Control of Online-Appointment Systems when the Booking Status signals Quality of Service

[blinded for review]

Abstract

We revisit a service provider's problem to match supply and demand via an online appointment system such as a doctor in the health care sector. We identify in a survey that an extensive set of available appointments leads to significantly less demand because customers infer a lower quality of the service, as part of an observational learning process. We capture the quality inference effect in a multinomial logit framework and present a Markov decision process for solving the problem of releasing available slots of the appointment system to optimality aiming at maximizing the expected profits. We further evaluate several simple decision rules and provide management insights on which rule to apply under different generic scenarios. Different from current literature, offering all available appointments may lead to suboptimal results when accounting for the quality inference effect. The profit-maximizing strategy then is to offer a subset of the available appointments.

Keywords: Appointment System, Choice-Based Optimization, Discrete Choice, Stochastic Dynamic Programming, Markov

1. Introduction

Appointment systems are widely used in the service industries to match customers' demand and a service provider's capacity, e.g., in the health care sector, gastronomy, or leisure industry. "An appointment system has completely altered our lives; it has brought order out of chaos and we can never cease to wonder how we endured our old ways so long" (Cardew 1967). First, common practices started with pen and paper and were later replaced by computer-based appointment systems, which are nowadays connected to the internet with increasing pace (see Zhao et al. 2017 for an overview of web-based medical appointment systems). Well-known platforms to make an appointment online are, e.g., www.opentable.com (restaurants), www.zocdoc.com (doctors), and www.opencare.com (dentists). Online appointments seem to be an emerging trend (see U.S. Government Accountability Office 2017) and are especially beneficial in times of a pandemic like COVID-19[1] (resp. Sars-CoV-2). In times of social distancing, well matched supply and demand is important to avoid unnecessary gathering of people. Online appointment systems can decrease the number of customers lining up in a long queue not knowing whether they will be treated after a reasonable waiting time, especially after a reopening (e.g., after a pandemic lockdown). The pandemic situation has even pushed virtual health care forward. "This crisis has forced us to change how we deliver health care more in 20 days than we had in 20 years" (Dr. Robert McLean in Span 2020) and Dr. Meeta Shah conjectures "kind of a turning point for virtual health care" (Dr. Meeta Shah in Abelson 2020). In any case (virtual or physical service), offering the booking possibility online comes with the decision which appointments to offer in detail. Some providers ask the potential customers about their time preferences before offering appointment times. Irrespective of possible time preferences, the following question is raised: Should the provider solely offer one appointment time, e.g., the one that is closest to the preference (if known), several times to choose from, or even all available appointments?

Nowadays, customers looking for an unknown service provider (e.g., for a new dentist or an escape room) often read ratings on portals like www.tripadvisor.com or www.yelp.com. Shukla et al. (2020) show by investigating clickstream data on online word-of-mouth that the number of rated doctors can influence customers' choice behavior. Observing the service provider's free capacity may also be part of the opinion formation (observational learning from previous customer choices). The consequence may be a negative or a positive effect. On the one hand, a small offer set may be negatively associated with e.g., longer waiting times, leading to a turned away customer. On the other hand, a small offer set may be positively associated with popularity, leading to an increased interest in the provider. We see this behavior (positive association) as an analogy to the offline empty restaurant syndrome, where

[1] https://www.who.int/emergencies/diseases/novel-coronavirus-2019, last checked on 23/06/2020.

herd behavior may lead to a queue if there are sufficiently informed customers who know that the quality of the service is high (see Becker 1991; Kremer and Debo 2016; Teraji 2003). The general concept is that uninformed customers follow the behavior of (seemingly) informed customers. One queue gets longer while the competitors possibly stay with a lack of demand (as long as the queue is not oversized). In terms of an online appointment system, customers may infer the quality of the service when observing the booking status, i.e., the number of offered time slots and the number of slots that are not offered (or booked). In that vein, a small set of offered appointments in the online world can be put on a level with a queue in the offline world.

We investigate whether this empty restaurant effect is also present in the online world by setting up a survey. We use the context of making an online appointment with a dentist. Different from shortages of specific physicians (in several regions), some dentists face a performance pressure, leading to many overtreatments to increase revenues (Heath et al. 2020), which can harm the perceived quality and the customers' satisfaction and may turn away customers. We find in the survey that customers are less likely to choose a dentist if the number of offered time slots is relatively high. As an example, one of our survey's participants states: "I chose by seeing the number of free appointments as a sign of how good the dentist is. So I preferred the one, that looks more preferred by other people". To our knowledge, this is the first survey showing that customers infer quality from the booking status.

Based on this observation, we model an online appointment system as a Markov decision process, following Gupta and Wang (2008b). The service provider aims at maximizing expected profits by controlling the number of time slots offered to a customer and by accepting/denying appointment requests. Denying requests enables the service provider to reserve capacity for short-term customers (e.g., acute patients). A booking limit defines the number of slots that are to be booked for requesting customers (see Gupta and Wang 2008b). A discrete choice model is embedded to cover the customers' preferences (see Train 2009 for an overview). We assume that the customer's utility function includes the number of offered time slots, as we observe in the survey that customers infer inferior quality of the service if too many time slots are offered. We analyze the conditions under which the visibility of free time slots should be actively managed by the service provider, i.e., offering only a specific subset of the available appointments.

As the state space of the Markov decision process grows exponentially with the number of differently preferred time slots (slot types), we analyze three decision rules and test them against the optimal solution. We find in a numerical study that the decision rules' performances mainly depend on two factors, i.e., how strong the quality inference affects choice behavior and how strong the preference for a time slot varies throughout the day. The decision rule that performs best in all instances is solving the combinatorial problem myopically by finding the offer set that minimizes the likelihood that a customer leaves the system without booking at any given point in time. We show that this rule is optimal in case the customers have no preferences concerning the timing of the appointment throughout the day. In other cases, the performance of the rule is close to optimal, while we also identify simpler rules than the myopic with similar performances.

