ORIGINAL PAPER

Empirical studies on the impact of booking status on customers' choice behavior in online appointment systems

Isabel Kaluza[1] · **Guido Voigt**[1] · **Friederike Paetz**[2]

Abstract

We consider customers' choice behavior in online appointment systems. In three online experiments, we investigate whether and to what extent customers are impacted by the number of available slots by asking subjects to choose between two providers of several service areas with different occupancy rates. In line with previous literature, we find some evidence that customers infer quality from a utilized system compared to an empty schedule; that is, any demand (very few booked appointments) is preferred to no demand (empty booking schedule). A too-small offer set, in contrast, shows congestion and leads to an opposed scarcity effect because customers expect the provider to be in a rush, the waiting room to be crowded, and further value flexibility in the choice of the appointment time. We contribute to the literature by analyzing the interaction of the two nonlinear effects and further find that the presence of the quality-inference effect depends on the type of service. For a standardized service, we find no evidence of the quality-inference effect. For non-standardized services, however, we find an inverse U-shaped preference in the number of offered slots, showing that customers prefer a medium utilization of the service provider. We find that this is a robust representation of customers' preferences, even if other quality signals, such as star ratings and prices, are available.

Keywords Behavioral operations management · Service operations · Customer choices · Observational learning · Discrete choice experiments

JEL Classification D91

✉ Isabel Kaluza
isabel.kaluza@uni-hamburg.de

1   Faculty of Business Administration, Institute of Logistics and Supply Chain Management, University of Hamburg, Hamburg, Germany

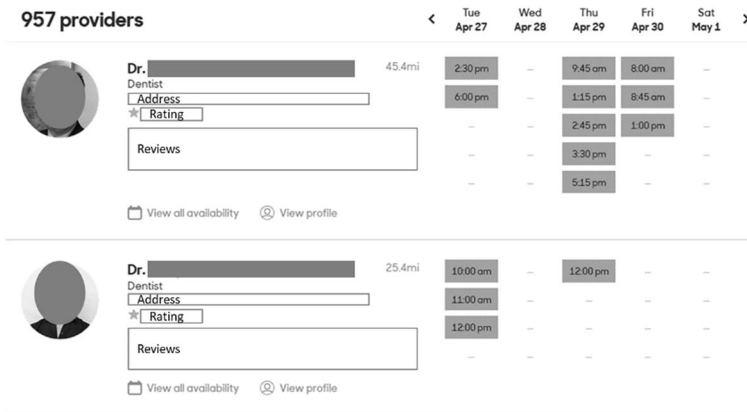2   Clausthal Executive School, Clausthal University of Technology, Clausthal-Zellerfeld, Germany

Fig. 1 Exemplary offer sets of two dentists on zocdoc.com

1 Introduction

The digitalization trend leads an increasing number of service providers to offer their services online via appointment systems, such as dentists on www.zocdoc.com and hairdressers on www.booksy.com. Figure 1 exemplarily shows the offer sets of two (anonymized) dentists. If a service provider aims to actively control the amount and timing of slot offers (see Kaluza et al. 2021), it appears crucial to understand how customers' willingness to book a service depends on varying offer sets. To this end, we show that the number of appointment slot offers (e.g., ten/four slots offered by the first/second provider in Fig. 1) impacts the likelihood that the service provider is chosen. We show that a very small offer set (indicating a high utilization of the service providers) makes the provider less attractive because consumers anticipate a disutility in higher-utilized systems (e.g., longer waiting times). We call this the "scarcity effect". In turn, very few offers (indicating a low utilization of the service provider) make the provider less attractive because consumers infer a low quality of service based on the lack of demand. We call this the "quality-inference effect". Considering these two effects altogether, we show that consumers have an inverse U-shaped preference curve for an increasing number of offered slots.

Customers looking for such services search online, gather information about the providers, and, at best, book an appointment. Some relevant aspects, such as sex, age, and address of the provider, can be observed directly. However, if a service provider is unknown, the provider's quality is uncertain. In a search process, the offer sets of providers may provide information to the uninformed customer about whether other customers have previously booked the services. Customers may infer quality from smaller offer sets when they presume that the other customers who previously booked the services know about the providers' quality. Such a learning process may generate positive externalities (e.g., Kremer and Debo 2016) that lead to an adaptation of the choice behavior of the other customers by also booking an appointment with a popular provider (i.e., choosing the provider with a well-booked

schedule). At the same time, it may also impose a negative externality (e.g., Jin et al. 2015), the scarcity effect. This effect leads to an opposed choice behavior if the provider is, for instance, (almost) fully booked and the appointment could be accompanied by a waiting time that is too long or customers do not have enough flexibility regarding the time of the appointment.

We analyze whether and to what extent customers infer quality or scarcity from the number of offered slots and how it impacts the interest in a service provider. Both effects, quality and scarcity, belong to the theory of observational learning (Banerjee 1992; Bikhchandani et al. 1992). The quality-inference effect is similar to the empty restaurant syndrome in the offline world, which has been examined in several studies (e.g., Becker 1991; Teraji 2003; Kremer and Debo 2016). Kaluza et al. (2021) show in an online experiment that customers infer quality from a moderately sized offer set compared to an empty schedule when booking an appointment with a dentist online. In the same set of experiments, it was also shown that a scarce offer set, indicating a high utilization of the dentist, evokes the scarcity effect, making the dentist comparably less attractive.

We first replicate those findings in online experiments by comparing three different occupancy rates (scarce, moderate, full offer set) in the same health care context. We then proceed by comparing a wider range of occupancy rates in the health care context to gather more insights into customers' choice behavior. We find non-linear effects on the interest in the provider. An increase in interest can already be observed with a partly booked schedule compared to an empty schedule. With more booked appointments, the interest increases degressively up to a turning point and then decreases again progressively if even fewer appointments are offered. Furthermore, we differentiate between standardized and non-standardized services while in both service areas, the customer can choose between several providers. In our context, standardized services offer services with an output that can be evaluated easily and objectively without much experience and know-how. The process of the service is often standardized and the quality of the output of each service provider is equivalent from an objective point of view. Examples may be a new identity card, a new credit card or the signing of certified copies. Output of non-standardized services, in contrast, can differ in the quality of the output. An investment consulting from two different providers may lead to two outputs of different quality. Other examples may be shooting for application photos or a massage. In our analysis, we consider different service areas, namely, a hairdresser with a non-standardized service and a government entity with a standardized service, to see whether the effects in the health sector are transferable to other contexts. We find that the type of service is decisive for whether the quality-inference effect is present. For the considered standardized service, we find no evidence of the quality effect. In contrast, customers prefer flexibility (scarcity effect) and do not consider the quality of the provider.

We then consider in two discrete choice experiments the relative importance of the booking status when other information (i.e., star rating, travel-distance, and prices) is available. Not surprisingly, we observe that those attributes also convey quality information; however, booking status still has a nonnegligible impact on the preference for a supplier. The inverse U-shaped preference for an increasing number of offered slots appears robust.

Our main contribution lies in the analysis of the interaction of the two nonlinear effects and how they may impact customers' choice behavior in online appointment systems when observing booking status. The managerial insights are primarily relevant for service providers with (many) first-time customers. We further find that customers do not infer quality from booking status when they search for a standardized service. Providers of standardized services should focus on enough appointment times for customers to choose from. Service providers with nonstandardized services should account for both effects, the quality-inference effect by offering fewer appointments than available if many appointments are left unbooked, and the scarcity effect by not offering too few appointments if possible. This way, the providers convey popularity without seemingly being rushed while providing customers flexibility in their choice.

The rest of this paper is structured as follows. In Sect. 2, we review the relevant literature regarding the empty-restaurant syndrome and the scarcity effect, while Sect. 3 outlines our problem description. In Sect. 4, we describe the design and hypotheses of our Online Experiment I on occupancy rates and present our results. Section 5 presents the design and hypothesis as well as the results of our Online Experiment II on different service areas. We summarize and critically discuss our approach and its limitations in Sect. 6 and conclude our paper in Sect. 7.

2 Related literature

We consider online appointment systems, similar to the approaches of, e.g., Gerchak et al. (1996), Gupta and Wang (2008), Feldman et al. (2014), Liu et al. (2019), and Kaluza et al. (2021), and investigate whether and to what extent the number of offered slots impacts customers' interest in unknown service providers. Associating quality with a smaller offer set may lead to an adaptation of previous customers' choice behavior (herd behavior, Banerjee 1992), i.e., uninformed customers are more likely to choose a provider if they observe that other customers booked the service before. In the case of an empty schedule, customers may rather choose a competitor (as other customers did) who is more in demand. An almost fully utilized provider, instead, harms this observational learning process and loses attractiveness due to congestion. Offering too few appointments may again lead to less demand. Both effects belong to the theory of observational learning (see, e.g., Banerjee 1992; Bikhchandani et al. 1992). The presence of the effects on online appointment systems was initially detected in an online experiment by Kaluza et al. (2021) but has also been investigated in other contexts (e.g., offline restaurants). The effect of inferring quality from observed demand is better known as the empty-restaurant syndrome in the offline world. According to herd behavior, uninformed customers expect informed customers in the queue to know about the provider's quality and may, thus, also line up in that queue (Kremer and Debo 2016). The opposite behavior could also be observed, e.g., in the case of congestion (e.g., Eyster et al. 2014; Jin et al. 2015), in which case customers choose a competitor if the considered booking system is (almost) fully booked.

Several experiments investigate the empty-restaurant effect in the offline world. Koo and Fishbach (2010) find in an experiment that the queue behind a person positively impacts the value of a product. Furthermore, waiting time can positively affect purchase intention (Giebelhausen et al. 2011) and increase perceived quality (Kremer and Debo 2016). However, the congestion level (too-long queue) may bias the learning process, tending toward a smaller queue (Jin et al. 2015). Kaluza et al. (2021) set the queue in the offline world equivalent to a (partly) utilized appointment schedule in the online world and the lack of demand (e.g., an empty restaurant) equivalent to an empty schedule. They compare three offer sets (scarce, moderate, and full offer set) for a service of a dentist in an online experiment and find that demand for a dentist increases with a smaller offer set, as long as the number of offered slots is not too scarce. We investigate customers' choice behavior regarding higher occupancy rates and different services. Liu et al. (2019) detect customers' (heterogeneous) preferences for specific time windows for an appointment with a doctor but do not consider the number of offered slots. We control these heterogeneous preferences by stating a predefined preference for a specific time slot but consider the impact of the quality-inference effect on the choice behavior.

The effect of a queue on the interest in the provider or the purchase intention has been included and investigated in several approaches (e.g., Hassin and Haviv 1997; Veeraraghavan and Debo 2009, 2011; Debo et al. 2012). In our online experiments and in line with, e.g., Jin et al. (2015), subjects observe the overall offer set. This setting is different from other approaches in which customers only observe the choice of the previous customer (e.g., Çelen and Kariv 2004), a subset of previous choices (e.g., Acemoglu et al. 2011; Smith and Sorensen 2013) or have full information (e.g., Banerjee 1992; Bikhchandani et al. 1992). We adapt the simple utility model from Kaluza et al. (2021) that solely captures the interest in the service provider with respect to the quality inferencequality-inference effect in a multinomial logit model (McFadden 1974; Ben-Akiva and Lerman 1985; Train 2009) and expand it to additionally consider the scarcity effect. Being overwhelmed by too many choice alternatives or too much information, e.g., choice overload or information overload, could also decrease customers' interest (e.g., Mushtaque and Pazour 2020). Choice overload can result from a large assortment of options, e.g., too many offered appointments (Chernev et al. 2015; Iyengar and Lepper 2000; Eppler and Mengis 2004) and is part of the construct of information overload, e.g., receiving too much information. It would be another explanation for decreased interest if too many appointments were offered, but this was not observed by Kaluza et al. (2021).

