more likely to have a positive dining experience (positive selection bias) whereas a negative coefficient on λ would indicate that selecting a restaurant according to our selection model is less likely to have a positive dining experience (negative selection bias).

Note that to obtain credible estimates from this two-stage regression it is crucial to have a valid exclusion restriction. That is, we need (at least) one variable that appears as a significant determinant in the selection stage but does not appear at all in the second stage. Fortunately, we have a set of variables that fulfil such exclusion restriction in our setting.

3.3 Variables

The unit of analysis is the visit at time t to a given focal restaurant in terms of (i) a consumer selecting or not such a restaurant to visit, and (ii) the perceived quality of the experience corresponding to such a visit from the consumer's viewpoint. Having said this, the last restaurant visited is also important because it is the main reference we use to predict both the probability of selecting the focal restaurant as well as the perceived quality of the experience in the focal restaurant.

3.3.1 Dependent variables

Restaurant visit. The dependent variable in our selection model is whether or not a focal restaurant is visited ($restaurant\_visit$). Note that the available selection set for a given consumer is comprised by the set of restaurants that were active at time t when the consumer selected restaurant j as the "next" restaurant to visit. In addition, because reference effects are important in our estimation we only consider restaurant visits from consumers that reported online her "last" dining experience.

Restaurant perceived quality. The dependent variable in the outcome model is a function of the perceived quality of the experience in the focal restaurant visited (and reported online) by the focal consumer. Specifically, our dependent variable captures whether or not the dining experience of the restaurant visited was "excellent" (*excellent_dining*). Given how positively skewed online reviews typically are (Dellarocas 2003, Ghose and Ipeirotis 2006), we focus our attention on understanding the determinants of nearly perfect dining experience, which typically accounts for nearly half of the consumer reviews: in the TripAdvisor samples, we define an excellent dining experience as one that yields a rating of 5 stars, which occurs in about 45% of the sample (in both Barcelona and Paris). The equivalent excellent dining experience in TheFork platform would be those receiving a mark of 9.5 or 10 which comprises 41% of the TheFork sample.

3.3.2 Independent variables

Because the last restaurant visit is the most concrete and recent empirical evidence that captures the location, price, and quality reference of a given consumer at time t, we use it as the proxy from which we measure deviations (from the reference) in location, price, and quality. As a result, all our key predictive variables in both the selection and outcome models are measured relative to the last restaurant visited.

Distance from last restaurant. Consumers might tend to visit restaurants which are in the same neighborhood, either because it is convenient for them (e.g., both restaurants are close to their home or workplace) or because the restaurants are in a more 'active' zone that they enjoy, where many social activities take place (e.g., a shopping mall or a busy commercial street). Because we expect that the probabilities of selecting the focal restaurant and having an excellent dining experience to decay with distance, we define *log_distance* as the logarithm of the distance between last restaurant visited and the possible focal restaurant to visit. Additionally, we define the dummy variable *same_zip* which indicates whether any two given restaurants have the same ZIP code to additionally control for restaurants sharing the same neighborhood in addition to being close to each other.

Same price category. Because we have data on the price category to which any restaurant belongs to, we are able to define *same_price* as a binary variable to capture whether the last restaurant visited and the possible focal restaurant are in the same price category (or not).

Common dining features. Because both TripAdvisor and The Fork list the food types associated with any restaurant listing they host based on dietary choices (e.g, Vegetarian, Vegan), food styles (e.g., Japanese, Mediterranean, Italian), or food restrictions (e.g., Gluten Free), we create a variable $common\_features$ that counts the number of common features between the last restaurant and the possible focal restaurant.

Expected restaurant quality. The expected quality of a focal restaurant is likely to significantly influence both the selection and outcome of a visit to such restaurant. Consumers may assess the expected quality of a restaurant by checking the focal restaurant's cumulative rating on the online recommendation platform they use. Hence, we define $restaurant\_rating\_next$ as the average rating of a focal restaurant on the first day of the month in which the potential visit may occur, i.e., by averaging all ratings given before that day.

Relative quality difference. Because a given restaurant rating may be perceived differently by different consumers, it is important to measure such a rating relative to a meaningful and recent reference of restaurant quality. We use the quality of the last restaurant visit as a proxy of such a quality reference. Hence, we define a variable which captures the difference in cumulative ratings of the focal and last restaurant visited. To do so, we define $restaurant\_rating\_last$ as the cumulative rating of the last restaurant visited (also measured on the first day of the month in which the visit to the next restaurant takes place). Then, we define $delta\_quality = restaurant\_rating\_next - restaurant\_rating\_last$. As a result, $delta\_quality$ defines two regimes which would allow us to test whether people are risk averse or gain seekers with respect to quality. That is, $positive\_delta\_quality$ captures when the potential focal restaurant is of better expected quality than their quality reference (as proxied by the quality of the last restaurant visited) whereas $negative\_delta\_quality$ captures when the potential focal restaurant is worse than their quality reference, as measured by the quality of last restaurant visited. Hence, someone avoiding visiting a worse restaurant and therefore visiting a better restaurant is both loss averse and gain seeking. To test for marginal effect of loss aversion in our models, we define $negative\_delta\_quality = delta\_quality$ when $delta\_quality \leq 0$, and zero otherwise.

Figure 2 shows the histogram of $delta\_quality$ for the TripAdvisor-Barcelona sample. Using other samples show similar histograms. When this variable is negative then the focal restaurant is of lower expected quality than the last restaurant visited; the opposite is captured by a positive $delta\_quality$.

Competition variables (exclusion restriction). To be able to estimate an outcome model conditional to a selection model in the first stage following the Heckman selection correction approach (Heckman 1979), it is critical to define an exclusion restriction. That is, a set of vari-