In sum, our main contribution lies in considering a new aspect of customers' choice behavior in online appointment systems. Including the possibility to offer only a subset of the available slots may improve the utilization and, thus, the expected profit of a service provider when customers infer quality from the booking status as in our survey. We propose rules and provide management insights on which policy seems most appropriate depending on the strength of the quality inference and the heterogeneity on time slot preferences.

The rest of this paper is structured as follows. In Section 2, we briefly review the relevant theoretical literature regarding related appointment systems and behavioral literature about the empty restaurant syndrome. In Section 3, we introduce the design and results of our survey on online appointments to investigate the empty restaurant syndrome in an online environment. In Section 4, we provide details about our theoretical model and explain the main properties in Section 5. Section 6 gives an overview of our numerical study, introduces our developed decision rules and shows the performance of our model and the decision rules. In Section 7, we critically discuss our approach and point out relevant limitations. Finally, we conclude our paper in Section 8.

2. Related Literature

In our paper, we study how time slots should be made available when customers infer quality from the number of available time slots in appointment systems. Revenue management, as an instrument for the allocation of service capacity among customers, has been extensively studied in the literature, especially airline revenue management (see McGill and van Ryzin 1999 for an overview). We focus on a setting, in which the service provider cannot differentiate the time slots by prices or other characteristics (contrary to, e.g., the standard two-fair class revenue management problem in Belobaba 1989). In contrast, we examine two customer groups, where prices do not necessarily have to be different, but our later customer group (short-term customers) must not be rejected, as it is common in the health sector (e.g., Gupta and Wang 2008b). See, e.g., DeCroix et al. (2020) for the consideration of dynamic personalized pricing when service variability reduces a provider's revenue.

Different decision levels are studied regarding appointment systems (see Ahmadi-Javid et al. 2017 for strategic, tactical and operational level). The strategic level primarily focuses on decisions such as how many servers to consider, the access policy, and how to deal with walk-ins in general. On the tactical level, the appointment design is to be determined, such as the interval, time window, and block size (Cayirli and Veral 2003), whereas the operational level, what we primarily focus on, refers to the individual customer level with fixed capacities. Patients are then either accepted and allocated to a day and time or rejected.

Approaches for multiple service providers in one appointment system exist (Gupta and Wang 2008b), whereas we focus on an appointment system for a single service provider for simplicity (and further show how to include competing providers in Appendix D). Liu et al. (2019) consider sequential offering, enabling interaction between the provider and the potential customer. This process is especially beneficial for telephone-based appointment scheduling. We focus on an online appointment system and exclude the possibility of iteratively offering different sets to one potential customer. Most of the literature considers intraday planning (e.g., Gupta and Wang 2008b; Talluri and van Ryzin 2004a), as we do. This means, we only consider and plan one workday separately from others. Only a few authors work on interday planning to consider several days of a week (e.g., Feldman et al. 2014, Wiesche et al. 2017, Zacharias et al. 2020). Some approaches for appointment systems, regarding the analytical methods on the operational level, make use of a queueing theory (see, e.g., Green 2006 and Zhou et al. 2021), while others, including our approach, use (stochastic) dynamic programming, like Gerchak et al. (1996), Green et al. (2006), Gupta and Wang (2008b), Feldman et al. (2014), and Truong (2015).

Concerning customer choices, discrete choice approaches (Train 2009) received great attention in the literature. Random utility models help consider individual preferences among alternatives. A considerable range of models has been developed and can be used for different contexts. Independent demand models, for example, are often found in airline revenue management studies (see Talluri and van Ryzin 2004b). Mackert (2019) considers dynamic slot management for profit-maximization in the context of attended home delivery using a general attraction model for customer choice behavior. We use a multinomial logit model (Ben-Akiva and Lerman 1985; McFadden 1974), which is a special case of the basic attraction model (Luce 1959), to take into account the size of the offer set (i.e., the number of offered appointments). Mushtaque and Pazour (2020) focus on the consideration set theory and study multinomial logit cardinality effect models to compare the benefits and costs of offering a specific subset on an entertainment subscription platform. In contrast to our approach, numerous customers can purchase the same service and customers are overwhelmed by too many information, wherefore a personalized subset is recommended. Customers do not infer quality from the offer set.

Several publications exist in which heterogeneous customer preferences are examined or modeled (e.g., Hole 2008, Liu et al. 2018, Liu et al. 2019), whereas we assume a homogenous customer group who requests service in advance. As demand is endogenous, our approach fits into the literature on choice-based optimization (see Haase and Müller 2013). Hole (2011) similarly models the decision on attributes in the choice endogenously, while other approaches (see for example Gupta and Wang 2008b, Liu et al. 2019) consider demand as exogenous.

Our work can be considered as an extension of Gupta and Wang (2008b). Additionally to the existing actions of either accepting or denying a slot request, we include an action that we call blocking. This action enables the service provider to actively manage the appointment offer set. Liu et al. (2019) also allow to hold back slot types. However, they block full slot types for later arriving customers, as heterogeneous customers are interested in specific slot types (binary choice model), and find several instances for which offering all slots is optimal. In contrast, we consider homogenous customers (requesting service during the booking horizon) and enable to block single slots of one slot type (set of slots that are equally interesting to the customers) to increase the customers' interest in the provider. Gupta and Wang (2008b) introduce a booking limit which indicates the optimal number of requests to be accepted, in other words, slots to be booked before the day of service (see Section 5.1 for more information). Thus, it is never optimal to reject a request as long as the booking limit is not reached. However, it may be worthwhile to limit the offer set.