Overall, our study is closest to Kaluza et al. (2021), who present the first empirical evidence on the impact of booking status on consumer choice behavior. The other abovementioned papers mainly look at herding behavior in financial settings or settings with physical queues. In contrast to Kaluza et al. (2021), in this study, we present evidence for an inverse U-shaped preference curve in the number of free slots, i.e., consumers prefer a medium occupancy rate of a service provider. We further show that the quality effect is negligible in standardized service settings (e.g., government), and we provide evidence that the quality effect is present even if other quality signals, such as star ratings or price signals, are available.

3 Problem formulation

We consider online appointment systems that are used by service providers to offer appointments to customers. We hereby build on the approach introduced by Kaluza et al. (2021). Each service provider has an overall capacity of κ appointments per workday. Customers visit the websites of the booking systems and observe the offered time slots, with $\mathcal{O}$ being the number of offered time slots. Depending on several aspects, e.g., reviews and the observed offered appointments, customers may choose one of the given appointments or another provider. We investigate what impact the number of offered appointments has on the choice of customers.

According to the theory of observational learning (Banerjee 1992; Bikhchandani et al. 1992), uninformed customers learn from other (presumed informed) customers. They may either adopt the behavior of the other customers (positive effect) or oppose their behavior (negative effect). In our Online Experiments I and II, customers receive no additional information about the provider's service quality, e.g., via ratings. They may only infer the provider's quality from the observed booking schedule (quality-inference effect), which increases their interest in the provider. They are, in turn, negatively impacted if the offer set is scarce. Note that we introduce additional factors, such as star rating and travel time, in our robustness check in Sect. 6.

We model the customers' choice behavior as a multinomial logit model, adjusting the formulation introduced by Kaluza et al. (2021) to account for both the quality and the scarcity effect and for nonlinear courses of the effect curves. In contrast, in Kaluza et al. (2021), the scarcity effect is neglected, while a linear relationship of the number of offered slots on quality perception is imposed. Equation (1) denotes the utility function with asc_j, the alternative specific constant of provider $j \in J$, and the Gumbel distributed error term ϵ_j. The strength of the quality-inference effect is denoted by β_{QIE}, and the strength of the scarcity effect by β_{SCAR}. Both effects impact customers' choice behavior relative to their overall capacity κ.

$$U_j = asc_j + \beta_{QIE} \cdot \left[\frac{\mathcal{O}^2}{\kappa} \right] + \beta_{SCAR} \cdot \left[ln\left(\frac{\mathcal{O}^2}{\kappa} \right) \right] + \epsilon_j \tag{1}$$

Figure 2 shows how the effects impact utility when increasing the number of offered appointments. When considering the quality-inference effect in isolation, decreasing the number of offered appointments leads to a degressive increase in utility, whereas increasing the number of offered appointments leads to a progressive decrease in utility. This means that if no appointment is booked and all appointments are offered, offering one appointment less leads to a higher increase in utility than when decreasing the number of offered appointments from, e.g., two to one appointment(s). An explanation for this behavior could be that any demand is better than no demand. Present demand shows a popular provider, irrespective of a well-utilized system or a system with even one more booked appointment, which harms the increase in interest. The highest utility, according to the quality-inference effect, is achieved with only one offered slot (see Fig. 2, dashed line). As customers do
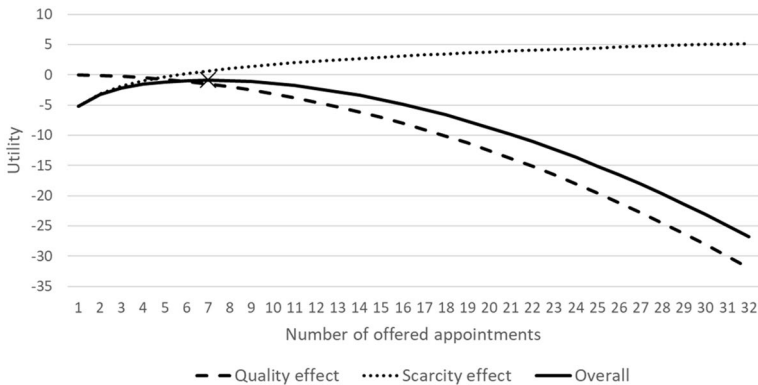
Fig. 2 Utility curves according to the quality-inference effect (dashed line), scarcity effect (dotted line) and overall utility (solid line, maximum marked with a x), with $\kappa = 32$, $\beta_{QIE} = 1$, $\beta_{SCAR} = 1.5$, and $asc_j = 0$

not have a choice when the service provider does not offer any appointment (zero offered appointments), we only consider nonempty offer sets.

The scarcity effect, considered in isolation, leads to a degressive increase in utility when increasing the number of offered appointments. A very scarce offer set (e.g., one offered appointment) may be associated with a crowded waiting room, too much waiting time, and a provider who is in a rush (Kaluza et al. 2021). Additionally, customers then do not have any flexibility in their choice of time. Offering one more appointment then has a great positive impact on the interest. A partly booked appointment system already gives flexibility in customers' choices. Offering one more appointment has a weaker effect on the interest. Thus, the greatest increase in utility is achieved when increasing the number of offered appointments from one to two and the highest utility with a full offer set (also see Fig. 2, dotted line).

Overall, Fig. 2 shows that utility increases up to a certain point when decreasing the number of offered appointments but then once again decreases when further decreasing the size of the offer set. As the relative utility is decisive for our analysis, it is irrelevant that the utility becomes negative at some point. Furthermore, an $asc_j > 0$ may lead to positive absolute utilities but solely shifts the respective curve and, thus, the overall utility, up.

4 Online experiment I: occupation rates

4.1 Online experiment design

Online Experiment I analyzes several occupancy rates to investigate the impact of the number of offered slots on customers' choice behavior in an online appointment system. The online experiment is set up with the software LimeSurvey and

distributed via Amazon Mechanical Turk (MTurk, www.mturk.com)[1] to exam-
ine seven different occupancy rates. A total of 1200 subjects completed the online
experiment in October 2020 (October 16 until October 30). Each subject was paid 1
USD for completing the questionnaire as an incentive.[2]

Online Experiment I is divided into three parts in line with the online experi-
ment presented in Kaluza et al. (2021). The first part comprises warm-up questions[3].
In the second (main) part, we ask subjects to imagine different scenarios in which
they choose between two service providers (providers A and B). In this context, the
subjects need a routine (i.e., nonurgent) appointment with the dentist and prefer an
appointment at 8 a.m. An exemplary scenario is given in Fig. 7 in Appendix A1. The
providers have opening hours from 8 a.m. to 4 p.m. with appointments in 15-min
cycles (this information is given to the subjects). We consider highly occupied sys-
tems (one, two, and four appointments offered), moderately occupied systems (eight,
16, and 24 offered appointments) and an empty schedule (32 appointments offered).
A smaller offer set is always a subset of the larger offer sets, and each provider offers
the preferred 8 a.m. appointment. The subjects are asked to choose between two
offer sets. Our only experimental factor is the size of the offer set with levels of 1,
2, 4, 8, 16, 24, and 32 offered appointments. The third part of the online experiment
comprises a question regarding the motives for the choices during the previous part
and several attention checks to see whether (1) the subjects read the tasks carefully
and (2) the subjects understand the given appointment system. An incorrect answer
to an attention check does not lead to aborted participation. Apart from the occu-
pancy rates and minor adjustments[4]due to the different subject pools, the question-
naire remains unchanged from Kaluza et al. (2021).

We consider eight treatments in a between-subject design; see Table 1. Three
scenarios are assigned to each treatment. Treatment T1, for example, first chooses
between service provider A with two offered appointments and service provider
B with eight offered appointments. The second choice in treatment T1 is between
four and 32 appointments, and the last choice is between one and 24 appoint-
ments. As no order effects are found in Kaluza et al. (2021), we fix the position
of the provider with the smaller (bigger) offer set to the left-hand side (right-hand

---

[1] MTurk is a controversial but valid platform for the distribution of behavioral surveys, see Lee et al.
2018.

[2] Initially, an MTurk master qualification was needed, which can be reached when successfully complet-
ing > 1000 surveys (not to be confused with the master's degree from the university). As only 267 mas-
ters completed the survey within two weeks, the requirement was then softened to > 50 successfully com-
pleted surveys (approved by the requesters) and an overall approval rate of > 95%. With a required MTurk
master qualification, a worker earns 5% extra, which was paid additionally to the first 267 subjects. Con-
tingency-analysis reveals that there are hardly any differences between the demographics of participants
after relaxing the criteria. There is only one significant effect for age Group 20-29, which was under-
represented with unrelaxed criteria (10.1%) and increased with relaxed criteria in the final sample (25%).

[3] The warm-up questions are as follows: "How frequently do you use the internet?"; "Have you already
booked an appointment or bought a ticket online? If so, what did you book or buy?"; "Which channel do
you prefer to use when booking an appointment?".

[4] E.g., net income of the household in USD instead of EUR.

Table 1 Treatment summary of Online Experiment I with offer sets of the two service providers (A vs. B)

Treatment	Within-subject		
	First choice	Second choice	Third choice
T1	2 vs. 8	4 vs. 32	1 vs. 24
T2	2 vs. 32	8 vs. 16	4 vs. 24
T3	8 vs. 32	2 vs. 4	16 vs. 24
T4	4 vs. 32	8 vs. 24	2 vs. 16
T5	16 vs. 32	4 vs. 8	2 vs. 24
T6	8 vs. 24	16 vs. 32	1 vs. 4
T7	2 vs. 4	24 vs. 32	1 vs. 8
T8	24 vs. 32	4 vs. 16	1 vs. 2

side) of the page. The response options (Provider A, Provider B) of each scenario are randomly displayed. A total of 150 subjects were randomly assigned to each treatment.

4.2 Hypotheses

The preferred 8 a.m. appointment is available in each offer set. Thus, a customer who gathers no information from the booking status would be indifferent to all the choices, resulting in random choices (50/50 splits). However, in line with previous literature, we expect that customers learn from other customers' choice behavior and interpret an empty schedule as a lack of demand. Offering a subset of all appointments may signal quality, similar to an offline queue (e.g., Banerjee 1992; Kremer and Debo 2016; Kaluza et al. 2021). We thus expect that the customer's choice probability for a service provider increases if the provider offers only a subset of the available appointments instead of all appointments. For our analysis, we primarily focus on the first choice of each treatment (see Table 1), as this is the independent observation of each subject. Replicating results from prior literature, we expect customers to prefer a provider who offers eight appointments to another provider with 32 offered appointments (first choice in T3, see Hypothesis 1a). In contrast, congestion may have a negative impact on customers' choice probability (e.g., Jin et al. 2015). Congestion may imply a crowded place and a lack of flexibility from a customer's perspective regarding appointment times. We thus expect that customers' choice probability for a service provider decreases again at some point if the offer set becomes too scarce. Again, replicating results from prior literature, a provider with only two offered appointments is no longer preferred to a bigger offer set of eight or 32 appointments (first choice in T1 and T2, see Hypothesis 1b).

Hypothesis 1: General quality-inference effect and scarcity effect

(a) The choice frequency of a service provider with a moderate offer set of eight appointments is higher than that of the provider with the larger offer set of 32 appointments in treatment T3.