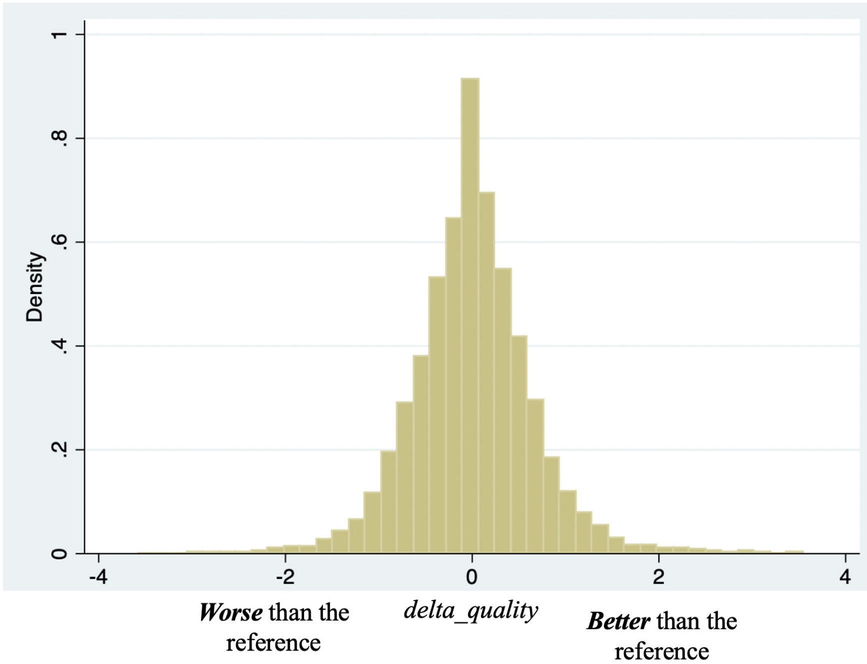

Figure 2: Histogram of *delta_quality* based on TripAdvisor-Barcelona sample

ables that significantly influence selection but not outcome. In our context, competitive choices provide the ideal exclusion restriction because the more competition a focal restaurant faces the less likely that it would be selected; however, once a focal restaurant is chosen then competition should not significantly influence the quality of the dining experience within the focal restaurant. As a result, we define three competition variables based on the number of restaurants in the vicinity of the focal restaurant. More specifically, we define *distance_competition* as the number of restaurants within 1 km of the focal restaurant; we define *same_price_competition* as the number of restaurants in the city within the same price category as the focal restaurant; we define *common_features_competition* as the number of restaurants in the city with the same set of restaurant features as the focal restaurant.

Contextual control variables. We include two contextual variables to control for time separation with respect to the last restaurant visit and popularity of the focal restaurant. In the outcome model, we account for the (logged) number of days between the last restaurant visit as reported online, and the visit to the focal restaurant, and denoted *log_timegap_days*. This is important because one could argue that the reference effect of the last restaurant may decay with time. (Note that this control variable is not defined for the restaurants in the selection set that are not chosen.) We also control, in both selection and outcome models, for the popularity of the focal restaurant by counting the (log) number of reviews that it has received on the online platform up on the first day of the month of the visit to the focal next restaurant, denoted *log_restaurant_comment_next*.

Past user experience. There might be consumers who are more critical than others, or who have a more sophisticated gastronomical taste, or have budgetary constraints. These user characteristics might affect the way a consumer experiences the quality in the focal restaurant and therefore are included as important controls in both selection and outcome models. Hence, we define *user_rating_last* as the consumer's past average rating given to all of the restaurants up to the last restaurant visit; we also define *user_comments_last* as the cumulated number of reviews posted by the consumer by until the last restaurant visit. We report in Table 1 summary descriptive statistics and pairwise correlations of all these variables.

Finally, note that the full samples are quite large in the first stage where the probability of each possible pairs (consumer i, restaurant j) is estimated. For this reason, in the estimation of the model, we used a random subsample of choices, of size 15,000 for TripAdvisor-Barcelona, 20,000 for TheFork-Barcelona and 12,000 for TripAdvisor-Paris (as large as our computing power permitted). This led to random samples of about 25 million, 12.5 million and 28 million observations for estimating the selection stage of the TripAdvisor-Barcelona, TheFork-Barcelona, and TripAdvisor-Paris respectively. We replicated the sampling several times leading to identical conclusions.

4 Results

4.1 Selection Model

In order to correct for restaurant selection while estimating the factors that determine the satisfaction of the dining experience in the selected restaurant, it is imperative to estimate a probit selection model (Heckman 1979). Table 2 shows the results of the selection models that predict the probit of selecting a possible focal restaurant. Models 1 to 4 are partial models, while model 5 is the full model estimated using the TripAdvisor-Barcelona sample. To show robustness of our results, models 6 and 7 are the full models for the TheFork-Barcelona and TripAdvisor-Paris samples, respectively.

Model 1 includes the effects of restaurant popularity and proximity to the last restaurant visited. As expected, those restaurants that are more popular as measured by a larger number of reviews are more likely to be selected as the next restaurant to visit (0.0730, $p < 0.001$). We measure proximity between the possible next restaurant (the focal restaurant) and the last restaurant visited along geographic, price, and restaurant features dimensions. Our results show that the more proximate the focal restaurant to the last restaurant is, the higher the likelihood of being selected for the next restaurant visit. More specifically, when the focal restaurant is in the same zip code as the last restaurant, it is more likely to be selected (0.177, $p < 0.001$); moreover the effect of geographic separation between the focal and the last restaurant is significantly negative and drops with an exponential decay (-0.185, $p < 0.001$). As robustness, we also created dummy variables, which

Table 1: Summary statistics and pairwise correlations based on TripAdvisor-Barcelona sample

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. excellent_dining	1														
2. log_distance	-0.09	1													
3. same_zip	0.08	-0.6	1												
4. same_price	-0.02	-0.07	0.09	1											
5. common_features	0.05	-0.1	0.14	0.12	1										
6. restaurant_rating_next	0.32	-0.05	0.01	-0.05	0.03	1									
7. delta_quality	0.19	0.00	0.00	-0.01	-0.01	0.62	1								
8. negative_delta_quality	0.2	-0.06	0.05	0.02	0.03	0.71	0.81	1							
9. distance_competition	0.02	-0.34	0.08	0.01	0.00	0.06	0.06	0.05	1						
10. price_competition	-0.06	0.00	0.01	0.45	0.08	-0.05	-0.02	-0.01	0.00	1					
11. common_features competition	-0.04	0.05	0.00	0.04	0.37	-0.1	-0.06	-0.07	-0.08	0.11	1				
12. log_restaurant_comments_next	0.02	-0.04	-0.01	0.04	0.06	0.03	0.01	0.08	0.18	0.06	-0.03	1			
13. log_timegap_days	0.01	0.07	0.00	0.00	0.03	0.03	-0.01	0.01	-0.14	-0.02	0.01	-0.08	1		
14. user_rating_last	0.3	-0.07	0.06	0.00	0.07	0.14	-0.19	-0.04	0.01	-0.02	-0.01	0.04	-0.02	1	
15. user_comments_last	-0.09	0.1	-0.06	-0.02	-0.04	0.00	-0.01	0.00	-0.07	0.00	0.01	-0.05	0.08	-0.03	1
Mean	0.45	1	0.17	0.64	0.55	4.05	0.01	-0.23	4.47	6.85	1.64	4.95	3.33	4.01	2.61
Standard deviation	0.5	0.53	0.38	0.48	1.07	0.48	0.65	0.38	1.02	1.54	2.39	1.36	1.93	0.97	2.51