Liu et al. (2019) find in a survey that patients have heterogeneous time windows preferences for an appointment with their primary care doctor. In our survey, we control these heterogeneous preferences by predefining the preferred appointment for the respondents and focus on the number of offered slots instead. We expect the customer's interest to decrease with a greater choice set in an online appointment system, as a positive effect from the theory of observational learning, meaning that customers adopt the behavior of previous customers. Thus, we expect to have a positive effect of observing the queue on the interest in the service provider. For further information on observational learning, see basic literature of Banerjee (1992) and Bikhchandani et al. (1992). The effect behind our expectation is also known as the empty restaurant syndrome, but rarely studied in the literature. A queue may be associated with quality by uninformed consumers (potential customers who do not know the provider) because they presume informed customers in the queue demanding service and knowing the provider (Kremer and Debo 2016). Thus, absent demand may evoke unpleasant associations (e.g., low quality of service). A service provider who offers many appointments at once may experience less demand if the effect is present. Therefore, it may be beneficial to not offer all available slots. Veeraraghavan and Debo (2009) for example include information about service quality and queue length in their model on customers' choice behavior.

They further investigate the impact of waiting cost in queues on the customers' behavior and find that customers behave according to the herd behavior as long as they do not want to minimize ex-post regret (Veeraraghavan and Debo 2011). Debo et al. (2012) consider a queueing system and assume that potential customers decide whether to buy the product after observing the queue. They conclude that a queue can be a signal of high quality. Özer and Zheng (2016) include the perceived probability that a product is available, which may impact the purchase behavior.

Empirical research finds the empty restaurant syndrome in offline scenarios. Koo and Fishbach (2010), for example, experimentally show that the presence and the length of a queue behind a person increase the value of a product. Giebelhausen et al. (2011) find in an experiment that waiting time can indicate quality and positively influences the purchase intention and experienced satisfaction. Kremer and Debo (2016) show in laboratory experiments that waiting times positively affect the uninformed consumers' purchase intention if informed consumers are present. Contrary, DeVries et al. (2018), e.g., find a negative impact of waiting time on the long-term customer behavior and revenue by analyzing the data collected from an Indian restaurant. Jin et al. (2015) observe both positive and negative effects for the choice between several locations and find that observational learning depends on the congestion level. In our survey, we focus on the examination of whether the empty restaurant syndrome may also occur in an online appointment system. In the following, we refer to it as the quality inference effect.

3. Survey

3.1. Design

We distributed the survey (Software: LimeSurvey) via a platform for students of the course [blinded for review] at the University of [blinded for review]. 248 undergraduates completed the survey in January 2020 (14 January until 23 January). As an incentive, subjects received bonus points for the exam if they completed the survey.

Our survey is organized into three parts. The first part includes warm-up questions regarding the subjects' internet usage. The second and main part consists of three scenarios in which the subjects choose between two service providers (see an exemplary scenario in Figure 3 in Appendix A). In each scenario, we ask participants to imagine the need for a yearly routine appointment at the dentist without having any toothaches, whereby the preferred time slot is at 8 a.m. With a complete appointment system of 32 slots (15-minute cycle from 8 a.m. to 4 p.m., known to the subjects), each dentist offers either a small (two slots), a moderate (eight slots) or a big set (32 slots) of appointments. The smaller offer set is always a subset of the bigger offer set and the preferred time slot (8 a.m.) is always offered by both providers. The dates and time until the day of service are excluded. The present (booking) day and the day of service do not play a role in the survey.

We conducted six treatments in a between-subject design (see Table 1). In each treatment, we test three scenarios in a within-subject design. As an example, in treatment T1a, subjects first choose between a dentist with 2 available slots and another dentist with 8 available slots, and in the second choice situation, between 2 and 32 slots, and so on. To check if the position of the choice option on the screen affects choice behavior, we have a second treatment T1b that is almost identical to treatment T1a but reverses the order of free slots. That is, in treatment T1a, the first choice is between 2 and 8, and in treatment T1b it is 8 vs. 2. Furthermore, we randomize in each treatment and each scenario the order of answers in the choice list (provider A and provider B).

Table 1: Treatment summary

Treatment		# of obs.	Within-subject: First choice	Within-subject: Second choice	Within-subject: Third choice	First independent choice matched data
T1	a:	46	2 vs. 8	2 vs. 32	8 vs. 32	2 vs. 8
	b:	47	8 vs. 2	32 vs. 2	32 vs. 8	
T2	a:	42	2 vs. 32	2 vs. 8	8 vs. 32	2 vs. 32
	b:	42	32 vs. 2	8 vs. 2	32 vs. 8	
T3	a:	41	8 vs. 32	8 vs. 2	2 vs. 32	8 vs. 32
	b:	30	32 vs. 8	2 vs. 8	2 vs. 32	

In the third and last part of the survey, we ask open questions to understand the subjects' decisions in the preceding scenarios. We integrated several attention checks making sure that subjects answer conscientiously. Note that we do not exclude those participants in our analysis who did not provide the correct answers. Only two subjects "failed" all attention checks. We did not find any critical contradictions when excluding those two subjects in our analysis.[2]

Our hypothesis follows the idea of the empty restaurant syndrome, i.e., a greater offer set may be associated with low quality and vice versa.

Hypothesis: A provider with a smaller offer set is preferred to a provider with a greater offer set.

[2] The two subjects marked "disagree" when they were asked to mark "strongly disagree" on a 5-point Likert scale. Regarding the answers to the remaining questions of the survey given by those two subjects, we could not find any specific pattern (e.g., only the first answer marked). Significant results from the data set without any attention check failure (197 respondents) are similar to the results given in Section 3.2.

3.2. Survey Results

We first analyze the choice frequencies in the first scenario of each treatment, see Table 2. We consider the (a) and (b) treatments jointly since we observe no significant order effects (fisher's exact test, $\alpha > 0.05$).