(b) The choice frequency of a service provider with a scarce offer set of two appointments is lower than that of the provider with the larger offer set of eight appointments in treatment T1 and 32 appointments in treatment T2, respectively.

We expect that the preference of the smaller offer set (when comparing eight with 32 appointments) does not hold for the comparison of 2 vs. 32 and 2 vs. 8 appointments, implying that the impact of the scarcity effect outweighs the impact of the quality-inference effect at some point. Below that number of offered appointments, the utility, and thus the interest, decreases with the decreasing offer set. To further analyze the turning point, we are interested in whether customers are more likely to choose a provider who offers four appointments than one with 32 offered appointments or vice versa. As four appointments do not give much flexibility to the customers regarding the appointment time, we expect the scarcity effect to still outweigh the quality-inference effect; see Hypothesis 2a. This would result in a preference for the larger offer set. If the customers prefer the smaller offer set instead (dominant quality-inference effect), it would be interesting to analyze whether it should be recommended to offer four or eight appointments when facing an empty schedule. The estimate of the effect coefficients may give more insights into the effect curves and the resulting change in the interest in a service provider when offering two, four, or eight appointments. Furthermore, we analyze how customers' choice behavior changes when offering more than eight appointments compared to another provider with an empty schedule (32 offered appointments). We expect that offering 16 appointments should also be preferred to 32 appointments, as a half-full system shows a seemingly in-demand provider while customers still have many appointment times to choose from. If no preference toward the smaller offer set is observed, this may mean that a service provider with an empty schedule could further increase interest by blocking (not offering) more than half of the appointments. The strength of the effects is then interesting for the decision support on how many slots to offer (whether customers value the flexibility with 16 appointments more than the higher quality with eight appointments or vice versa).

Hypothesis 2: Bounds and course of the effect curve

(a) The choice frequency of the provider with four offered appointments is lower than that of the provider with 32 offered appointments in T4.

(b) The choice frequency of the provider with a half-full schedule (16 offered appointments) is higher than that of the provider with an empty schedule (32 offeredavailable appointments) in T5.

Kaluza et al. (2021) explore choices between great differences in the number of offered slots. We decrease the differences between the sets to study whether smaller

changes still evoke behavior based on scarcity or the quality-inference effect. Therefore, we further analyze the customers' choice behavior when observing two partly utilized systems (two well and two hardly utilized systems). Following Hypothesis 1, we expect the customers to prefer a higher utilized service provider if both providers are partly (hardly) booked (Hypothesis 3a and Hypothesis 3c, respectively). We further expect that customers prefer a larger offer set when both service providers are almost fully booked (Hypothesis 3b). Both providers then seem to be popular, and customers may not perceive that they are giving up quality by choosing the more flexible provider.

Hypothesis 3: Small differences in occupancy rates

(a) If two service providers both have partly booked systems (i.e., eight and 24 appointments in treatment T6, respectively), the choice frequency of the smaller offer set is higher than that of the larger offer set.
(b) If two service providers both have almost fully booked systems (i.e., two and four appointments in treatment T7, respectively), the choice frequency of the smaller offer set is lower than that of the larger offer set.
(c) If two service providers both have hardly booked systems (i.e., 24 and 32 appointments in treatment T8, respectively), the choice frequency of the smaller offer set is higher than that of the larger offer set.

Kaluza et al. (2021) detect several main motives for customers' choices. We analyze whether the subjects in our online experiment also mention the scarcity and the quality-inference effect as the main drivers behind their choice behavior and summarize this assumption in Hypothesis 4.

Hypothesis 4: The quality-inference effect and the scarcity effect are the main drivers of the customer's choice of a service provider when observing the number of offered appointments.

4.3 Results

4.3.1 Choice frequencies

We first review the answers regarding the attention checks, as subjects could continue the questionnaire even with a wrong answer. We consider comprehension questions and open questions and checked for specific patterns. For example, one comprehension question failed by more than 26%. We further find data that clearly showed nonserious answers, i.e., unsuitable answers to open questions. We therefore clean up the data.[5] The cleaned-up and further analyzed dataset

---

[5] Subjects were excluded who (1) failed the instruction to give a predetermined answer, (2) failed any of the two comprehension questions about the displayed offered appointments, (3) disregarded the predetermined 8 a.m. preference ("no way I'm getting up that early to go to a dentist appointment I do not desperately need", respond. 567), (4) gave groundless statements in open-text answers (which were often identically given by several subjects) and/or (5) contra-dictionary answers that followed a specific pattern, e.g., always first answer option chosen, which deliberately led to contradictory answers.

Table 2 Descriptive statistics of the first choice, varying occupancy rates

Treatments	2 vs. 8 (T1)	2 vs. 32 (T2)	8 vs. 32 (T3)	4 vs. 32 (T4)	16 vs. 32 (T5)	8 vs. 24 (T6)	2 vs. 4 (T7)	24 vs. 32 (T8)
Choice frequency in %	47: 53	40: 60**	59: 41**	45: 55	48: 52	56: 44	43: 57*	61: 39**
(p-value)	(0.301)	(0.034)	(0.041)	(0.182)	(0.379)	(0.150)	(0.099)	(0.019)
# Subj	92	97	96	98	94	93	102	93

*Significant difference from random choice (50% each) on a 90%-level

**Significant at the 95% level

consists of 765 subjects, with 41% female and 57% male respondents (2% others). The ages range from 14 to 81 years, while 87% are between 20 and 49 years old (25% 20–29 years, 44% 30–39 years, 18% 40–49 years). The majority live in the United States of America (78%), 11% in India and 5% in Brazil (the remaining 6% come from all over the world).

Analyzing the cleared-up dataset, we focus on the independent observations (first choice) and present the choice frequency of each treatment in Table 2. The descriptive statistics of all choices are summarized in Table 10 in Appendix A2.

4.3.1.1 Basic quality-inference effect (Hypothesis 1) Comparing the choices of treatment T3 between the provider with a moderately sized offer set (eight appointments) and an empty schedule (32 appointments), 59% prefer the smaller offer set (see Table 2). If no effect was seen, we would observe a 50% choice frequency for each provider. Choice-by-chance can be rejected for T3 (binomial test, $\alpha < 0.05$), and we observe a significant preference for the smaller offer set, which supports Hypothesis 1a. In treatment T2, subjects were asked to decide between a scarce and a full offer set (2 vs. 32). Subjects preferred the bigger (full) offer set (60%). We again observe a significant difference from the 50% choice frequency for treatment T2 (binomial test, $\alpha < 0.05$), although in the opposite direction. Subjects significantly prefer the bigger offer set, which supports Hypothesis 1b. Treatment T1 (2 vs. 8) also shows a slight but nonsignificant preference (53%) for the larger offer set of eight appointments compared to the scarce offer set of two (binomial test, $\alpha > 0.1$). In summary, Hypotheses 1a and 1b are supported significantly by our results and directionally supported by the nonsignificant result.

4.3.1.2 Bounds and course of the effect curve (Hypothesis 2) Comparing an offer set containing four appointments with a full offer set (T4), 55% prefer the provider with the larger offer set. This shows a nonsignificant tendency toward the full offer set (binomial test, $\alpha > 0.1$), which directionally supports Hypothesis 2a. This further indicates that an offer set with four offered slots may still be too scarce, i.e., the scarcity effect superimposes the quality-inference effect for the majority.

For the comparison of 16 offered appointments with the empty schedule (16 vs. 32, T5), we observe an almost 50% choice frequency, which does not significantly differ from random choice (binomial test, $\alpha > 0.1$). Some subjects seem to prefer greater availability, while others interpret the smaller offer as a signal of quality. We cannot detect a dominant effect and, thus, find no support for Hypothesis 2b with treatment T5.

4.3.1.3 Small differences in occupancy rates (Hypothesis 3) When having to choose between two moderately utilized dentists (8 vs. 24, T6), we find a nonsignificant tendency of 56% toward the smaller offer set. This indicates that customers infer quality from the smaller offer set of 8 appointments more than how they value

the greater flexibility with the bigger offer set of 24 appointments and directionally supports our Hypothesis 3a (binomial test, $\alpha > 0.1$).

As expected, if the customers only observe scarce numbers of offered appointments (2 vs. 4, T7), the majority (57%) associate the larger offer set with less crowded waiting rooms and less rush. We find a weak significant tendency toward the larger offer set, which indicates the scarcity effect to be dominant and supports our Hypothesis 3b (binomial test, $\alpha < 0.1$). Customers may not perceive giving up quality by choosing the dentist who provides more appointment times to choose from, as both dentists seem to be in demand, again indicating that the scarcity effect dominates when observing such high occupancy rates.

For the comparison of two hardly utilized dentists (T8, 24 vs. 32), we find a significant preference of 61% for the dentist that has at least some capacity utilized (binomial test, $\alpha < 0.05$). Having some booked appointments seems to indicate quality by still giving the customers enough flexibility in choosing an appointment time. This finding supports Hypothesis 3c. However, the results of T8 show an interesting tendency compared to T5 (16 vs. 32, Hypothesis 2b). It is surprising, as we see no significant difference in the comparison of 16 slots with the empty schedule (T5). Thus, an open question remains of whether it may be more beneficial to offer only a few slots (eight appointments) or many (24 appointments) but not half of the system size (16 appointments) if the appointment schedule is still empty.

4.3.2 Motives for the choice behavior

We observe a preference for a smaller (but not scarce) offer set if the alternative offer set is very large. In turn, a preference for the bigger offer set is observed in case one provider offers only a few slots. We hypothesize that the quality-inference effect and the scarcity effect mainly drive this behavior; see Hypothesis 4. A content analysis gives further insights into the reasons for the subjects' response behavior. Two independent raters assessed an open answer regarding the motive for the choice behavior and coded on a binary scale, blinded to the hypotheses, whether a motive was given (yes: 1, no: 0). Cohen's kappa (Byrt et al. 1993) is given to state the interrater reliability. We detect three dominant variables (motives): "quality", "scarcity", and "flexibility".[6]

Quality: One explanation of the behavior is the quality motive, as expected. Customers associate a smaller number of offered slots with a more popular provider. Other customers seem to request the service of the respective provider, which may mean that the provider is the better dentist. "FewerLess appointments available means more popular and a better option." (respond. 1022). Furthermore, a customer may not want to visit a dentist who treats no other customers. "The dentist with the all or most of the day free seemed strange to me—either he has no patients (not good), or

---

[6] Further explanations that are sporadically mentioned, such as a misunderstood offer and single entries are aggregated to the category others and no longer considered in this analysis. Several subjects did not want to explain their choices or the answer was not clear enough to be categorized ("just reliability" respond. 605).

Table 3 Frequency of the motives for the choice behavior

	Quality	Flexibility	Scarcity	Other
Frequency in % (Cohen's kappa)	34% (0.7918)	47% (0.7013)	11% (0.3191)	24% (0.4005)

The frequency is the geometric mean of the raters' results. Multiple reasons were possible, leading to a sum of the frequencies > 100%

the system double-books people (also not good)" (respond. 236). In summary, customers infer quality from a smaller availability set.

Scarcity: Another explanation of the behavior is the scarcity motive, which again concurs with our expectations. If a service provider offers too little, potential customers expect the provider to be overbooked and/or the waiting room to be crowded. "I chose the one that didn['] t look like he was overbooked. [S]ometimes appointments run longer than expected." (respond. 311). Furthermore, they expected the provider to be in a rush, which would lead to a dissatisfying service. "I didn't want a dentist that was going to feel rushed because their day was packed" (respond. 85). Thus, potential customers have negative associations with a scarce offer set and choose another provider (probably a competitor) with more offered slots.