$N = 14,900$ restaurant visits. Correlations greater than $|0.02|$ are significant at 0.05 level.

classified the distance between any two restaurants into four categories: 0 < distance < 500 m, distance < 1000 m, distance < 1500 m, distance > 1500 m. We confirm with this analysis that the exponential decay effect of distance also appears when using such indicator variables, instead of *log_distance*. Furthermore, when the focal restaurant is positioned in the same price category as the last restaurant visited, then it is more likely to be selected (0.0416, $p < 0.001$), and finally the larger the number of restaurant features that are common between the focal and last restaurant visited, the higher the chances of the focal restaurant being selected (0.0419, $p < 0.001$). This also indicates that the last restaurant visited indeed seems to be a valid reference when estimating the effects of deviating from a restaurant reference on the probability of a focal restaurant being selected as the next restaurant to visit. Moreover, proximity to such a reference increases the probability of being selected as the next restaurant to visit. All these effects are consistent across all the models shown in Table 2.

Model 2 adds competition variables to the simpler selection models. Our results show, as expected, negative and significant coefficients for the effects of geographic, price, and restaurant feature competition. This indicates that, keeping all other factors constant, the more competition a focal restaurant faces (measured by the number of other restaurants in the proximity of the focal restaurant), the less likely it is to be selected for the next visit. Even though these effects are stable in models 2-5, the effect of $common\_feature\_competition$ is negative but not significant in the TheFork-Barcelona sample (Model 6). This suggests that the effect of competition is more robust when measured based on geographic and price proximity (to a given reference).

Model 3 adds the effect of the expected quality of the focal restaurant. Model 3 shows a positive and significant coefficient for *restaurant_rating_next* (0.0758, $p < 0.001$) which indicates that the higher the expected quality of the focal restaurant (as captured by the average rating of the restaurant at the beginning of the month when the restaurant selection was to be made), the more likely that it will be selected for the next visit. Note that this baseline effect is also positive and significant when estimating analogous models with TheFork-Barcelona (0.0546, $p < 0.001$) and TripAdvisor-Paris (0.0347, $p < 0.001$), not shown in Table 2 for brevity.

In Model 4, we include the effect of *delta_quality* which depends on the quality of the last restaurant visited. By including the effect of *delta_quality* we are capturing the effect of *restaurant_rating_next* due to the difference in quality with the last restaurant visited. Note that *delta_quality* is positive for cases in which the focal restaurant has expected quality better than the last restaurant visited, and negative otherwise. Model 4 includes a positive and significant effect of *delta_quality* (0.0126, $p < 0.013$) which indicates that there is a significant positive effect attributed to the difference in quality between the next possible restaurant and the last one visited. That is, *delta_quality* increases the chances of being selected for relatively better restaurants and reduces such chances for relatively worse restaurants.

Model 5 finally considers the effect of *delta_quality* in the two regimes defined by positive and

Table 2: Probit models predicting the probability of selecting a focal restaurant

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	TripAdvisor	TripAdvisor	TripAdvisor	TripAdvisor	TripAdvisor	TheFork	TripAdvisor
	Barcelona	Barcelona	Barcelona	Barcelona	Barcelona	Barcelona	Paris
user_	0.00178	0.00190*	0.00187*	0.00195*	0.00172	0.000855***	0.000949
comments_last	(-0.000945)	(-0.000946)	(-0.000948)	(-0.000947)	(-0.000949)	(-0.00018)	(-0.000871)
user_rating_last	-0.00520*	-0.00558*	-0.00600*	-0.00365	-0.00376	-0.00983***	-0.0036
	(-0.00242)	(-0.00243)	(-0.00243)	(-0.00261)	(-0.0026)	(-0.00236)	(-0.00298)
log_restaurant_	0.0730***	0.0801***	0.0799***	0.0800***	0.0778***	0.142***	0.0366***
comments_next	(-0.00209)	(-0.00213)	(-0.00212)	(-0.00212)	(-0.00213)	(-0.00174)	(-0.00254)
same_zip	0.177***	0.163***	0.166***	0.165***	0.165***	0.271***	0.215***
	(-0.00791)	(-0.00795)	(-0.00796)	(-0.00796)	(-0.00797)	(-0.0083)	(-0.0088)
log_distance	-0.185***	-0.237***	-0.236***	-0.236***	-0.235***	-0.239***	-0.200***
	(-0.00591)	(-0.00646)	(-0.00647)	(-0.00647)	(-0.00648)	(-0.00607)	(-0.00614)
same_price	0.0416***	0.0628***	0.0684***	0.0676***	0.0647***	0.0650***	0.0743***
	(-0.00491)	(-0.00564)	(-0.00566)	(-0.00567)	(-0.00568)	(-0.00593)	(-0.006)
common_features	0.0419***	0.0490***	0.0464***	0.0465***	0.0465***	0.0480***	0.0580***
	(-0.00237)	(-0.00262)	(-0.00262)	(-0.00262)	(-0.00262)	(-0.00235)	(-0.00444)
distance_		-0.0564***	-0.0566***	-0.0567***	-0.0559***	-0.0497***	-0.0499***
competition		(-0.00261)	(-0.00262)	(-0.00262)	(-0.00263)	(-0.00266)	(-0.00316)
same_price_		-0.0135***	-0.0137***	-0.0136***	-0.0137***	-0.0288***	-0.0219***
competition		(-0.00174)	(-0.00176)	(-0.00176)	(-0.00177)	(-0.00143)	(-0.00173)
common_features_		-0.00753***	-0.00625***	-0.00627***	-0.00635***	-0.00074	-0.00508***
competition		(-0.00108)	(-0.00108)	(-0.00108)	(-0.00108)	(-0.00102)	(-0.00102)
restaurant_			0.0758***	0.0634***	0.0253**	-0.0330***	0.00131
rating_next			(-0.00509)	(-0.00712)	(-0.00796)	(-0.00763)	(-0.00845)
delta_quality				0.0126*	-0.0334***	-0.0408***	-0.0547***
				(-0.00505)	(-0.00686)	(-0.00782)	(-0.0082)
negative_					0.130***	0.238***	0.148***
delta_quality					(-0.0124)	(-0.0131)	(-0.0138)
Constant	-3.429***	-3.063***	-3.369***	-3.328***	-3.132***	-2.673***	-2.895***
	(-0.0165)	(-0.0232)	(-0.0312)	(-0.0352)	(-0.0397)	(-0.0684)	(-0.0428)
AIC	246,103	245,558	245,327	245,323	245,212	269,782	205,984
BIC	246,224	245,723	245,508	245,518	245,423	269,983	206,196
Observations	24,995,369	24,995,369	24,995,369	24,995,369	24,995,369	12,486,756	28,259,563

Standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure 3: Probit of selecting a restaurant as a function of its expected quality (based on Model 5)

negative *delta_quality*. By including the term *negative_delta_quality*, we see that the effect of *delta_quality* in the "worse than before" regime is positive and significant (and much larger in absolute terms than the other terms that comprise the overall effect of *restaurant_rating_next*).