Table 2: Descriptive statistics of the first choice

	2 vs. 8 slots (T1)	2 vs. 32 slots (T2)	8 vs 32 slots (T3)
Choice frequency	47% vs. 53%	57% vs. 43%	75% vs. 25%
# observations	93	84	71

Having to choose between the moderate and the big offer set (T3), 75% choose the option with less available slots. The narrow majority chooses the option with fewer available slots when opting between the small and the big offer set (T2). Slightly less than one-half chooses the small offer set compared to the moderate offer set (T1). Choice by chance (i.e., no effect of the number of offered slots) would predict a 50/50 split, which we could not reject in T1 and T2 (binomial test, $\alpha > 0.1$), but which is clearly rejected in T3 ($\alpha < 0.001$). Choosing the provider with eight slots differs highly significantly with the given alternative (fisher's exact test, $\alpha < 0.01$).

3.3. Choice Motives

A too-big offer set leads to less demand in treatment T3. This effect is in line with our hypothesis. However, the choice behavior of preferring a smaller offer set can be explained by different motives besides our considered quality inference effect, e.g., by choice overload or by simply being easier to find the predefined preferred appointment at 8 a.m. if less slots are displayed. Therefore, we have a closer look at how the respondents explained their choices in an open question after the three choices ("What aspect(s) did you consider when deciding which of the providers to choose in the scenarios?"). We let two independent raters code the open question on a binary scale whether a motive is mentioned by a respondent or not (yes: 1, no: 0). Besides the frequency of the mentioned motives, the Cohen's kappa coefficient (Byrt et al. 1993) is stated for the inter-rater reliability. We provide five explanations detected from the data set: quality, choice overload, flexibility, scarcity and less waiting time.[3]

In Table 3, we show the raters' mean percentage (geometric mean) of respondents in T3 (8 vs. 32, significant results) who name the respective motives and the Cohen's kappa for the inter-rater

[3] Further explanations that are sporadically mentioned, such as a misunderstood offer set and single entries are aggregated to the category others and no longer considered in this analysis (e.g., "safety" respond. 67, "Breaks […]" respond. 242, "Intuitively" respond. 290).

reliability. For the sake of completeness, we provide the frequency of the motives of all respondents in Table 7 in Appendix B. However, those results do not deviate significantly.

Table 3: Frequency of the motives for the choice behavior in T3 (71 respondents)

	Quality	Choice overload	Flexibility	Scarcity	Less wait time	Other
Frequency in %	42.95	4.88	38.03	10.35	9.13	23.78
(Cohen's kappa)	(0.9713)	(0.5497)	(0.9402)	(0.6290)	(0.9154)	(0.6920)

Notes. The frequency is the geometric mean of the raters' results. Multiple answers possible, which leads to a sum of the frequencies >100%.

The predominant explanation of the behavior is the quality motive, which is mentioned by 43% of the respondents in treatment T3 and in line with our hypothesis. Customers expect the service provider with a smaller offer set to be more demanded and thus more popular, which leads to the perception that it may be the better dentist. “Less appointments may mean that the dentist is more popular among the clients and so better.” (respond. 225) Further, a too-big offer set seems to be suspicious due to absent demand. “[H]ow many appointment choices were available because I think that too many options show that the doctor doesn't have many patients, so maybe he's not that good in what he's doing” (respond. 110). Summarizing, a smaller offer set is mostly associated with higher quality, while a too-big offer set is negatively associated with lower quality. Customers thus rather choose a service provider with a smaller offer set.

An alternative explanation to the quality motive for a preferred smaller offer set is choice overload (see Eppler and Mengis 2004 and Scheibehenne et al. 2010 for a literature review). Only 5% give an overload-related motive. “It should be clearly structured and not too overloaded” (respond. 283). This motive is considered explicitly even though it has only single entries and a poor Cohen's kappa.

An explanation for opting the larger offer set is the “flexibility”, and given by 38% of the respondents in T3. This motive is also found by Rubin et al. (2006) as a driver. They investigate waiting times and choice of time and doctor in a discrete choice experiment and find that, e.g., employees are willing to wait longer in order to get their preferred time. Even with a predetermined preference of the 8 a.m. appointment in our survey, customers like to have the flexibility and greater availability. “The more available possibilities[,] the more freedom I have […]” (respond. 245). Further, some subjects considered the possibility that their preferred 8 a.m. appointment would not take place. They imagined being late or shifted by the provider and preferred flexibility for the postponed appointment. “[…] possibilities to change to a[n] appointment a bit later in case that the appointment at 8am can[]not take place […]” (respond. 138). Summarizing, a greater offer set gives more flexibility and may thus be positively associated. However, we want to mention that some subjects only mentioned "availability", which is not always a clear motive for a preferred greater offer set. By simply stating "availability", it is not clear whether the respondent prefers more options or also infers quality from a smaller availability. Further note that the open question is given for all three choices, even though we focus on the first choice of treatment T3 (8 vs. 32 slots). Thus, the explanation could also be a reason for a different choice behavior (of the second or third choice, e.g., 2 vs. 32) and does not need to be contradictory.

Two less mentioned but clearer motives are scarcity (10%), see for example Denier (2008), and less waiting time (9%), see for example Rubin et al. (2006). If a service provider offers too few slots, potential customers expect the provider to be in a rush and/or the waiting room to be crowded, which leads to a longer wait time. "[…] I don't want to visit a dentist, who is super stressed […]" (respond. 10). Potential customers may have negative associations with a scarce offer set (i.e., very few offered slots) and may rather choose another provider with more available slots. We included the two motives in our analysis as alternatives to "flexibility" despite the lower frequency and the poorer Cohen's kappa for scarcity.

The remaining motives (24%) are single entries (e.g., safety, breaks). Note that the open question gives the possibility to mention several aspects. Thus, it sums up to more than 100%.

Summarizing, the open-ended question indicates that customers mainly infer quality from a smaller offer set. In contrast, a too-small offer set may also lead to less demand, although these results are not significant and may need further research. A scarce offer set may induce concerns such as longer waiting times and too little flexibility. We refer to this effect as the scarcity effect. This does not solely mean shortage-associations, but more aspects such as flexibility.