Flexibility: The third motive detected is flexibility. The subjects were asked to imagine a scenario in which they wanted an 8 a.m. appointment with a dentist. Both alternatives offered the preferred appointment time. However, many subjects wanted a greater assortment. "[F]lexible hours. [G]reater number of alternatives" (respond. 1180). Furthermore, many subjects were worried that they would be late for the early appointment or that it may have to be shifted to a later time. "How many appointments they had available, in case I was running late or needed to reschedule to a later time that day" (respond. 1267). Some others take into account the number of offered appointments close to the preferred time, no matter how many slots are offered in total. "How many time openings were close to the 8 AM timeslot that I wanted" (respond. 173). We pool those reasons in the flexibility motive. Note that future research may help further analyze this motive. The provider with the smaller offer set could offer more options around the predetermined slot to see whether behavior then changes.

Some respondents weigh several aspects. "[I]nitially, I imagined A being more sought after, and as such better (e[.]g[.,] herd mentality), but then again the cons are too many: less flexible timing, possibly overworked/tired, lack of information on both professionals" (respond. 476). "Well, it's weird. When the[re were] many appointments available, I was hesitant because that seemed to signal they were of low quality. When they had a medium amount, I appreciated that because it offered flexibility. So it was really a balance there. Too few and too much availability were both concerning for different reasons." (respond. 90). Nevertheless, one of the motives individually outweighs the other.

In Table 3, we summarize how frequently the main motives are mentioned as the general reasons for the choice behavior. Note that the question about motive is only asked once per subject for all three choices. The category "Other" comprises single entries and categories with entries that only one of the raters considered relevant.
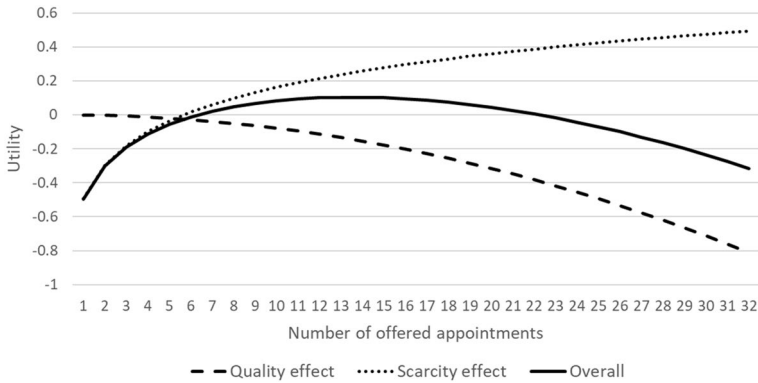
Fig. 3 Estimated impact of the quality-inference effect (dashed line), the scarcity effect (dotted line) and the overall utility (solid line) according to the independent observations, with $\widehat{\beta}_{QIE}^{indep} = -0.0253$ and $\widehat{\beta}_{SCAR}^{indep} = 0.1426$

For example, one of the raters detected single entries of choice overload-related motives ($<1\%$). Less waiting was not considered relevant by either coder and was thus also assigned to the category "Other". Overall, quality was mentioned by 34% of the subjects as a reason for their choice. The dominant motive for a preference of a smaller offer set seems to be due to the association with higher quality and with the more popular dentist. Being overwhelmed by too many offered appointments seems not to be relevant in the context of a dentist appointment. Forty-seven percent of the subjects considered flexibility. Scarcity shows a poor Cohen's kappa. The reason for the preference for a larger offer set seems to be the desired flexibility instead of crowded waiting rooms. The two main motives represent different behaviors (preference for a smaller/larger offer set) and, thus, support Hypothesis 1 and Hypothesis 4. Note that the scarcity effect in our context implies that customers (1) like to have flexibility and (2) dislike providers who are too busy.

4.3.3 Strength of the quality-inference and scarcity effects

We next estimate the strength of the quality-inference effect and the scarcity effect with R (version 4.1.0). Therefore, we use the multinomial model introduced in Sect. 3. Each provider has a capacity of $\kappa = 32$ appointments. $\mathcal{O}$ denotes the number of offered slots per alternative (provider A and provider B). Note that asc_j is set to zero for all $j \in J$ ($asc_j = 0 \quad \forall j \in J$), as the only noticeable difference between provider A and provider B is the offer set, which is already included elsewhere in the utility function.

Estimating the strength of the effects, we consider all independent observations (i.e., the first choice of each subject) and label the estimated coefficients by the superscript ^{indep}. We estimate weakly significant effects of $\widehat{\beta}_{QIE}^{indep} = -0.0253$ ($\sigma = 0.0125$, t-ratio $= -2.03$) and $\widehat{\beta}_{SCAR}^{indep} = 0.1426$ ($\sigma = 0.0581$, t-ratio $= 2.45$). Both effects show the expected directions, i.e., a negative impact of the quality-inference effect and a positive impact of the scarcity effect on the interest in the provider when

Fig. 4 Visualization of the service areas of the considered services, dentist (left), hairdresser (middle), government (right)

increasing the number of offered appointments. Figure 3 shows the curve of the interest in the provider when increasing the number of offered appointments according to the independent observations. Following the estimation and our utility function (see Sect. 3), we obtain the maximum utility when offering 13 appointments. In Appendix A3, we provide further estimates of the effect coefficients when building a reference category of having no booked appointments and when considering all observations. The estimates and with this the recommended number of offered appointments only marginally differ from the recommendation for all independent observations.

In summary, the results indicate the quality-inference effect as well as the scarcity effect when booking a general check-up appointment with the dentist. If a dentist faces an (almost) empty schedule, it is recommended that only a subset of the available appointments be offered to increase the customers' interest and the probability that they book an appointment with the dentist. Nevertheless, enough appointments should be offered to avoid negative associations and to give customers enough appointment times to choose from. According to our results, any offer set with $4 < \mathcal{O} < 32$ may be beneficial compared to offering all available slots.

All results refer to the analysis of a specific health care context and are thus not necessarily generalizable for all contexts. We therefore consider a second online experiment in which we investigate different service areas.

5 Online experiment II: service areas

5.1 Online experimental design and hypothesis

Our second online experiment analyzes the quality-inference effect and the scarcity effect for different services in several service areas. Our Online Experiment I focuses on one specific service in the health care sector, more precisely, on a service of a dentist. We investigate the transferability of those results by including two other service areas, namely, a hairdresser and a government entity.

Our related online experiment is again set up with the software LimeSurvey and distributed via Amazon Mechanical Turk. A total of 450 subjects completed the online experiment in November 2020 (November 27 until November 28). Again, each subject was paid 1 USD for completing the questionnaire.

Table 4 Treatment summary of Online Experiment II with fixed offers of A = 8 vs. B = 32 but different services

Treatment	Within-subject		
	First choice	Second choice	Third choice
T9	Dentist	Hairdresser	Government
T10	Hairdresser	Government	Dentist
T11	Government	Dentist	Hairdresser

Compared to Online Experiment I, we modify the second part and leave the first and third parts unchanged. A significant indication of the quality-inference effect for the comparison of a moderately sized offer set of eight appointments compared to an empty schedule (8 vs. 32 appointments) was already discovered. Therefore, we fix our comparison to these offer sets and only vary the service that is considered. In addition to the follow-up appointment with the dentist (minor changes in the text from Online Experiment I to obtain comparable contexts), we also consider an appointment with a hairdresser for a new haircut and an appointment at a government office for a new identity card. The scenarios for each service are given in Figs. 10, 11 and 12 in Appendix B1. Each subject is asked about all three services, but in a different order. To highlight the considered service, each scenario is accompanied by a picture (see Fig. 4). As no order effects are found in Kaluza et al. (2021), service provider A, on the left-hand side of the page, always offers the same eight appointments, while service provider B, on the right-hand side of the page, always offers all 32 appointments. Again, the response options (provider A, provider B) are displayed in random order. All other aspects of the scenario(s) remain unchanged from the prior online experiments. By changing only one aspect per online experiment, we can better trace differences in results back to differences in the considered services.

In Online Experiment II, which examines service areas, we consider three treatments—T9, T10 and T11—again in a between-subject design. A total of 150 subjects were randomly assigned to each treatment (no repeated participation in either online experiment). Table 4 presents the three treatments in order of the services. Subjects in treatment T9, for example, first decide whether to choose dentist A (with eight appointments) or dentist B (with 32 appointments). In the second scenario, subjects in treatment T9 choose between hairdresser A (with eight appointments) and hairdresser B (with 32 appointments), and in the third scenario, they choose between government office A (with eight appointments) and government office B (with 32 appointments). We vary the service of the appointment and use the dentist scenario as our base to replicate the results from Online Experiment I.

We expect the type of service to be decisive regarding the strength of the quality-inference effect. Services may vary in the quality of the output as explained in Sect. 1. We analyze a standardized and a non-standardized service: If you need a new identity card, the output will be a new identity card, no matter which government office you choose, provided this service is offered. A form may have to be filled out and handled. Nevertheless, the output of this process should not depend on the specific employee. Thus, the output of this standardized service from different

Table 5 Descriptive statistics of the first choice, varying service areas

Service area treatments	Dentist (T9)	Hairdresser (T10)	Government (T11)
Choice frequency in %	46: 54	58: 42*	28: 72***
(p value)	(0.278)	(0.088)	(0.000)
# Subj	72	79	81

*Significance at the 90% level

***Significance at the 99% level

employees should be equivalent (i.e., a new identity card). If you want a new haircut instead, you get a different haircut from different salons (or even from different hairdressers of one salon). For this non-standardized service, the quality of the provider plays an important role. These two services show that not each service may evoke the quality-inference effect the same way. We thus hypothesize that the results from Kaluza et al. (2021) and our Online Experiment I are not generalizable to all services. We expect the quality-inference effect to be present (but possibly with different strengths) for nonstandardized services (e.g., haircut) if the offer set in the appointment system is not scarce. However, we expect the quality-inference effect to be absent and only the scarcity effect to be present for standard services where the output does not differ (e.g., new identity card). We assume that different types of service areas exist: (1) areas in which some of the services evoke the quality-inference effect (e.g., financial institutions, credit card vs. investment advisory) and (2) areas in which quality either has an impact on the output throughout the services within the service area or not (e.g., wellness, government).

We chose our considered service areas and services that somehow represent the related service area regarding the importance of the output quality (type (2) service areas). Kaluza et al. (2021) already show that dentists may face the quality-inference effect and scarcity effect. We expect the hairdresser to reveal similar results. We further expect the service at a government office to be independent of the quality aspect. Thus, we hypothesize that the quality-inference effect is present for the hairdresser and the dentist scenarios but not for the government scenario; see Hypothesis 5.

Hypothesis 5: The quality-inference effect is present if quality affects the output, i.e., in the dentist and hairdresser scenarios, but not if the output is independent of the provider, i.e., in the government scenario.

5.2 Results

We next analyze the online experiment on booking an appointment with service providers in different service areas. To obtain comparable results, we clear the dataset in the same manner as that of Online Experiment I (see Sect. 4.3). Again, we exclude respondents if an attention check fails, if the answers are contradictory due to a specific answering pattern, if they have no relation to the online experiment

Table 6 Frequency of the motives for the choice behavior (232 subj.)