Figure 3 shows the effect of *restaurant_rating_next* on the probit of selecting a next possible restaurant based on Model 5. The plot shows a positive effect that captures the overall effect of *restaurant_rating_next* (red line with a positive and significant slope of 0.0253, $p < 0.01$). Based on Model 5, the plot also shows the significantly different effects of *restaurant_rating_next* in the "worse" and "better" regions. For the "better" region, the net effect of *restaurant_rating_next* is not significantly different than zero (-0.0081, $p < 0.396$) while for the "worse" region the effect of *restaurant_rating_next* is positive and significant (0.122, $p < 0.001$).

Overall, our probit models support the notion that people show a disproportional stronger tendency to avoid selecting restaurants of "worse" (than their quality reference) expected quality than to seek restaurants of "better" (than the quality reference) expected quality. Models 6 and 7, based on alternative samples, show fully consistent results. This corroborates the existence of loss aversion, as predicted in §2.4.

4.2 Outcome Model

We can now estimate the probit models that predict the probability of having an excellent dining experience in a given selected restaurant, conditional to having selected such a restaurant instead of any other one in the sample. For the TripAdvisor samples, an excellent dining experience corresponds to a 5-star dining experience, whereas an excellent dining experience in TheFork platform would be those receiving a mark of 9.5 or 10.

Table 3 shows the results of our regression models corresponding to the second (outcome) stage of the Heckman estimation procedure. Models 8 to 10 are partial models using the TripAdvisor-

Table 3: Probit models predicting the probability of having an excellent dining experience

	(8) TripAdvisor Barcelona	(9) TripAdvisor Barcelona	(10) TripAdvisor Barcelona	(11) TripAdvisor Barcelona	(12) TheFork Barcelona	(13) TripAdvisor Paris
user_comments_last	-0.0498***	-0.0466***	-0.0449***	-0.0455***	-0.00445***	-0.0209***
	(-0.00471)	(-0.00481)	(-0.00481)	(-0.00482)	(-0.000827)	(-0.00436)
user_rating_last	0.414***	0.391***	0.448***	0.454***	0.505***	0.477***
	(-0.0116)	(-0.012)	(-0.013)	(-0.0131)	(-0.0113)	(-0.0157)
log_restaurant_comments_next	-0.211***	-0.0333*	-0.0252	-0.0550***	-0.011	-0.013
	(-0.014)	(-0.0155)	(-0.0156)	(-0.0164)	(-0.0214)	(-0.0131)
log_timegap_days	0.0054	0.00915	0.0127*	0.0121*	0.0116	-0.00564
	(-0.00565)	(-0.00581)	(-0.00584)	(-0.00584)	(-0.00627)	(-0.00671)
same_zip	-0.431***	0.0639	0.0691	-0.0171	0.109	-0.074
	(-0.0468)	(-0.0509)	(-0.0511)	(-0.0532)	(-0.0579)	(-0.0704)
log_distance	0.440***	0.00966	-0.021	0.0686	0.0319	0.0141
	(-0.0398)	(-0.0437)	(-0.044)	(-0.0465)	(-0.0375)	(-0.0493)
same_price	-0.209***	-0.0540*	-0.0591*	-0.0727**	-0.00252	-0.00285
	(-0.0236)	(-0.0248)	(-0.0249)	(-0.025)	(-0.0211)	(-0.0282)
common_features	-0.0907***	0.00791	0.0141	-0.00465	-0.000497	0.0233
	(-0.0122)	(-0.013)	(-0.013)	(-0.0134)	(-0.0106)	(-0.022)
restaurant_rating_next		0.933***	0.706***	0.766***	0.519***	0.758***
		(-0.0292)	(-0.0348)	(-0.0363)	(-0.0311)	(-0.0394)
delta_quality			0.281***	0.413***	0.316***	0.382***
			(-0.0238)	(-0.0327)	(-0.0325)	(-0.0412)
negative_delta_quality				-0.399***	-0.207**	-0.290***
				(-0.0675)	(-0.069)	(-0.0786)
Mills ratio	-3.064***	-0.492*	-0.386*	-0.914***	-0.241	-0.616*
	(-0.168)	(-0.192)	(-0.193)	(-0.213)	(-0.158)	(-0.25)
Constant	9.749***	-3.547***	-3.253***	-1.695*	-8.409***	-2.992***
	(-0.624)	(-0.768)	(-0.771)	(-0.815)	(-0.595)	(-0.908)
AIC	18,558	17,445	17,306	17,273	22,313	13,684
BIC	18,634	17,529	17,397	17,372	22,415	13,780
Observations	14,900	14,900	14,900	14,900	19,063	12,003

Standard errors in parentheses. $*p < 0.05, **\,p < 0.01, ***p < 0.001$.

Barcelona sample while models 11, 12, and 13 are full models for the TripAdvisor-Barcelona, TheFork-Barcelona, and TripAdvisor-Paris samples, respectively. Model 8 estimates a baseline model that controls for the past consumer experience in the reviewing platform as well as the popularity of the restaurant visited. We see that consumers who are more active on review platforms exhibit a lower tendency to report an excellent dining experience (-0.0498, $p < 0.001$), however those who have reported excellent experience in the past are more likely to report an excellent experience in the selected restaurant (0.414, $p < 0.001$). In addition, when the selected restaurant is "popular" (as measured by the number of comments received in the review platform by the time of the visit), then the likelihood of having an excellent experience is significantly lower than average (-0.211, $p < 0.001$). As for the rest of the control variables, we account for the influence of (contextual) factors of the selected restaurant relatively to the last one visited. Specifically, these models include for the effect of the time elapsed since the last visit, and the effects of geographic, price, and restaurant features proximity. The non-significant effect of *log_timegap_days* since last restaurant visit suggests that the time since last restaurant visit (which was also reviewed on the platform) is not associated with any higher chances of reporting an excellent dining experience. Similarly, the other "proximity" control variables measured show non-consistent (across models) significant effects on the probit of having an excellent dining experience. This further suggests that contextual reference effects have little influence on the dining experience.