3.4. Strength of the Quality Inference Effect

We next estimate the effect of the number of displayed slots using a multinomial logit model in R (version 4.0.2) with the utility function specified in equation (1). The number of offered slots are denoted by σ. We use a relative formulation in which $\frac{\sigma-a}{b-a}$, $b > a > 0$, becomes 1 if all slots are offered (upper bound b) and 0 if the number of offered slots is equal to the lower bound a which is to be specified. We set the alternative specific constant to zero, $asc = 0$, as no differences between the alternatives are given in our survey, apart from the number of slots that are offered. We thus cannot find any alternative specific anchor that requires the integration of an $asc \neq 0$. The upper bound is $b = 32$ and the lower bound $a = 2$ (the lowest offer set we consider in our survey, thus $b \geq \sigma \geq a$ in our setting). The $\hat{\beta}$-coefficient measures the strength of the effect relative to asc and the error term ϵ.

$$U = asc - \hat{\beta} \cdot \frac{\sigma - a}{b - a} + \epsilon \tag{1}$$

We present our analysis regarding the first decision of each subject (248 observations) in order to account for the dependencies of the within-subject variations. We receive similar results when running the analysis on the whole data set, see Table 8 and Table 9 in Appendix C.

The overall estimated $\hat{\beta}^{all}$-value is 0.6007 (σ=0.18, t-ratio= -3.34), indicating a slight but significant negative tendency of the utility for an increasing offer set at the 95% confidence level, which confirms our hypothesis on an aggregated level. The more slots are offered, the less attractive the provider gets.

We also estimate $\hat{\beta}^{i}$ for the three treatments separately, $i \in \{T1, T2, T3\}$. Contrary to our hypothesis, we estimate $\hat{\beta}^{T1} = -0.5382$ ($\sigma = 1.0385, \text{t-ratio} = 0.52$), indicating a non-significant increase of interest in a provider with an increasing offer set (2 vs. 8). Note that we model a negative impact of the size of the offer set on the interest in the provider, see Equation (1). With a negative $\hat{\beta}$-estimation, it results in a positive impact of the effect on the utility. For T2 (2 vs. 32), we get a non-significant estimate of $\hat{\beta}^{T2}$ =0.2877 ($\sigma = 0.2205, \text{t-ratio} = -1.3$). For T3 (8 vs. 32), we estimate a significant effect of $\hat{\beta}^{T3} = 1.35$ ($\sigma = 0.341, \text{t-ratio} = -3.96$).

In sum, we find evidence that customers primarily infer quality from the booking status. We assume that in realistic scenarios, customers go through a search process when looking for a service provider. Within such a process, customers observe one offer set after another of different providers and decides each time whether to book an appointment with that provider or search for an alternative. The process ends with an appointment request. In our stochastic dynamic program, we consider one of those service providers and include the positive externalities of the search process. We further assume that the scarcity problem is out of the service provider's control and focus on those situations where $\hat{\beta} > 0$ by setting the lower bound a such that the quality-effect clearly exceeds the scarcity-effect. Due to the scarcity effect, this quality inference effect only affects choice behavior if the number of offered slots is sufficiently large. We next introduce our stochastic dynamic program with the integrated discrete choice model that accounts for customers' quality inferences from the booking status and solely focuses on one provider (disregarding competitors).

4. Model Formulation

We focus on an appointment system for one workday of a single service provider. In Appendix D, we further show how we derive the customers' choice behavior regarding the considered provider when several competing providers are observed (e.g., two providers in our survey, see Section 3). The service provider faces two homogenous groups of uncertain customer demand. The first group requests slots in advance during the booking horizon. We hereby focus on customers who have not booked the provider's service yet and are thus not familiar with the provider's quality. The second group only contains customers who place their requests on the day of service, so-called same-day requests. We formulate our problem as a discrete-time, finite-horizon Markov decision process, following Gupta and Wang (2008b). To better cope with the curse of dimensionality, we adopt the formulation of Liu et al. (2019) for our state structure.

Appointment system & action space

The service provider has a fixed capacity of κ slots on the workday. As the customer group who asks for service in advance may prefer some appointment times to others, we consider N different slot types, with $n = 1 \dots N$. Slots within one slot type are equally attractive to the customers; slots of different slot types are differently attractive. Appointments from slot type $n = 1$ are the most preferred, followed by appointments from slot type $n = 2$, and so on. The fixed capacity of slot type n is denoted by κ_n, with $\kappa = \sum_{n \in N} \kappa_n$. Each slot belongs to one unique slot type n and can be booked for at most one customer. The booking status of the workday is denoted by $\vec{s} = (d_1, \dots, d_N, q_1, \dots, q_N)$, whereby $d_n \in \{0,1, \dots, \kappa_n\}$ states how many slots of slot type n are already booked for a customer, with $\vec{d} = (d_1, \dots, d_N)$. It further states by $q_n \in \{0,1, \dots, \kappa_n - d_n\}$ how many slots of slot type n are temporarily blocked, with $\vec{q} = (q_1, \dots, q_N)$.[4] A blocked slot is neither booked for a customer nor offered to the customer but held back. If an offered slot of type n is booked for a customer, d_n is increased by 1. A potential customer cannot see whether a slot is blocked or booked, but only if offered. $\vec{o} = (o_1, \dots, o_N)$ states the resulting number of offered slots per slot type, with $d_n + q_n + o_n = \kappa_n$ for all $n \in N$ and $\sigma = \sum_{n \in N} o_n = \sum_{n \in N} (\kappa_n - d_n - q_n)$ the size of the overall offer set. $\vec{u} = (u_1, \dots, u_N)$ further states the number of unbooked slots per type, with $d_n + u_n = \kappa_n$ for all $n \in N$, $\mathfrak{u} = \sum_{n \in N} u_n = \sum_{n \in N} (\kappa_n - d_n)$ and $0 \leq o_n \leq u_n$ for all $n \in N$.