	Quality	Flexibility	Scarcity	Other
Frequency in % (Cohen's kappa)				
Dentist	34% (0.7459)	40% (0.8292)	15% (0.4746)	17% (0.3928)
Hairdresser	30% (0.6857)	37% (0.7226)	15% (0.5420)	26% (0.2352)
Government	9%*** (0.7498)	40% (0.6689)	30%*** (0.8159)	28% (0.6726)

The frequency is the geometric mean of the raters' results. Multiple reasons were possible, leading to a sum of the frequencies > 100%

***Significant differences of the motive frequency regarding the government service compared to the dentist and the hairdresser at the 99% level

or the respondent mentions that he or she does not have time in the morning. The cleaned and further analyzed dataset consists of 232 subjects, with 44% female and 56% male respondents. The age ranges from 19 to 70 years, while 84% are between 20 and 49 years old (33% 20–29 years, 35% 30–39 years, 14% 40–49 years). The majority live in the United States of America (82%), 7% in India and 6% in Brazil (the remaining 6% come from all over the world).

Table 5 summarizes the choice frequency per service area, considering the first choice of each subject (independent observations). We cannot replicate the previous results on the choice preference considering a dentist. Instead, we find a contradictory but nonsignificant tendency toward the larger offer set when choosing between two dentists (T9). In Sect. 6, we discuss possible reasons and explanations. However, we find support for Hypothesis 5 regarding the services of the hairdresser and the government office. The majority prefers the hairdresser with the smaller offer set (58%, T10). This choice frequency shows a weakly significant difference from choice-by-chance (binomial test, $\alpha < 0.1$). In contrast, the government office with a full offer set is preferred by 72% of the subjects in treatment T11. This highly significant result indicates that the quality-inference effect has no or hardly any influence on standardized services such as a new identity card from a government office (binomial test, $\alpha < 0.01$). Instead, scarcity and desired flexibility dominate. When considering all three observations of each subject, we obtain similar results for the dentist and the government services but no longer significant support for the hairdressing service. We provide the descriptive statistics for all choices in Table 12 in Appendix B2.

In summary, we find some evidence for the quality-inference effect for services where quality influences output. While we find choice frequencies that clearly point in this direction (Hairdresser, T10), we also see choice frequencies that are not significantly different for random choice. Most importantly, in scenarios where quality apparently does not matter (Government, T11), we observe choice frequencies that point substantially and highly significantly toward the less utilized provider, indicating that a countervailing quality-inference effect is not present.

5.3 Motives for the choice behavior

The results for the appointment at a government office show that customers significantly prefer a bigger offer set to a smaller offer. We next investigate the motives for the subjects' choices of all 232 subjects regarding the different service areas. Note that we asked for the motive separately for each service area. Similar to Online Experiment I, we let the two independent raters code the explanations for the subjects' choice behavior per service area (see Sect. 4.3.2), again blinded to the hypothesis. We depict the frequency of the motives for the choice behavior of all subjects in Table 6.

For the dentist and the hairdresser appointments, quality is relevant for approximately one-third of the subjects (34% and 30%, respectively). Flexibility is also the main motive (40% and 37%, respectively), whereas scarcity is not as relevant (15% each). Considering the motives for the choice behavior when booking an appointment with an government office, the quality motive is significantly less relevant (9%) than the other contexts (Fisher's exact test, $\alpha < 0.01$). Instead, scarcity is much more relevant (Fisher's exact test, $\alpha < 0.01$). Flexibility is relevant regardless of the service area (37% to 40%). Cohen's kappa is acceptable for all service areas and motives apart from the scarcity motive for the service of a dentist and from the collective category for single entries.

We do not estimate the strength of the effect, as the dataset for each service is too small to obtain meaningful results and only considers the comparison of eight and 32 available appointments. This online experiment mainly focuses on the presence of the effects in different service areas and not on the overall strength of the effect for each service.

6 Robustness check with follow-up discrete choice experiments

To check the robustness of our results, we conduct further empirical follow-up studies. In particular, we perform discrete choice experiments (DCE) and estimated individual preferences for several service attribute levels with Hierarchical Bayes-Multinomial Logit (HB-MNL) models (Allenby and Rossi 2006; Hein et al. 2020). Obviously, from a practical point of view, the use of a segmentation model like the Latent Class MNL model, in contrast to an individual-level model, i.e., HB-MNL model, may be helpful to derive specific marketing implications. However, since we are interested in a deeper understanding of customers' choice behavior in online appointment systems and in particular in the investigation of the presence of the quality-inference effect and the scarcity effect, an individual model is best appropriated here. In contrast to individual models, segmentation models estimate aggregated preferences on segment levels and therefore conceal—to some extent—preference heterogeneity within segments.

6.1 Hierarchical Bayes-Multinomial Logit Model

In accordance with our previous approach, we assume that respondent i chooses alternative j, which provides the largest utility U_{ij} to him or her. However, we do not aim to

estimate aggregated scarcity effects $\widehat{\beta}_{SCAR}^{indep}$ and aggregated quality-inference effects $\widehat{\beta}_{QIE}^{indep}$ as before. Rather, we are interested in the individual preferences of respondents for certain attribute levels that describe a certain booking alternative. Therefore, the utility function reads as follows:

$$U_{ij} = x_j \cdot \beta_i + \varepsilon_i \tag{2}$$

where $i = 1, \ldots, I$ denotes respondent i, and x_j displays the effect coding vector of the booking alternative $j = 1, \ldots, J$. The preference vector β_i in Eq. (2) is individual to each respondent i, and ε_i describes a Gumbel distributed error term. This random error term captures all those effects that are not included in the deterministic part of the utility. Following the resulting MNL model, the probability of respondent i choosing alternative j from choice set t resembles:

$$P_{itj} = \frac{\exp\left(\mu \cdot x_{tj} \cdot \beta_i\right)}{\sum_{r=1}^{R} \exp\left(\mu \cdot x_{tr} \cdot \beta_i\right)}, \mu > 0 \tag{3}$$

where μ is a positive scale parameter that is commonly fixed to one and R is the number of alternatives in choice set t, $t = 1, \ldots, T$. To achieve the HB-MNL model, we assume β_i in Eq. (3) to be a multivariate Gaussian distribution, i.e., $\beta_i \sim N(\sigma, \Sigma)$, where the covariance matrix Σ in turn is an inverse Wishart distribution. The estimation of the covariance matrix Σ and the mean σ is performed by Gibbs sampling, while the conditional posteriori distribution of the individual parameter vectors β_i are estimated via a Metropolis–Hastings algorithm.[7]

6.2 Data

For our robustness check, we focus on respondents' booking choice behavior for dentists' practices and hairdressing salons and abstained from the investigation of respondents' booking choices for government offices. The main reason lies in difficulties in the description of government offices by a certain number of appropriate attributes, which is mandatory for DCEs (Zwerina et al. 1996).

For our two DCEs (here: dentist and hairdresser), we consider the following attributes and associated levels for dentist's and hairdresser's DCE: number of free slots (2/8/16/24/32), travel time to dentist's practice and hairdresser salon (up to 15 min/15 to 30 min/more than 30 min), star rating (1/2/3/4/5), gender of dentist (male/female), and price level of haircut (lower than average/average/higher than average). We consider attributes that were stated as further important attributes for the final booking decision of a dentist's practice and a hairdresser salon by the respondents in our previous studies, in which we consider the number of free slots only. In addition, these attributes match the attributes from recent literature in the field of choice of physicians/dentists (Kiiskinen et al. 2010; Victoor

---

[7] A detailed explanation on the estimation procedure of the HB-MNL model is out of scope of this paper. We refer the interested reader to the description given by Train (2001).

Table 7 Stated attribute importance

Dentist study [in %]		Hairdresser study [in %]	
Number of free slots:	20.11 (sd: 14.12)	Number of free slots:	15.23 (sd: 14.39)
Travel time:	28.29 (sd: 17.22)	Travel time:	24.25 (sd: 17.21)
Star rating:	40.55 (sd: 21.61)	Star rating:	28.38 (sd: 19.97)
Gender of dentist:	11.04 (sd: 14.77)	Price level:	32.14 (sd: 17.87)

et al. 2012; Clark et al. 2014) and service facilities such as hairdressing salons (Ostrom and Iacobucci 1995; Wu et al. 2011). In particular, several studies prove the importance physician's gender for patients' choice, e.g. female patients prefer female physicians (Alyahya et al. 2019) or the attribution of certain gender-specific characteristic, like highly social skills of female physicians increase patients' preferences for female physicians (Kerssens et al. 1997). For the price attribute in the hairdresser study, we abstained from the notation of concrete prices, because the price for a dry hair cut differs dramatically between gender, hairdresser salons and locations of hairdresser salons. Instead we used more general levels, e.g., lower than average, average. This approach is straight forward for our study, because we are not interested in concrete values for consumers' willingness-pay-for dry haircut, but rather in the importance of the price attribute for the final booking decision.

Based on these attributes and associated levels, we build 16 choice sets that included four hypothetical booking alternatives as well as a no-purchase option. The choice sets were generated with the SSI module of Sawtooth Software and consider individualized choice designs. In particular we used the default method "Balanced overlap", which allows for a modest degree of attribute level repeating and is recommend by Sawtooth Software (2009). We set up two distinctive questionnaires: one focused on the booking choice for hairdressing salons, and the other focused on the booking choice for dentist practices. In both questionnaires, the choice task built one part of our questionnaire, while the other part comprised some demographic questions regarding gender and age of the respondents as well as a constant sum scale on the importance of different attributes for the final booking decision. The questionnaires were distributed by a German market research institute in July 2022 to a German sample, which was representative w.r.t the gender and age distribution of German adults.

For the dentist study, 314 respondents completed the questionnaire, while 311 answered the hairdresser study. We follow common practice and cleaned the data for non e-straightliners, i.e., respondents who chose the no-purchase option in at least 75% of all choice sets (Paetz 2021). Such straightlining behavior suggests that such respondents either did not take the choice task seriously or abstained from consulting a dentist or a hairdresser per se. While the former behavior could trigger biased results, the latter leads to results that do not match our target group. Hence, an elimination of the data of those respondents is straightforward.

For our dentist study, the final sample consists of 302 respondents, with 154 females and 158 males. The mean age was 47.76 (sd: 16.21). The choice share of the no-purchase option was 4.85%, lower than the recommended upper bound of 15% for a sound database (Johnson and Orme 2003, p. 4). In the hairdresser study, the final sample included 291 respondents with 148 females and 143 males. The mean age was 47.00 (sd: 15.75). The choice data yield a no-purchase-option's share of 9.01%.

6.3 Results

First, we inspect the directly stated importance of the study's attributes for the final booking decision. Table 7 yields the results.

Obviously, respondents stated that they consider all attributes to be important for their final booking decision. This maintains a sound setup of our DCEs, because the respondents feel all attributes as important for their choice building. Furthermore, the stated importance of our focal attribute "number of free slots" differs across studies. While respondents stated that the number of free slots was less important when booking an appointment at a hairdresser (15.23%), it was nearly twice as important than the gender of the dentist (20.11%). However, we obtain high standard deviations, indicating that some people actually care significantly about the number of free slots, while other respondents do not. This observed heterogeneity gives a first hint on prevalent preference heterogeneity that could be captured by the estimation of a model, such as the HB-MNL model, that accommodates heterogeneity on an individual level.