Models 9, 10, and 11 include the effects of expected quality of the selected restaurant. Model 9 includes the base line of *restaurant_rating_next* which is positive and significant across the models indicating that the probability of having an excellent experience in the selected restaurant is directly proportional to its expected quality (as captured by the aggregated rating in the platform during the month of the visit). Model 10 adds a positive and significant effect of *delta_quality* which indicates that above the base line effect of *restaurant_rating_next* consumers have a higher probability of having an excellent dining experience when the quality of the selected restaurant is better than the last one visited. Conversely, because the effect of *delta_quality* is positive when the visited focal restaurant had a worse quality than the last one visited, then such a positive *delta_quality* effect reduces the probability of having an excellent dining experience when visiting a restaurant of worse quality than the last one visited. Hence, according to this model there is a "gain" in the probability of having an excellent experience by visiting a restaurant of expected quality being "better" than the consumer's quality reference; while there is a decrease in the probability of having an excellent experience by visiting a restaurant of expected quality being "worse" than the consumer's quality reference. Interestingly, and consistent with our theoretical arguments put forward in §2.4, Model 11 shows that the effect of *delta_quality* is not linear. The negative and significant effect of *negative_delta_quality* reduces the effect of *delta_quality* for the cases in which the selected restaurant is worse than the last one visited. That is, the quality lost for visiting a worse restaurant is not as deep as the quality gain to go to a better restaurant.

Figure 4 plots the effect of expected quality on the probability of reporting an excellent dining experience on the selected restaurant based on Model 11. As indicated by the dashed line, there is a positive and significant baseline effect of *restaurant_rating_next* (0.766, $p < 0.001$). Yet, when the selected restaurant is better than the last one visited there is a positive and significant increment (0.413, $p < 0.001$) of the effect of *restaurant_rating_next*; such an increment is due to the effect of positive *delta_quality*. However, we do not see an equivalent loss due to the effect of negative *delta_quality*, because the negative and significant coefficient of *negative_delta_quality* (-0.399, $p < 0.001$) reduces significantly the effect of the possible effect of *delta_quality* for the regime of "worse" than expected restaurants. That is, there is no significant penalty (0.014, not significant) for selecting a restaurant that has an expected quality that is worse than the last one visited. This is fully consistent with our theoretical argument suggesting that because there is a negative expectation built in when visiting a restaurant with "worse than the quality reference" then the chances of being "pleasantly surprised" increases the chances of having an excellent dining experience. These results are qualitatively similar in the full models estimated with TheFork-Barcelona and TripAdvisor-Paris samples (Models 12 and 13 respectively).

Figure 4: Probit of having an excellent dining experience as a function of expected quality of selected restaurant (based on Model 11)

All models shown in Table 3 control for any selection bias by including a Mills ratio (as specified by Heckman correction estimation). The Mills ratio controls for any selection bias associated with selecting a restaurant according to the corresponding full models shown in Table 2 (either Model 5, 6 or 7, depending on the sample used). A positive Mills ratio would indicate that selecting a restaurant that would be predicted by our selection model would increase the quality of the experience (i.e., a positive selection bias) whereas a negative Mills ratio indicates that dining in a restaurant that is expected to be selected reduces the expected quality of the dining experience. All models in Table 3 that are estimated using the TripAdvisor samples exhibit negative and significant

Mills ratio. This suggests that there is, a priori, a negative selection bias. However, this negative bias becomes non-significant in the TheFork-Barcelona sample. Interestingly, it is important to note that in no case there is evidence of a positive selection bias. That is, there is no empirical evidence suggesting that people who select a restaurant according to our selection model are more likely to experience an excellent dining experience. This highlights the importance of not confounding selection and outcome. Because people's expectations contribute to both selection and outcome in different ways, it is crucially important not to assume that selecting a less likely item would lead to lower utility.

4.3 Additional robustness

We test the robustness of our results by estimating various alternative model specifications in the three samples we analyzed. First, we test that the probit selection model yields consistent results against an analogous specification based on a multinomial logistic (MNL) formulation. Second, we estimate alternative outcome models that excludes proximity controls. Finally, we estimate alternative selection and outcome models that include consumer-specific random effects to control for any unobserved heterogeneity across consumers. All our results are robust against all these alternative specifications.

Although a MNL model is the typical specification to model "choice" in the operations management literature (Anderson et al. 1992), we estimate a probit model in the selection stage of our regressions. We do this because the Heckman correction approach requires the error terms of the first and second stage to be jointly normal, which is assured with a probit selection model. Having said this, it was important to confirm that the probit selection models shown in Table 2 were fully consistent with equivalent conditional logit models that estimate the logit of selecting a possible next restaurant conditional to having visited the last restaurant; the conditional logit formulation we estimate is equivalent to the multinomial logit model, see McFadden (1974). Note that because these models are conditioned on the last restaurant visited we cannot include any variable that does not show within-group variation. That is, any variable that is constant with respect to the last restaurant, such as *user_comments_last*, *user_rating_last*, and *restaurant_rating_last*, is automatically dropped from the estimation because it would be constant to all the possible restaurant choices conditional to having visited the last restaurant. Because of this, the variation effect of *delta_quality* in these models is driven entirely by the variation in *restaurant_rating_next*. Overall, these conditional logit models, shown in Table 4, are fully consistent with the probit models shown in Table 2.

Table 5 mirrors the models shown in Table 3 but all of them exclude the "proximity" controls (to the last restaurant visited). These models examine the robustness of our main results to the consideration of contextual proximity variables to be part of the exclusion restriction. That is, since geographic, price, and restaurant features proximity seem to significantly influence selection but not outcome, we test whether our results are robust to the exclusion of such variables from the outcome model. As expected, the results are fully consistent across all three samples.

Table 4: Conditional Logit Models predicting the probability of selecting a focal restaurant