The booking horizon starts with opening the slots for requests, ends with the beginning of the planned workday, and is divided into τ discrete time periods, with $t = 1, \dots, \tau$. Time is counted backwards, thus, we denote the planned workday by $t = 0$. In each period $t \geq 1$, the service provider has to decide which of the yet unbooked slots per slot type to offer and whether or not to accept a request if present. At most, one request occurs per period and only a slot of a slot type n with $o_n > 0$ can be

[4] Another formulation appears in the literature in which the booking status is denoted by the vector $\vec{s} = (s_1, s_2, \dots, s_\kappa)$, where $s_j = 0$ if slot *j* is offered to the potential customer (regardless of the slot type), $s_j = 1$ if slot *j* is already booked for a customer and $s_j = 2$ if slot *j* is temporarily blocked. That formulation leads to 3^κ possible booking status. Our formulation outperforms this formulation in most cases as $\sum_{x_1=0}^{\kappa_1} [\dots] \sum_{x_N=0}^{\kappa_N} (\kappa_1 + 1 - x_1) \cdot [\dots] \cdot (\kappa_N + 1 - x_N) \leq 3^\kappa$.

requested. The service provider aims at an optimal capacity utilization via controlling the appointment system's booking status to maximize her expected profit.

Thus, in each period, two decisions have to be made (actions). (1) With a present request, the decision is to either accept or deny the request. (2) The second action is to decide how many slots per slot type to offer resp. block in the subsequent period. Each available slot can be offered or blocked.

Customer group 1: Planning requests and booking status

In each period $t \geq 1$, there is at most one potential customer with an independent arrival rate $0 < \alpha_t \leq 1$. The customers' demand for time slots of any slot type is random and follows a multinomial logit choice rule provided that a customer arrives in the respective period. $\mathcal{P}_n(\vec{o})$ is the conditional probability that any slot of slot type n is requested given offer set $\vec{o}$. $\mathcal{P}_0(\vec{o})$ denotes the probability that a customer does not request any of the offered slots ("no choice"). $\mathcal{P}_n(\vec{o}) = 0$, if no slot of slot type n is offered. For this specification, $\sum_{n \in N} \mathcal{P}_n(\vec{o}) + \mathcal{P}_0(\vec{o}) = 1$ holds true.

Equation (2a) states the request probability $\mathcal{P}_n(\vec{o})$ for any slot of slot type n given the offer set $\vec{o}$, assuming that the unobserved attributes of slot type n, denoted by ϵ_n, with the utility $\mathcal{U}_n = w_n - \beta \cdot \left(\frac{\sigma - a}{b - a}\right)^+ + \epsilon_n$, follow a Gumbel distribution (Train 2009). The weight of the slots of slot type n, w_n, captures its mean utility, also known as the alternative specific constant. We assume the weights to be exogenous. A survey similar to Liu et al. (2019) could help gather further information on time preferences. In order to capture quality inferences from the booking status (i.e., the number of offered slots, see Section 3), we assume that the size of the offer set σ has a negative impact on the choice probability of all offered slots. Thus, the observable attribute of each slots of slot type n consists of the weight w_n and the size of the offer set (number of offered slots) in a relative formulation, multiplied by the parameter β, which indicates the strength of the effect. As already mentioned in Section 3.4, we have a lower bound $a \geq 0$, which indicates the minimum number of slots to be offered for the quality inference effect (exceeding the scarcity effect), and an upper bound $b \leq \kappa$, with $a \leq b$. The more slots are offered for $\sigma \geq a$, the less attractive the offered slots of all slot types become, if $\beta > 0$. The system size (capacity) can be seen as a natural upper bound for the respective provider. As the increase in the number of offered slots by one might at some point no longer increase the effect significantly, the upper bound can also be lower than the overall capacity. Without the lower bound ($a = 0$), we would already evoke the quality inference effect with the smallest possible offer set of one appointment, $\sigma = 1$. With $b = a$, we would eliminate the considered effect. An upper bound of $b = \kappa$ in combination with an empty schedule and all appointments offered ($\sigma = \kappa$) would lead to a full impact of the effect. $(\cdot)^+$ is short for $\max\{\cdot, 0\}$ and formalizes that we only consider the negative quality inference effect on slot requests. A negative term $\left(\frac{\sigma-a}{b-a}\right)$ would result in a scarcity effect, similarly to $\beta < 0$, which is briefly discussed in Sections 3.4 and 7 and not considered in our analysis. As each slot of slot type n has the utility $\mathcal{U}_n$, we aggregate it to the slot type by multiplying each exponentiated utility of slot type n by the number of offered slots o_n within the slot type.

$$\mathcal{P}_n(\vec{o}) = \begin{cases} \dfrac{o_n \cdot e^{w_n - \beta\cdot\left(\frac{\sigma-a}{b-a}\right)^+}}{\sum_{m\in N} o_m \cdot e^{w_m - \beta\cdot\left(\frac{\sigma-a}{b-a}\right)^+} + e^{w_0}} & \text{if } o_n > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2a}$$

We finally simplify (2a) by capturing the quality inference effect in the no-choice option, instead of subtracting it from each slot type option, see equation (2b).

$$\mathcal{P}_n(\vec{o}) = \begin{cases} \dfrac{o_n \cdot e^{w_n}}{\sum_{m\in N} o_m \cdot e^{w_m} + e^{w_0 + \beta\cdot\left(\frac{\sigma-a}{b-a}\right)^+}} & \text{if } o_n > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2b}$$

A customer request can either be accepted or denied to reserve capacity for the second customer group. Accepting a request, which means booking any slot (regardless of the slot type), gives a revenue of r_p (for the planned customer). Denying a request (rather theoretical, see Section 5.1 for practical application) and having a potential customer who chooses the no-choice option both come with a penalty π as both occurrences lead to a loss of a potential customer. Neither conversation (interaction) with the service provider nor multiple requests per potential customer are considered.