We estimate the data of both studies by using the CBC/HB software of Sawtooth Software (Sawtooth 2021). We abstain from setting constraints and consider 20,000 iterations and skipped the first 10,000 iterations as the burn-in phase.[8] We use 14 choice sets for the estimation and left two choice sets as hold out choice sets, respectively. The fit statistics yield well-fitted models in both studies. In the dentist study, the log-likelihood value of the HB-MNL model is -2061, the RLH is 0.614 and the percent certainty is 0.697. The values of both the RLH and the percent certainty are high (for a detailed explanation of these fit statistics, see Paetz et al. 2019) in both studies, which displays a good fit of the models. In the hairdresser study, the log-likelihood value is -2491, the RLH is 0.543 and the percent certainty is 0.620. These results also argue for a well-fitted model. Both model yield a well predictive validity with first-choice hit rates of 71.52% for the dentist study and 65.98% for the hair dresser study. The hit rates are more than 3.5 resp. 3 times larger than the probabilities by chance, i.e., 20%. Furthermore, we calculated an aggregated measure, i.e., the mean absolute error (MAE). The MAE was 0.013 for the dentist study and 0.017 for the hairdresser study. The MAE values in both studies are very close to 0 representing excellent (ex-post) forecasts (cp. Paetz et al. 2019).

---

[8] For both studies, we checked the convergence plots in order and can assure convergence for the parameters.

Table 8 Dentist study's population mean parameters

	Mean beta	Std-err
Number of free slots		
2 of 32	0.031	0.110
4 of 32	0.428	0.100
8 of 32	0.300	0.080
16 of 32	− 0.028	0.090
32 of 32	− 0.731	0.120
Star rating		
1 star	− 4.663	0.430
2 stars	− 2.632	0.330
3 stars	0.276	0.130
4 stars	3.038	0.200
5 stars	3.981	0.260
Travel time		
Less than 15 min	1.551	0.130
15–30 min	0.615	0.090
More than 30 min	− 2.166	0.170
Gender of dentist		
Male	0.124	0.070
Female	− 0.124	0.070
None	− 0.909	0.350

The estimated population means of the part-worth parameters as well as the standard errors for our dentist study are displayed in Table 8. The estimates for our hairdresser study are shown in Table 9.

First, we use the results and calculated the relative importance of attributes for the final booking decision. The star rating is most important for the booking decision for both a dentist (62.78%) and a hairdresser (43.35%). In contrast, the number of free slots is less important in both studies (dentist: 8.42%, hairdresser: 5.03%). However, both attributes, e.g., the number of free slots and star ratings, correspond to respondents' perceptions of the quality of the offered services. We, therefore, must keep two things in mind. (1) It is likely that the respondents are much more familiar with the star rating attribute as a direct quality signal than with the number of free slots and, therefore, attach greater importance to the star rating attribute. However, both attributes are shown in real booking systems such as zocdoc.com, and it is therefore straightforward to incorporate both attributes in our DCE, even though the importance of the number of free slots is reduced due to the display of a star rating. (2) We already detect a great amount of preference heterogeneity in our studies, and an inspection of the estimated covariance matrices maintains this result in both studies. There are indeed people who attach great importance to the number of free slots, while others do not care much about this attribute: We counted 39 (out of 291) respondents in the hairdresser study and 29 (out of 302) respondents in the dentist study who attach a higher importance than expected, i.e., higher than $\frac{1}{4} = 25\%$, to the number of tree slots. Furthermore, the individual relative importance for the

Table 9 Hairdresser study's
population mean parameters

	Mean beta	Std-err
Number of free slots		
2 of 32	− 0.076	0.090
4 of 32	0.247	0.080
8 of 32	0.110	0.090
16 of 32	0.102	0.090
32 of 32	− 0.383	0.110
Star rating		
1 Star	− 2.968	0.230
2 Stars	− 1.997	0.160
3 Stars	0.575	0.100
4 Stars	1.934	0.140
5 Stars	2.455	0.170
Travel time		
Less than 15 min	1.493	0.120
15–30 min	0.405	0.080
More than 30 min	− 1.897	0.140
Price level		
Lower than average	1.050	0.110
On average	0.968	0.090
Above average	− 2.018	0.150
None	0.388	0.210
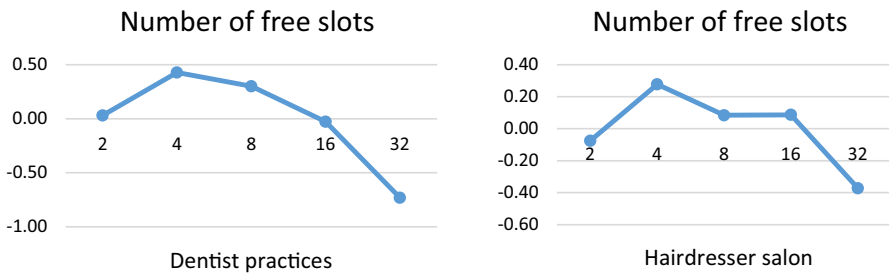
Fig. 5 Plots of population mean parameters

number of free slots ranges between 1.53 to 87.09% in the dentist study and between
1.60 and 72.26% in the hairdresser study.[9]

---

[9] For the purpose of comparability, we also estimated a LC-MNL model for both studies. Although, in both studies, the estimation of a LC-MNL model yields one segment, that attaches a higher attribute importance to the attribute "number of free slots" than the other segments, it is not able to capture the entire extent of preference heterogeneity. Once more, this justifies the use of the state-of-the-art individual model, i.e., the HB-MNL model.
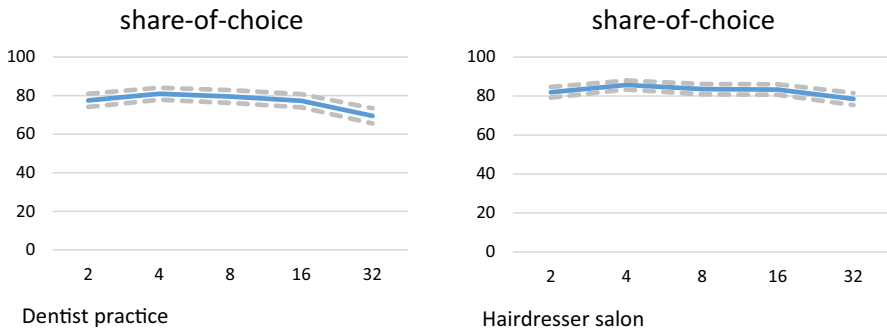
Fig. 6 Share-of-choices of services with varying number of free slots versus outside good (solid line) including 95% confidence intervals (dashed lines)

Overall, we are able to replicate the results from our previous studies in the context of preferences for a lower or higher number of free slots. For both studies, we find an inverse U-shaped curve for respondents' preferences for an increasing number of free slots. Figure 5 displays the corresponding plots.

In both studies, an empty schedule (32 free slots) is clearly rejected, while four (out of 32) free slots are mostly preferred.

To underline these results in a more practical manner and to abstain from an over-interpretation of utility estimates, we set up market simulation studies of certain dentist practices resp. hairdresser salons against the outside option, respectively. For both studies, we varied the number of free slots and fixed the levels of the other attributes. For the hairdresser study, we choose "3 stars, 15 min travel time or less, average price level" and for the dentist study "3 stars, 15 min travel time or less, male" as fixed attribute levels. We then calculated the preferences based on the results of the estimated models and used the logit model's share-of-choice rule. Figure 6 displays the results:

For both type of services, i.e., dentist practices and hairdresser salons, we observe inverse U-shape curves for the share-of-choices for a increasing number of free slots. For example, the choice shares of a dentist service with an attribute level combination of "32 free slots, 3 stars, 15 min or less, male" shows a share of choices of 69.49%, (i.e., the outside option achives a share of choices of 30.51%) and a dentist service with attribute level combination "2 free slots, 3 stars, 15 min or less, male" holds a share of choices of 77.55%. In contrast, a moderate number of free slots, i.e., 4, 8 and 16 free slots, are more preferred and result into choice shares of 80.97% (4 (out of 32) free slots). A similar picture becomes obvious for the hair dresser study: Here, the choice shares of a hairdresser (against the outside option) are lowest for 2 and 32 (out of 32) free slots, i.e., 81.93% and 78.48%, respectively. Obviously, too many and too less numbers of free slots lead to lower choice shares of the respective dentist practices or hairdresser salon versus the respective outside option in comparison to medium numbers of free slots.

Considering this result and the inverse U-shapes of the preference curves, we conclude that the scarcity effect and the quality-inference effect are robust behavioral phenomena, even if customers account for information other than the booking status.

7 Discussion and limitations

We investigate customers' choice behavior in online appointment systems. In Online Experiment I, we consider several occupancy rates. Fixing the overall capacity per provider to 32 appointments helped us replicate the main results from Kaluza et al. (2021) and compare the results. Varying the overall capacity in future research would be interesting to see whether the results are similar. We expect our main insights to remain unchanged. On an absolute or relative basis, the results may differ slightly from ours. Nevertheless, having an empty schedule shows absent demand, and a fully utilized schedule comes with a lack of flexibility in the customer's choice. Both appointment schedules may lead to a decrease in the customers' interest in the provider, no matter the overall number of appointments in the system. Furthermore, we do not include a full-factorial design to compare all possible occupancy rates but rather consider different specific comparisons. Further research is needed to determine whether our results are replicable and whether different comparisons of occupancy rates lead to similar results.

We consider three different service areas in Online Experiment II. This analysis gives interesting insights into the quality-inference effect and the scarcity effect. However, we only consider one specific context (service) of each service area. This gives insights into specific services but is neither necessarily generalizable to other services of the considered areas nor necessarily transferable to other service areas. We choose simple examples to analyze whether customer choices are related to the number of slot offers for standardized and nonstandardized services. Further research may help understand to what extent the results are transferable to other settings and contexts.

In Online Experiments I and II, we use a layout in which only the offered appointments are displayed. However, some subjects had difficulties understanding that the remaining (nondisplayed) appointments exist but are simply not offered to them. It would be interesting to see whether a different layout, e.g., with unavailable appointments grayed out but visible, would lead to similar insights. In fact, subjects may have all sorts of associations surrounding the reasons for unavailable slots, e.g., service hours, holidays, and booked slots. Since we assign subjects randomly, this nuisance variable should not confound our main effect. However, a better understanding consumers' perceptions might be of interest for the management of online appointment systems. We further find that some subjects disregarded the predetermined preferred 8 a.m. appointment that was given to avoid biases due to individual daily schedules. We conjecture that highlighting that information about the preferred 8 a.m. appointment would give a higher quality of the dataset, which would lead to less deleted data but not to a change in the significant results. Other subjects focused on the appointments close to the 8 a.m. appointment in case the appointment had to be postponed. It would be interesting to see how results would change with different offer sets (but the same

number of offered appointments) focusing on less variation in the appointments close to the preferred one. We conjecture that it could slightly weaken the flexibility effect. Another possible online experiment would be to vary the preferred appointment time and compare the results. Overall, we see in our discrete choice experiments in Sect. 6 where we do not fix the time of the appointment and do not show appointment schedules, that the inverse U-shaped preference curve is robust.

We replicate the results from Kaluza et al. (2021) regarding the health context (dentist, 8 vs. 32) in Online Experiment I but not in Online Experiment II, which could be further evaluated. Several reasons are conceivable. Marginal changes, such as the added pictures, the change in the scenario text, possible changes in the subject pool (sex, age, qualification, etc.), or a type-I error could evoke the nonreplicated results. It is an exciting avenue of future research to further investigate the robustness of the quality-inference context both within and between service areas. Note that the health sector may be a controversial sector, as in some regions and medical specialties, there is a lack of physicians and health systems that vary between countries. Overall, the robustness check in Sect. 6 provides additional evidence that both the quality-inference effect and the scarcity effect are at work when consumers observe the utilization of a service provider.