	(14) TripAdvisor Barcelona	(15) TripAdvisor Barcelona	(16) TripAdvisor Barcelona	(17) TripAdvisor Barcelona	(18) TheFork Barcelona	(19) TripAdvisor Paris
log_restaurant_	0.403***	0.446***	0.443***	0.435***	0.511***	0.318***
comments_next	(-0.00791)	(-0.00815)	(-0.00804)	(-0.00812)	(-0.00552)	(-0.0107)
same_zip	0.513***	0.440***	0.445***	0.440***	0.593***	0.579***
	(-0.0277)	(-0.0276)	(-0.0277)	(-0.0277)	(-0.0257)	(-0.0306)
log_distance	-0.838***	-1.046***	-1.042***	-1.032***	-0.939***	-0.887***
	(-0.0241)	(-0.0239)	(-0.0241)	(-0.024)	(-0.0197)	(-0.0231)
same_price	0.288***	0.390***	0.416***	0.405***	0.468***	0.436***
	(-0.0215)	(-0.0244)	(-0.0242)	(-0.0242)	(-0.0231)	(-0.0258)
common_features	0.185***	0.224***	0.212***	0.210***	0.223***	0.237***
	(-0.00988)	(-0.0108)	(-0.0109)	(-0.0108)	(-0.00864)	(-0.0186)
log_distance_		-0.267***	-0.269***	-0.266***	-0.151***	-0.217***
competition		(-0.00884)	(-0.0089)	(-0.00891)	(-0.00793)	(-0.0112)
log_same_price_		-0.0461***	-0.0469***	-0.0480***	-0.0855***	-0.0730***
competition		(-0.00607)	(-0.00615)	(-0.00615)	(-0.00458)	(-0.00614)
log_comon_features_		-0.0380***	-0.0316***	-0.0317***	-0.0197***	-0.0259***
competition		(-0.0038)	(-0.00382)	(-0.00382)	(-0.00325)	(-0.00368)
delta_quality			0.309***	-0.121**	-0.0985**	-0.254***
			(-0.0184)	(-0.037)	(-0.0325)	(-0.0393)
negative_delta_quality				0.680***	0.956***	0.727***
				(-0.0511)	(-0.0499)	(-0.0569)
AIC	212,815	211,834	211,534	211,356	222,054	179,849
BIC	212,890	211,955	211,670	211,506	222,197	180,001
Observations	24,900,073	24,900,073	24,900,073	24,900,073	11,783,117	28,136,684

Standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Although our outcome models control for consumer-specific attributes such as the number of online reviews and rating of the last review provided by the focal consumer, there might be some unobservable attributes of the consumer which might influence their tendency to give high (or low) ratings to restaurants. To account for these, we estimate consumer-specific random effect probit models in both selection and outcome stages of our Heckman estimation procedure. We do not estimate consumer-specific fixed effects models for two reasons. First, including fixed effects in probit models can lead to incidental parameter problems in particular in samples with very low number of observations per consumer (Wooldridge 2010). Second, we do have, on average, very few observations for a given user in all our samples. Table 6 show two sets of models with consumer-

Table 5: Models after excluding proximity controls in the outcome stage

	(20) TripAdvisor Barcelona	(21) TripAdvisor Barcelona	(22) TripAdvisor Barcelona	(23) TripAdvisor Barcelona	(24) TheFork Barcelona	(25) TripAdvisor Paris
user_comments_last	-0.0419***	-0.0466***	-0.0453***	-0.0446***	-0.00449***	-0.0206***
	(-0.00462)	(-0.00477)	(-0.00477)	(-0.00477)	(-0.000815)	(-0.00433)
user_rating_last	0.409***	0.392***	0.449***	0.453***	0.506***	0.477***
	(-0.0115)	(-0.012)	(-0.013)	(-0.013)	(-0.0112)	(-0.0157)
log_restaurant_	-0.0668***	-0.0388***	-0.0372***	-0.0380***	-0.0218*	-0.00726
comments_next	(-0.00913)	(-0.0095)	(-0.00952)	(-0.00953)	(-0.00889)	(-0.0106)
log_timegap_days	0.0134*	0.00938	0.0124*	0.0133*	0.0119	-0.00519
	(-0.00557)	(-0.00576)	(-0.00578)	(-0.00579)	(-0.00626)	(-0.00669)
restaurant_rating_next		0.931***	0.702***	0.779***	0.521***	0.764***
		(-0.0272)	(-0.0333)	(-0.0358)	(-0.0306)	(-0.0389)
delta_quality			0.279***	0.404***	0.316***	0.371***
			(-0.0237)	(-0.0319)	(-0.0321)	(-0.0392)
negative_delta_quality				-0.375***	-0.219***	-0.268***
				(-0.0626)	(-0.0609)	(-0.0715)
Mills ratio	-0.938***	-0.559***	-0.551***	-0.647***	-0.325***	-0.467***
	(-0.0627)	(-0.0656)	(-0.0657)	(-0.0677)	(-0.0444)	(-0.0656)
Constant	1.861***	-3.296***	-2.651***	-2.733***	-8.071***	-3.557***
	(-0.249)	(-0.296)	(-0.302)	(-0.302)	(-0.311)	(-0.291)
AIC	18,763	17,445	17,308	17,274	22,311	13,680
BIC	18,809	17,499	17,369	17,343	22,382	13,747
Observations	14,900	14,900	14,900	14,900	19,063	12,003

Standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

Table 6: Outcome models with consumer-specific random effects in the selection stage only (26-28) and in both stages (29-31)

	(26) TripAdvisor Barcelona	(27) TheFork Barcelona	(28) TripAdvisor Paris	(29) TripAdvisor Barcelona	(30) TheFork Barcelona	(31) TripAdvisor Paris
user_comments_last	-0.0455***	-0.00445***	-0.0209***	-0.0484***	-0.00365***	-0.0221***
	(-0.00482)	(-0.000827)	(-0.00436)	(-0.00546)	(-0.00106)	(-0.00499)
user_rating_last	0.454***	0.505***	0.477***	0.483***	0.538***	0.521***
	(-0.0131)	(-0.0113)	(-0.0157)	(-0.0164)	(-0.0136)	(-0.0217)
log_restaurant_	-0.0550***	-0.011	-0.013	-0.0589***	-0.00898	-0.0149
comments_next	(-0.0164)	(-0.0214)	(-0.0131)	(-0.0178)	(-0.0239)	(-0.0145)
log_timegap_days	0.0121*	0.0116	-0.00564	0.0127*	0.0146*	-0.00634
	(-0.00584)	(-0.00627)	(-0.00671)	(-0.00633)	(-0.00701)	(-0.00742)
same_zip	-0.0171	0.109	-0.074	-0.0197	0.122	-0.0839
	(-0.0532)	(-0.0579)	(-0.0704)	(-0.0575)	(-0.0644)	(-0.0777)
log_distance	0.0686	0.0319	0.0141	0.0796	0.0291	0.0212
	(-0.0465)	(-0.0375)	(-0.0493)	(-0.0504)	(-0.0419)	(-0.0544)
same_price	-0.0727**	-0.00252	-0.00285	-0.0801**	-0.00609	-0.00486
	(-0.025)	(-0.0211)	(-0.0282)	(-0.0271)	(-0.0236)	(-0.0312)
common_features	-0.00465	-0.000497	0.0233	-0.00594	0.00106	0.0237
	(-0.0134)	(-0.0106)	(-0.022)	(-0.0145)	(-0.0119)	(-0.0244)
restaurant_rating_next	0.766***	0.519***	0.758***	0.832***	0.591***	0.839***
	(-0.0363)	(-0.0311)	(-0.0394)	(-0.0435)	(-0.0359)	(-0.0495)
delta_quality	0.413***	0.316***	0.382***	0.439***	0.344***	0.421***
	(-0.0327)	(-0.0325)	(-0.0412)	(-0.0361)	(-0.0363)	(-0.0468)
negative_delta_quality	-0.399***	-0.207**	-0.290***	-0.426***	-0.225**	-0.325***
	(-0.0675)	(-0.069)	(-0.0786)	(-0.0733)	(-0.0767)	(-0.0873)
Mills ratio	-0.914***	-0.241	-0.616*	-0.999***	-0.242	-0.701*
	(-0.213)	(-0.158)	(-0.25)	(-0.232)	(-0.176)	(-0.277)
Constant	-1.695*	-8.409***	-2.992***	-1.783*	-9.376***	-3.219**
	(-0.815)	(-0.595)	(-0.908)	(-0.882)	(-0.674)	(-1.005)
log_sigma2u				-1.764***	-1.419***	-1.516***
(log random effect variance)				(-0.289)	(-0.127)	(-0.309)
AIC	17,273	22,313	13,684	17,258	22,197	13,670
BIC	17,372	22,415	13,780	17,364	22,307	13,773
Observations	14,900	19,063	12,003	14,900	19,063	12,003