Customer group 2: Same-day requests

Short-term customers do not ask for a slot of a specific slot type but for service at any time of the day (e.g., acute patients in health care). Our consideration of same-day requests is in line with Gupta and Wang (2008b). A short-term customer must not be denied and generates a revenue of r_s. We model short-term customers Poisson-distributed, $X \sim \mathrm{P}(\lambda)$, with the parameter λ. X denotes a random number of short-term customers who appear at the beginning of $t = 0$, with F_X the cumulative distribution function (c.d.f.) of X. The service provider has to work overtime which causes overtime cost of c_o per same-day customer if there are not sufficient time slots available. In contrast, each unused slot causes underage cost of c_u.

Stochastic dynamic program

Let $v_t(\vec{s})$ be the maximum expected profit from t onwards given the booking status $\vec{s}$, see equation (3), and $v_0(\vec{s})$ the expected provider's profit from short-term customers, see equation (4). The problem is solved recursively with the overall maximized expected profit stated in $v_t(\vec{s})$.

$$v_t(\vec{s}) = \alpha_t \cdot \left(\sum_{n \in N} \mathcal{P}_n(\vec{o}) \cdot \max\left\{ r_p + \max_{\vec{g} \in G}\{v_{t-1}(\vec{s} + \vec{e} + \vec{g})\}, -\pi + \max_{\vec{g} \in G}\{v_{t-1}(\vec{s} + \vec{g})\} \right\} + \mathcal{P}_0(\vec{o}) \cdot \left(-\pi + \max_{\vec{g} \in G}\{v_{t-1}(\vec{s} + \vec{g})\} \right) \right) + (1 - \alpha_t) \cdot \max_{\vec{g} \in G}\{v_{t-1}(\vec{s} + \vec{g})\} \tag{3}$$

$$v_0(\vec{s}) = E[r_s \cdot \min\{\sigma, X\} - c_o \cdot (X - \sigma)^+ - c_u \cdot (\sigma - X)^+] \tag{4}$$

Vector $\vec{e} = (e_1, \dots, e_N, e_{N+1}, \dots, e_{2N}) = (e_1, \dots, e_N, 0, \dots, 0)$ tracks a requested and accepted slot of slot type n. If a slot of slot type n is requested and accepted, component n is one while all other components are zero. Thus, $\vec{e}$ states the change in $\vec{d}$. Vector $\vec{g} = (g_1, \dots, g_N, g_{N+1}, \dots, g_{2N})$ $= (0, \dots, 0, g_{N+1}, \dots, g_{2N})$ stores information about how many slots are blocked or unblocked per slot type in the following period, with $G(\vec{u})$ the set of all possible blocking options. Component $(N + n) \in (-q_n, \dots 0, \dots, o_n)$, e.g., is (-1) if one less slot of slot type n is blocked in the subsequent period, (+1) if one more slot of slot type n is blocked, and (0) if the amount of blocked slots of slot type n remains unchanged. Consequently, $\vec{g}$ states the change in $\vec{q}$. The operator max{·,·} ensures that, in each period, the actions (accept/deny request and offer/block slots) are taken that maximize the overall expected profit.

5. Properties

5.1. Booking Limit

Gupta and Wang (2008b) show that accepting all requests up to a booking limit b^*, i.e., a specific number of slots booked, regardless of the slot type, maximizes profits (see equation (5)). The optimal booking limit in our context is identical to the booking limit in Gupta and Wang (2008b) (see Appendix E) and defined by

$$b^* = \left[\kappa - F^{-1}\left(\frac{r_s + c_o - r_p - \pi}{r_s + c_u + c_o} \right) \right]^+ . \tag{5}$$

Intuitively, the aim of blocking slots is to trigger demand by signaling high quality. When the booking limit is reached, no further requests of the first customer group are desired, as the provider will save the remaining capacity for the second customer group and will therefore deny all further requests of the first customer group. Thus, it is no longer beneficial to block slots. However, in practice, the available slots may then all be blocked to make it impossible for customers to request any of the unbooked slots.

Note that it only matters how many requests to accept to reach b^* but not which specific slots. This is because slots are homogenous from the service provider's perspective, i.e., each slot yields identical revenues.

5.2. Threshold Value

We found no general static rule for the blocking action. The optimal policy on whether and which slots per slot type to block is state-dependent. We next show that blocking is beneficial if the impact of the quality inference effect is beyond a threshold value (i.e., if the effect is sufficiently strong).

Consider two (remaining) unbooked slots of different slot types and two (remaining) periods. We define $\Delta V(\vec{s})$ as the difference of the overall expected profit with the blocking of one slot in period 1 compared to the overall expected profit without blocking, see equation (6). If $\Delta V(\vec{s}) > 0$, it is beneficial to block at least one slot temporarily. We show in Appendix F that $\Delta V(\vec{s}) > 0$ always holds if β exceeds the threshold value $\tilde{\beta}$.

$$\begin{aligned} \Delta V(\vec{s}) = {} & E[V(\vec{s}|\vec{s} = (d_1, d_2, q_1, q_2), d_n \in \{0,1\}, q_1 = 1 \text{ for } t = 1, q_n = 0 \text{ otherwise})] \\ & - E[V(\vec{s}|\vec{s} = (d_1, d_2, q_1, q_2), d_n \in \{0,1\}, q_n \in \{0\})] \end{aligned} \quad (6)$$

We analyze in Appendix F the properties of $\Delta V(\vec{s})$ analytically for $\mathfrak{u} = 2$ and $b^* = \kappa$, $c_u = c_o = \lambda = 0$ and $\alpha_1 = \alpha_2 = 1$. $\Delta \mathrm{V}(\vec{s}) \leq 0$ holds if quality is not inferred from the booking status, i.e., $\beta = 0$. Intuitively, blocking increases the cumulated no-choice probability in both periods leading to lower expected profits. From $\Delta \mathrm{V}(\vec{s}) > 0$, if $\beta \to +\infty$, it follows that blocking is beneficial if β is sufficiently large. Increasing the costs (c_u and c_o) or the expected number of same-day requests λ may shift the booking limit b^* and the expected profit but does not impact the existence of the threshold value $\tilde{\beta}$ beyond which blocking becomes beneficial.