As some of our results are not statistically significant, further research may help understand the findings in detail. One reason may be a decisive aspect for several subjects. Many subjects chose the dentist that had more opportunities in case the appointment had to be postponed for any reason (e.g., the subject was late, as 8 a.m. is very early, or the dentist does not accept the appointment). The possibility of a delay or cancellation of the appointment was not considered in the scenario given to the subjects. However, it was not explicitly mentioned in the scenario that the chosen appointment would definitely take place. It may also be interesting to change the scenario by noting that delays are not considered and an offered appointment that is taken by the customer is guaranteed to take place as agreed upon. Furthermore, a different choice of displayed appointments keeping the number of offered appointments unchanged could give more insights into the choice due to the postponing motive. More appointments around the predetermined 8 a.m. preference may weaken the (postponing) flexibility motive and strengthen the quality motive. Further, it would be interesting to learn more about the interaction of the importance of the quality aspect in the service provider choice, the providers' utilization rate, and choice behavior, e.g., via mediation analysis.

The follow-up discrete choice experiments come certainly with the limitation of their stated-choice character. However, no market data is available so far and the efficiency of (stylized) discrete choice experiments for the evaluation of consumer preferences is well-established. We did not incorporate textual consumer reviews as an attribute, but rather used a star-rating as it is common practice on online booking platforms. Investigating the role of reviews appears to be an interesting avenue for future research. Overall our discrete choice experiments revealed only minor aggregated importances for the attribute level "number of free slots", but we showed great preferences heterogeneity for this attribute, too. Hence, future research may investigate this phenomea in more detail and may seek for differences between varying service categories as well as consumers' individual variables, e.g., personality, cultural context etc. Obviously, due to our desire to mirror real choice decisions in online booking systems, we used attributes in the discrete choice study that are related to some extent, e.g., star rating and free

slots. Future studies may therefore incorporate interaction effects in the estimations to tackle this point in more detail.

Finally, all experiments rely on hypothetical choices. It is conceivable that the importance of the travel time attribute and the price attribute are understated, since the decisions have no real consequences in these dimensions. Yet, it is also conceivable that the quality effect is even more pronounced if low quality will translate to a poor service in reality (e.g., a bad haircut). Testing the prevalence of the identified quality effects and scarcity effects in incentivized experiments or, even better, in a real-word environment is an exiting opportunity for future research, but beyond the scope of this study.

8 Conclusion

We analyze customers' choice behavior in online appointment systems to investigate the presence of the quality-inference effect and the scarcity effect. In two stylized online experiments, we let subjects decide between two service providers of different service areas who only differ in the number of offered appointments. We show that the two effects (quality and scarcity) have a nonlinear impact on customers' interest in the service provider. We find some evidence for the quality-inference effect, meaning that customers prefer providers who offer fewer appointments, as they often associate a smaller offer set with higher quality. Already with a rarely utilized booking system compared to an empty schedule, customers' interest is increased. However, customers also like to have flexibility when choosing an appointment. Thus, a too-scarce offer set may, in contrast, lead to deterred customers. The strength of the quality effect depends on the service and other available quality signals (e.g., star ratings). For a standardized service with a government office, customers apparently solely focus on flexibility, not quality, when considering booking status. Overall, we find that the number of free slots appears to have a minor importance for the service choice on an aggregate level, when other attributes such as prices, star-ratings, or travel time are taken into consideration. However, we also find a large degree of heterogeneity and the relative importance ranges e.g. between 1.53 to 87.09% (dentist study). From a managerial perspective, our results therefore indicate that managing the number of free slots actively is of minor importance when (a) the service provider cannot target those consumers who put a relatively high importance to the booking status individually and (b) other quality information (such as star-ratings) are available.

Appendix A: Online experiment I supplements

A1 Design

Figure 7 shows an exemplary scenario of Online Experiment I.

Table 10 Descriptive statistics of all choices varying the occupancy rates in Online Experiment I

Treatments	1 vs. 2 (T8)	1 vs. 4 (T6)	1 vs. 8 (T7)	1 vs. 24 (T1)	2 vs. 4 (T3, T7)	2 vs. 8 (T1)	2 vs. 16 (T4)	2 vs. 24 (T5)	2 vs. 32 (T2)
Choice freq. %	35: 65***	35: 65***	40: 60**	51: 48	44: 56*	47: 53	38: 62***	35: 65***	40: 60**
(p-value)	(0.003)	(0.003)	(0.030)	(0.459)	(0.051)	(0.301)	(0.010)	(0.003)	(0.034)
# Subj	93	93	102	92	198	92	98	94	97

Treatments	4 vs. 8 (T5)	4 vs. 16 (T8)	4 vs. 24 (T2)	4 vs. 32 (T1, T4)	8 vs. 16 (T2)	8 vs. 24 (T4, T6)	8 vs. 32 (T3)	16 vs. 24 (T3)	16 vs. 32 (T5, T6)	24 vs. 32 (T7, T8)
Choice freq. %	39: 61**	33: 67***	46: 54	48: 52	45: 55	52: 48	59: 41**	59: 41**	50: 50	55: 45*
(p-value)	(0.025)	(0.001)	(0.271)	(0.358)	(0.208)	(0.332)	(0.041)	(0.041)	(0.500)	(0.076)
# Subj	94	93	97	190	97	191	96	96	187	195

*Significance at the 90% level

**Significance at the 95% level

***Significance at the 99% level
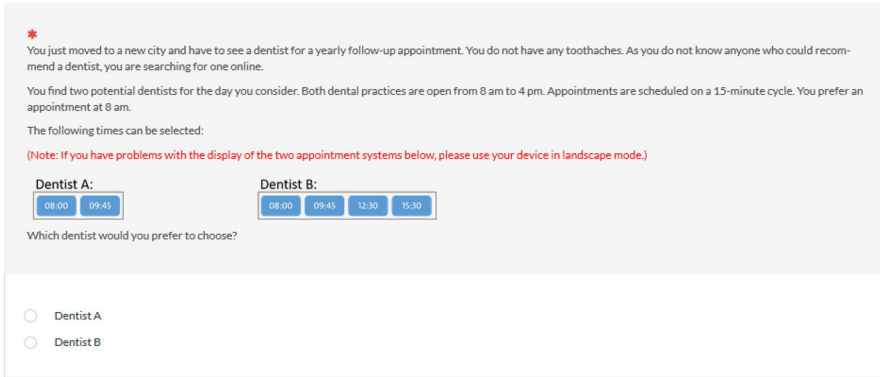
Fig. 7 Exemplary scenario of Online Experiment I (T7)

A2 Descriptive results of all observations

Table 10 summarizes the descriptive statistics of all observations of Online Experiment I with varying occupancy rates. We use a binomial test to check for random choice (50% each).

A3 Further estimations of the effect coefficients

We next provide further estimations of the effect coefficients. Focusing on the independent observations, we consider all treatments that include the full offer set (T2, T3, T4, T5, T8) to fix one offer set and only vary the comparison. Hereby, we create a reference category and label the estimated coefficients by the superscript ^{ref}. While the estimation of the effect coefficients considering all independent observations in Sect. 4.3.3 gives overall decision support on the number of appointments to offer, this regression may give more precise decision support for providers who still face an empty schedule. Our estimate for the quality effect, $\widehat{\beta}_{QIE}^{ref} = -0.0335$ ($\sigma = 0.0206$,
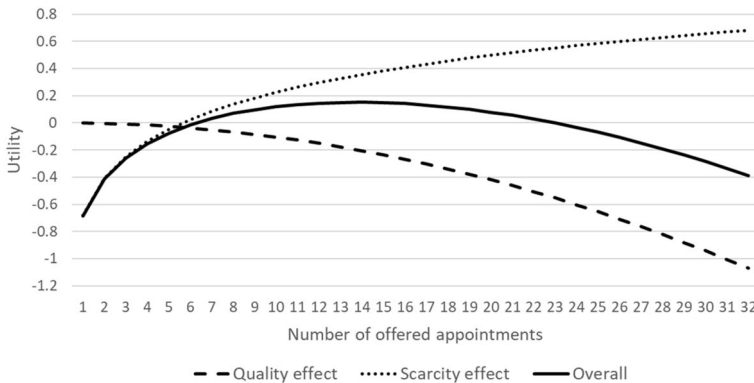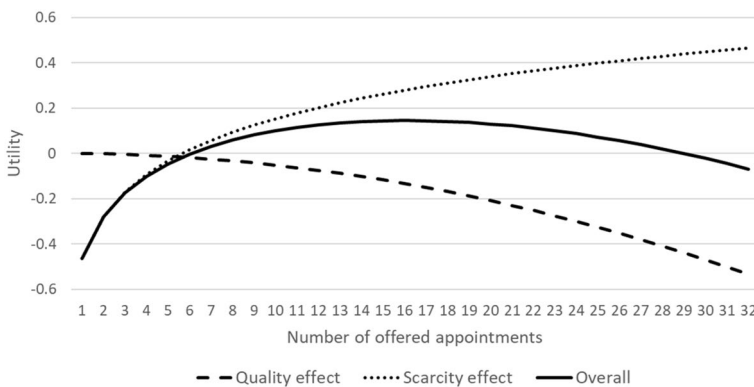

Fig. 8 Estimated impact of the quality-inference effect (dashed line), the scarcity effect (dotted line) and the overall utility (solid line) according to the independent observations of T2, T3, T4, T5, T8, with $\widehat{\beta}_{QIE}^{ref} = -0.0335$ and $\widehat{\beta}_{SCAR}^{ref} = 0.1970$

Table 11 Estimates of $\widehat{\beta}$-coefficients for all observations of Online Experiment I, $\widehat{\beta}_{T1}$ is the base for the fixed effect, size = size of the offer set

		$\widehat{\beta}$	Std. error	t-ratio	Rob. std. error	Rob. t-ratio
All observations	QIE	−0.0167	0.0040	−4.13	0.0034	−4.86
	SCAR	+0.1341	0.0226	+5.93	0.0222	+6.04
	T2	+0.0411	0.1105	+0.37	0.1876	+0.22
	T3	+0.0244	0.1109	+0.22	0.1893	+0.13
	T4	+0.0466	0.1103	+0.42	0.1860	+0.25
	T5	−0.0444	0.1123	−0.40	0.1959	−0.23
	T6	−0.0586	0.1127	−0.52	0.1985	−0.30
	T7	+0.1184	0.1091	+1.09	0.1794	+0.66
	T8	+0.0794	0.1125	−0.71	0.1972	−0.40

t-ratio = −1.62), is nonsignificant, whereas the estimate for the scarcity effect is weakly significant with $\widehat{\beta}_{SCAR}^{ref} = 0.1970$ ($\sigma = 0.0987$, t-ratio = 2.00). These estimated coefficients lead to a maximized utility when offering 14 appointments (see Fig. 8) and do not differ substantially from the results of all independent observations in Sect. 4.3.3.

We next estimate the effect coefficients considering all observations (see Table 11) to check whether our results are robust. To account for treatment effects, we add fixed order effects in our utility model (one $\widehat{\beta}$-coefficient per treatment, $\widehat{\beta}_{T1}, \widehat{\beta}_{T2}, \widehat{\beta}_{T3}, \widehat{\beta}_{T4}, \widehat{\beta}_{T5}, \widehat{\beta}_{T6}, \widehat{\beta}_{T7}, \widehat{\beta}_{T8}$). The individual error term is also considered and included in the (higher) robust standard error. Considering all observations (i.e., all three choices per subject), the estimated coefficients labeled with the superscript ^{all}, we estimate significant effects of $\widehat{\beta}_{QIE}^{all} = -0.0167$ ($\sigma = 0.0040$, t-ratio = −4.13) and $\widehat{\beta}_{SCAR}^{all} = 0.1341$ ($\sigma = 0.0226$, t-ratio = 5.93). Both effects again show the expected directions and hardly change from the analysis of the independent observations in Sect. 4.3.3.