Standard errors in parentheses. $*p < 0.05, **p < 0.01, ***p < 0.001$.

specific random effects. Models 26-28 show the results of including random effects in the selection stage only; Models 29-31 show the results after including random effects in both stages. Results are fully robust to the inclusion of such consumer-specific random effects.

5 Discussion

Our empirical study indicates that restaurant sequential choices are subject to reference effects, both in the selection of future restaurants to visit, and in the satisfaction experienced by the consumer in such a visit. Specifically, we find that consumers are loss-averse in selection, i.e., they disproportionally avoid choosing restaurants that are evaluated lower than their previous choice (which serves as a proxy of the consumer's quality reference). At the same time, outcomes associated with such relatively "lower-quality" restaurants are relatively more satisfactory, in the sense that the likelihood of an excellent dining experience is not as sensitive to restaurant quality as for higher-evaluation restaurants. This effect is consistent with the fact that consumers reduce quality expectations for restaurants that are worse than their previous dining experiences. We call this the *value of low expectations*.

While our results provide evidence of quality loss aversion when choosing hedonic goods, the fact that consumer satisfaction does not exhibit loss aversion has profound implications for the design of recommendation systems. Indeed, if satisfaction was subject to loss aversion, a consumer would need to avoid trajectories (i.e., sequence of hedonic choices) that exhibit a drop in quality because they would involve a disproportionate drop in the overall satisfaction of the consumption path. Towards that end, it would be optimal to visit the restaurants in the order of increasing rating, leaving the best one for the end, as suggested by existing literature (e.g., Das Gupta et al. 2015). In contrast, when consumer satisfaction is not affected by loss aversion and a visit to a lower-quality restaurant does not imply a steep satisfaction penalty, the structure of exploration is fundamentally affected. According to our results, we see that the satisfaction derived from a low-quality restaurant is relatively insensitive to the rating of that restaurant, while the satisfaction from visiting a high-quality restaurant strongly increases with its rating. Within this regime, it becomes optimal to choose a "zigzagging" trajectory where a consumer alternates between a high-quality restaurant and a low-quality one: this way, the consumer receives a high satisfaction jump every time the better restaurant is visited, and received no penalty when the worse restaurant is chosen. This is a behavior similar to Aflaki and Popescu (2014) with the difference that, in our case, it is initiated by the consumer, while in Aflaki and Popescu (2014) service level fluctuation is a decision made by the firm.

To illustrate this effect through an example, let us consider a consumer who chooses to visit three restaurants during a trip, after having visited a given restaurant (which we call restaurant 0) with rating 3.9 out of 5.0 in TripAdvisor (or equivalently 7.8 out of 10.0 in Fork). The three restaurants have average values in terms of relative location, dining features, price, and competition; they only differ in their quality rating: restaurant A has rating of 4.5 (out of 5.0), while restaurant B's rating is 4.1 (out of 5.0) and restaurant C's restaurant is 3.3 (out of 5.0). What is the optimal sequence in which the restaurants should be visited after having visited restaurant 0 in the first place? We estimate the sequence - even if unlikely according to our choice model - that would give the consumer the highest expected satisfaction path. For simplicity, and considering that most of the literature considers total path satisfaction as additive over experience epochs (e.g., Das Gupta et al. 2015), we consider here the total expected number of excellent dining experiences, that is, the sum of probabilities of having an excellent experience over the sequence of three choices. Based on the full models shown in Tables 2 and 3, Table 7 reports the distribution of the expected number of excellent dining experiences, for different visit sequences for the three samples analyzed.

Table 7: Total expected number of excellent dining experiences per sequence of restaurants to visit

Sequence	C→B→A	C→A→B	B→C→A	B→A→C	A→C→B	A→B→C
TripAdvisor-Barcelona	1.2514	1.2308	**1.2705**	1.1459	1.2672	1.1405
Fork-Barcelona	0.6839	0.7033	**0.7592**	0.5616	0.7226	0.5545
TripAdvisor-Paris	1.1597	1.1373	**1.179**	1.0578	1.168	1.0499

As we can see, the sequence that maximizes expected satisfaction (measured as the expected number of excellent experiences) is B $\rightarrow$ C $\rightarrow$ A, in all of our three samples. Hence, when there is no loss aversion when estimating the satisfaction of the consumer with a given restaurant, taking into account the possible reference effect of the current visit on future restaurant visits may lead to zigzagging choices. Note that the second-best sequence in all three sample is A $\rightarrow$ C $\rightarrow$ B, which also exhibits zigzagging. This suggests that recommending trajectories of monotonically increasing quality as suggested by the literature might result in suboptimal prescriptions and recommendation systems should refrain from such simple heuristics that are largely based on consumer choices while ignoring the consumer satisfaction obtained from such choices. In contrast, our results highlight the importance of identifying the possible quality reference effects that affect choice likelihood and outcome in order to suggest recommendations that ultimately lead higher consumer satisfaction.

Finally, our paper would not be complete without mentioning limitations and future research opportunities. We believe that our analysis, and specifically the separation of choices and outcomes, can be applied to settings where consumers consume services repeatedly, like movies, music, performing arts (Tereyağoğlu et al. 2017), museums (Martínez-de Albéniz and Valdivia 2019), sports (Veeraraghavan and Vaidyanathan 2012), news (Besbes et al. 2015), or retail in general (Caro et al. 2020). In these contexts, if consumer-level longitudinal data is available, e.g., from loyalty cards or annual ticket passes, together with quality perceptions, our model can be directly applied, and expect to find similar insights where choices tend to avoid losses, while satisfaction is relatively less affected by them due to expectation adjustments. Furthermore, our results also pave the way for more theoretical work around the dynamics of expectation adjustments. It would be interesting to track consumers with finer time granularity, so that expectation can be explicitly measured and used as a covariate in our models.

References

Adomavicius, G., and A. Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering* 17 (6): 734–749.

Aflaki, S., and I. Popescu. 2014. Managing retention in service relationships. *Management Science* 60 (2): 415–433.

Allon, G., and D. Zhang. 2017. Managing service systems in the presence of social networks. SSRN.

Anderson, E. W., and M. W. Sullivan. 1993. The antecedents and consequences of customer satisfaction for firms. *Marketing science* 12 (2): 125–143.

Anderson, S. P., A. De Palma, and J.-F. Thisse. 1992. *Discrete choice theory of product differentiation.* MIT Press.

Avery, C., P. Resnick, and R. Zeckhauser. 1999. The market for evaluations. *American economic review* 89 (3): 564–584.

Baucells, M., and R. Sarin. 2012. *Engineering happiness: A new approach for building a joyful life.* University of California Pr.

Besbes, O., Y. Gur, and A. Zeevi. 2015. Optimization in online content recommendation services: Beyond click-through rates. *Manufacturing & Service Operations Management* 18 (1): 15–33.

Caro, F., A. G. Kök, and V. Martínez-de Albéniz. 2020. The future of retail operations. *Manufacturing & Service Operations Management* 22 (1): 47–58.

Das Gupta, A., U. S. Karmarkar, and G. Roels. 2015. The design of experiential services with acclimation and memory decay: Optimal sequence and duration. *Management Science* 62 (5): 1278–1296.

DeCroix, G., X. Long, and J. Tong. 2019. How Service Quality Variability Hurts Revenue When Customers Learn: Implications for Dynamic Personalized Pricing. Available at SSRN 3082370.

Dellarocas, C. 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science* 49 (10): 1407–1424.

Dellarocas, C. 2006. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management science* 52 (10): 1577–1593.

Festinger, L. 1954. A theory of social comparison processes. *Human relations* 7 (2): 117–140.

Ghose, A., and P. G. Ipeirotis. 2006. Designing ranking systems for consumer reviews: The impact of review subjectivity on product sales and review quality. In *Proceedings of the 16th annual workshop on information technology and systems*, 303–310.

Heckman, J. J. 1979. Sample selection as a specification error. *Econometrica* 47:153–161.

Heese, H. S., and V. Martínez-de Albéniz. 2018. Effects of assortment breadth announcements on manufacturer competition. *Manufacturing & Service Operations Management* 20 (2): 302–316.

Hirschman, E. C., and M. B. Holbrook. 1982. Hedonic consumption: emerging concepts, methods and propositions. *Journal of marketing* 46 (3): 92–101.

Holbrook, M. B., and E. C. Hirschman. 1982. The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of consumer research* 9 (2): 132–140.

Hollenbeck, B., S. Moorthy, and D. Proserpio. 2019. Advertising strategy in the presence of reviews: An empirical analysis. *Marketing Science* 38 (5): 793–811.

Kahneman, D. 2011. *Thinking, fast and slow.* Macmillan.

Kalyanaram, G., and R. S. Winer. 1995. Empirical generalizations from reference price research. *Marketing science* 14 (3_supplement): G161–G169.

L'Ecuyer, P., P. Maillé, N. s. E. Stier-Moses, and B. Tuffin. 2017. Revenue-maximizing rankings for online platforms with quality-sensitive consumers. *Operations Research* 65 (2): 408–423.

Love Exploring 2019, 22 July. Ranked: the 50 best food cities of 2019. https://www.loveexploring.com/gallerylist/69944/the-50-best-food-cities.

Magno, F., F. Cassia, and A. Bruni. 2018. "Please write a (great) online review for my hotel!" Guests reactions to solicited reviews. *Journal of Vacation Marketing* 24 (2): 148–158.

Maister, D. H. 1984. *The psychology of waiting lines.* Harvard Business School Boston, MA.

Martínez-de Albéniz, V., and A. Valdivia. 2019. Measuring and exploiting the impact of exhibition scheduling on museum attendance. *Manufacturing & Service Operations Management* 21 (4): 761–779.

Mazumdar, T., S. P. Raj, and I. Sinha. 2005. Reference price research: Review and propositions. *Journal of marketing* 69 (4): 84–102.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, ed. P. Zarembka. Academic Press, New York.

Nasiry, J., and I. Popescu. 2011. Dynamic pricing with loss-averse consumers and peak-end anchoring. *Operations research* 59 (6): 1361–1368.

Oliver, R. L. 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research* 17 (4): 460–469.

Parasuraman, A., V. A. Zeithaml, and L. L. Berry. 1988. Servqual: A multiple-item scale for measuring consumer perc. *Journal of retailing* 64 (1): 12.

Popescu, I., and Y. Wu. 2007. Dynamic pricing strategies with reference effects. *Operations Research* 55 (3): 413–429.

Ricci, F., L. Rokach, and B. Shapira. 2011. Introduction to recommender systems handbook. In *Recommender systems handbook*, 1–35. Springer.

Rust, R. T., and T. S. Chung. 2006. Marketing models of service and relationships. *Marketing science* 25 (6): 560–580.

Rust, R. T., J. J. Inman, J. Jia, and A. Zahorik. 1999. What you don't know about customer-perceived quality: The role of customer expectation distributions. *Marketing Science* 18 (1): 77–92.

Tereyağoğlu, N., P. S. Fader, and S. Veeraraghavan. 2017. Multiattribute Loss Aversion and Reference Dependence: Evidence from the Performing Arts Industry. *Management Science* Forthcoming:NA.

Tse, D. K., and P. C. Wilton. 1988. Models of consumer satisfaction formation: An extension. *Journal of marketing research* 25 (2): 204–212.

Tversky, A., and D. Kahneman. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2): 263–291.

Tversky, A., and D. Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics* 106 (4): 1039–1061.

Tversky, A., and D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty* 5 (4): 297–323.

Van de Ven, W. P., and B. M. Van Praag. 1981. The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of econometrics* 17 (2): 229–252.

Veeraraghavan, S., and R. Vaidyanathan. 2012. Measuring seat value in stadiums and theaters. *Production and Operations Management* 21 (1): 49–68.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data.* MIT press.