The estimates of all observations lead to a maximized utility when offering 16 appointments; see Fig. 9.

Fig. 9 Estimated impact of the quality inference effect (dashed line), the scarcity effect (dotted line) and the overall utility (solid line) according to all observations, with $\widehat{\beta}_{QIE}^{all} = -0.0167$ and $\widehat{\beta}_{SCAR}^{all} = 0.1341$

Appendix B: online experiment II supplements

B1 Design

Figures 10, 11 and 12 show the scenarios for the dentist (the hairdresser and the government) of Online Experiment II.

Fig. 10 Scenario for the dentist in Online Experiment II
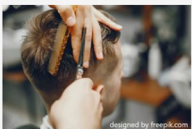

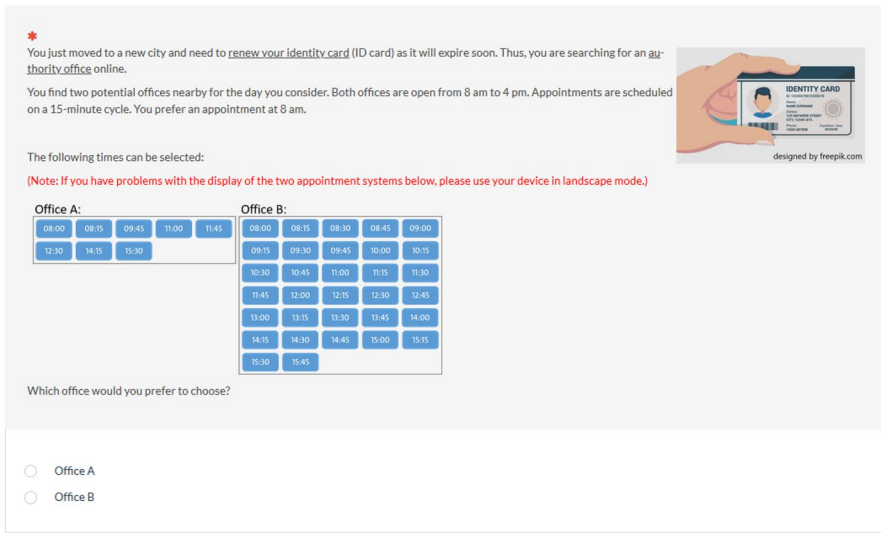**Fig. 11** Scenario for the hairdresser in Online Experiment II

Fig. 12 Scenario for the government office in Online Experiment II

Table 12 Descriptive statistics of all observations in Online Experiment II, varying the services

Treatments	Dentist (T9)	Hairdresser (T10)	Government (T11)
Choice frequency %	45: 55	49: 51	27: 73***
(p-value)	(0.084)	(0.371)	(0.000)
# Subj	232	232	232

***Significant on a 99%-level

B.2 Further results

Table 12 summarizes the descriptive statistics of all observations of Online Experiment II with varying service areas. Again, we use a binomial test to check whether the frequency deviates from choice-by-chance.

Declarations

Conflict of interest Not applicable.

References

Acemoglu D, Dahleh MA, Lobel I, Ozdaglar A (2011) Bayesian learning in social networks. Rev Econ Stud 78(4):1201–1236. https://doi.org/10.1093/restud/rdr004

Allenby GM, Rossi PE (2006) Hierarchical bayes models. In: Grover R, Vriens M (eds) The handbook of marketing research: Uses, misuses, and future advances. Sage Publications, London, pp 418–440

Alyahya G, Almohanna H, Alyahya A, Aldosari M, Mathkour L, Aldhibaib A, Al-Namshan Y, Al-Mously N (2019) Does physicians' gender have any influence on patients' choice of their treating physicians? J Nat Sci Med 2(1):29–34. https://doi.org/10.4103/JNSM.JNSM_28_18

Banerjee AV (1992) A simple model of herd behavior. Q J Econ 107(3):797–817. https://doi.org/10.2307/2118364

Becker GS (1991) A note on restaurant pricing and other examples of social influences on price. J Polit Econ 99(5):1109–1116. https://doi.org/10.1086/261791

Ben-Akiva M, Lerman SR (1985) Discrete choice analysis: theory and application to travel demand. MIT Press, Cambridge

Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. J Polit Econ 100(5):992–1026. https://doi.org/10.1086/261849

Byrt T, Bishop J, Carlin JB (1993) Bias, prevalence and kappa. J Clin Epidemiol 46(5):423–429. https://doi.org/10.1016/0895-4356(93)90018-V

Çelen B, Kariv S (2004) Observational learning under imperfect information. Games Econ Behav 94(3):484–498. https://doi.org/10.1016/S0899-8256(03)00179-9

Chernev A, Böckenholt Ulf, Goodman J (2015) Choice overload: a conceptual review and meta-analysis. J Consum Psychol 25(2):333–358

Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW (2014) Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics 32(9):883–902. https://doi.org/10.1007/s40273-014-0170-x

Debo LG, Parlour C, Rajan U (2012) Signaling quality via queues. Manag Sci 58(5):876–891. https://doi.org/10.1287/mnsc.1110.1454

Eppler MJ, Mengis J (2004) The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. Inf Soc 20(5):325–344. https://doi.org/10.1080/01972240490507974

Eyster E, Andrea G, Kartik N, Rabin M (2014) Congested observational learning. Games Econ Behav 87:519–538. https://doi.org/10.1016/j.geb.2014.06.006

Feldman J, Liu N, Topaloglu H, Ziya S (2014) Appointment scheduling under patient preference and no-show behavior. Oper Res 62(4):794–811. https://doi.org/10.1287/opre.2014.1286

Gerchak Y, Gupta D, Henig M (1996) Reservation planning for elective surgery under uncertain demand for emergency surgery. Manag Sci 42(3):321–334. https://doi.org/10.1287/mnsc.42.3.321

Giebelhausen MD, Robinson SG, Cronin JJ (2011) Worth waiting for: increasing satisfaction by making consumers wait. J Acad Mark Sci 39(6):889–905. https://doi.org/10.1007/s11747-010-0222-5

Gupta D, Wang L (2008) Revenue management for a primary-care clinic in the presence of patient choice. Oper Res 56(3):576–592. https://doi.org/10.1287/opre.1080.0542

Hassin R, Haviv M (1997) Equilibrium threshold strategies: the case of queues with priorities. Oper Res 45(6):966–973. https://doi.org/10.1287/opre.45.6.966

Hein M, Kurz P, Steiner W (2020) Analyzing the capabilities of the HB logit model for choice-based conjoint analysis: a simulation study. J Bus Econ 90:1–36. https://doi.org/10.1007/s11573-019-00927-4

Iyengar S, Lepper M (2000) When choice is demotivating: can one desire too much of a good thing? J Pers Soc Psychol 79:995–1006. https://doi.org/10.1037//0022-3514.79.6.995

Jin C, Debo L, Iravani S, Kremer M (2015) Observational learning in congested environments with multiple choice options: the wisdom of majorities and minorities. Chicago Booth Research Paper 8 **(Working paper)**

Johnson R, Orme B (2003) Getting the most from CBC, Sawtooth Software Inc. https://sawtoothsoftware.com/resources/technical-papers/getting-the-most-from-cbc

Kaluza I, Voigt G, Haase K (2021) Control of online-appointment systems when the booking status signals quality of service. Working Paper. University of Hamburg. Hamburg

Kerssens JJ, Bensing JM, Andela MG (1997) Patient preference for genders of health professionals. Soc Sci Med (1982) 44(10):1531–1540. https://doi.org/10.1016/s0277-9536(96)00272-9

Kiiskinen U, Suominen-Taipale AL, Cairns JA (2010) Think twice before you book? Modelling the choice of public vs private dentist in a choice experiment. Health Econ 19(6):670–682. https://doi.org/10.1002/hec.1504

Koo M, Fishbach A (2010) A silver lining of standing in line: queuing increases value of products. J Mark Res 47(4):713–724. https://doi.org/10.1509/jmkr.47.4.713

Kremer M, Debo L (2016) Inferring quality from wait time. Manag Sci 62(10):3023–3038. https://doi.org/10.1287/mnsc.2015.2264

Lee YS, Seo YW, Siemsen E (2018) running behavioral operations experiments using amazon's mechanical turk. Prod Oper Manag 27(5):973–989. https://doi.org/10.1111/poms.12841

Liu N, van de Ven PM, Zhang Bo (2019) Managing appointment booking under customer choices. Manag Sci 65(9):4280–4298. https://doi.org/10.1287/mnsc.2018.3150

McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in economics. Academic Press, New York, pp 105–142

Mushtaque U, Pazour JA (2020) Random utility models with cardinality context effects for online subscription service platforms. J Revenue Pricing Manag 19(4):276–290. https://doi.org/10.1057/s41272-019-00227-0

Ostrom A, Iacobucci D (1995) Consumer trade-offs and the evaluation of services. J Mark 59(1):17–28. https://doi.org/10.2307/1252011

Paetz F (2021) Personality traits as drivers of social preferences: a mixed logit model application. J Bus Econ 91:303–332. https://doi.org/10.1007/s11573-020-01001-0

Paetz F, Hein M, Kurz P, Steiner W (2019) Latent class conjoint choice models: a guide for model selection, estimation, validation, and interpretation of results. Mark ZFP J Res Manag 41(4):3–20. https://doi.org/10.15358/0344-1369-2019-4-3

Sawtooth (2009) Fine-Tuning CBC and Adaptive CBC Questionnaires. https://content.sawtoothsoftware.com/assets/cf6f1953-db50-4174-8672-06bc04a0d5ef

Sawtooth (2021) The CBC HB System for Hierarchical Bayesian Estimation, Technical Paper V5.6. https://sawtoothsoftware.com/resources/technical-papers/cbc-hb-technical-paper

Smith L, Sorensen PN (2013) Rational social learning with random sampling. https://ssrn.com/abstract=1138095 or https://doi.org/10.2139/ssrn.1138095

Teraji S (2003) Herd behavior and the quality of opinions. J Socio-Econ 32(6):661–673. https://doi.org/10.1016/j.socec.2003.10.004

Train KE (2001) A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit. Working Paper, University of California Berkley

Train KE (2009) Discrete choice methods with simulation, 2nd edn. Cambrigde University Press, Cambrigde

Veeraraghavan S, Debo L (2009) Joining longer queues: information externalities in queue choice. Manuf Serv Oper Manag 11(4):543–562. https://doi.org/10.1287/msom.1080.0239

Veeraraghavan SK, Debo LG (2011) Herding in queues with waiting costs: rationality and regret. Manuf Serv Oper Manag 13(3):329–346. https://doi.org/10.1287/msom.1100.0322

Victoor A, Delnoij DM, Friele RD, Rademakers JJ (2012) Determinants of patient choice of healthcare providers: a scoping review. BMC Health Serv Res 12(1):1–16. https://doi.org/10.1186/1472-6963-12-272

Wu C-C, Liao S-H, Chen Y-J, Hsu W-L (2011) Service quality, brand image and price fairness impact on the customer satisfaction and loyalty. In: 2011 IEEE international conference on industrial engineering and engineering management, pp 1160–1164. https://doi.org/10.1109/IEEM.2011.6118097

Zwerina K, Huber J, Kuhfeld WF (1996) A general method for constructing efficient choice designs, White Paper SAS. http://support.sas.com/techsup/technote/mr2010e.pdf

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.