In case more slots are available ($\mathfrak{u} > 2$ with $\kappa - \mathfrak{u} < b^*$), we refer to our stochastic dynamic program in Section 4 and the related results in Section 6.2. We further note that the booking system will reach $\mathfrak{u} = 2$ in expectation if sufficient periods are left and enough potential customers request slots. Further, if only one slot is left unbooked ($\mathfrak{u} = 1$ with $b^* = \kappa$), no more blocking is advisable.

6. Numerical Study

6.1. Setup

We first analyze in which scenarios blocking is beneficial when making optimal decisions (see Section 6.2). Since the state space of the Markov decision process grows exponentially in the number of slot types, we further test three decision rules in Section 6.3.

We let the overall capacity vary from three to six ($\kappa = 3, \dots, 6$), with a fixed number of two slot types. The time horizon ($\hat{\tau}$, incl. the day of service) is kept proportional to the system size with 200% and 300% (e.g., for $\kappa = 6$, we set $\hat{\tau} = 12, 18$). Similarly, we keep the expected number of short-term customers proportional to the system size with 50% and 75% (e.g., for $\kappa = 6$, we set $\lambda = 3, 4.5$).

Without loss of generality, we assume a revenue of $r_p = 1$ for accepting a request during the booking horizon ($t \geq 1$). As both customer groups can generate different revenues, we fix the revenue for accepted requests during the booking horizon r_p to 1 while varying the revenue for same-day requests r_s such that we generate the relations $r_p > r_s$, by setting $r_s = 0.5$, $r_p = r_s$ by setting $r_s = 1$ and $r_p < r_s$ by setting $r_s = 2$. We set the cost for denying a request in $t \geq 1$ to $\pi \in \{0.5, 1, 2\}$. The costs for unused slots c_u are kept to 1 while varying overtime costs relatively, i.e., $c_o \in \{0.5, 1, 2\}$. In addition, we vary the coefficient of the quality inference effect, $\beta \in \{0, 0.5, 1, 1.5, 3, 6\}$. Besides the value of zero, which means that there is no impact, we test a value for a small impact (0.5), a small and a moderate value around the value of the survey (1 and 1.5), another moderate impact (3), and a high impact (6). The constant arrival rate of a potential customer varies from rather small ($\alpha = 20\%$), moderate ($\alpha = 50\%$) to rather high ($\alpha = 80\%$).

Our survey results show that customers only infer quality from the booking status if the number of offered slots exceeds a certain value a (i.e., the lower bound in equation (2b)). We set the lower bound relative to the number of available slots whereas imposing a lower bound of one, i.e., $a = \max\{0.25 \cdot \kappa; 1\}$. Further, we consider two slot types. Slot type $n = 1$ contains all preferred slots, whereas slot type $n = 2$ contains the less preferred slots. We refer to the weights of the two slot types as w_h for highly preferred and w_l for less preferred instead of w_1 and w_2. The weights of the preferred slots (of slot type 1) are varied, $w_h \in \{1,2,5\}$, while keeping the weights of the less preferred slots (of slot type 2), w_l, equal to 1. This results in instances where slots are homogenous to the customer ($w_h = w_l$), in other words, where we have one single slot type, and instances with heterogeneous slot preferences ($w_h > w_l$). Note that w_0 is normalized to 0. Besides varying the overall capacity and the weights, we build different ratios of the capacities of the two slot types. The assignment of the slots to the two slot types follows a typical time-of-day preference distribution, in which appointments in the morning, at noon and after work are preferred to appointments in the forenoon and in the afternoon. Accordingly, we get $\kappa_1 = 3$ and $\kappa_2 = 2$ for a system with $\kappa = 5$ appointments. The capacities per slot type of all considered system sizes are given in Table 4.

Table 4: Capacity of high-weighted slots (w_h) vs. low-weighted slots (w_l), in absolute values

κ	3	4	5	6
Ratio $w_h : w_l$	2 : 1	2 : 2	3 : 2	4 : 2

All parameters are examined in a full factorial design and we consider, in total, 23,328 instances. The stochastic dynamic programming (sdp) and the baseline without blocking were implemented in Matlab R2017a for the numerical study.

6.2. Results: Stochastic Dynamic Program

Note that expected profits may turn negative in our numerical study since we varied the cost and revenue parameters relative to others. As such, low-revenue parameters meet high-cost parameters in the full factorial design. Since it is unlikely that firms may economically survive in such scenarios, we focus on the instances with positive values (12,604 instances) and report the analysis of all instances in Table 10 in Appendix G.

Table 5 reports the most interesting extract of the percentage improvement when the optimal blocking strategy is followed (compared to a baseline without blocking). We find that blocking becomes more beneficial with a greater impact of the quality inference effect and mostly when customers differentiate between appointment times (two slot types) without a large discrepancy between the slot weights.

Table 5: Percentage improvement ($+\Delta\%$) of our sdp compared to the baseline without blocking varying β, slot weight (w_h) and reward for short-term customers (r_s) for all positive instances

Parameter	Overall	$\beta = 0$	$\beta = 0.5$	$\beta = 1$	$\beta = 1.5$	$\beta = 3$	$\beta = 6$
Overall	8.65	0.00	0.03	0.64	2.99	21.93	72.22
$w_h = w_l = 1$	10.52	0.00	0.00	0.46	4.46	45.72	115.95
$w_h = 2, w_l = 1$	10.77	0.00	0.04	1.32	4.62	34.01	126.51
$w_h = 5, w_l = 1$	5.58	0.00	0.05	0.17	0.41	3.02	45.31
$r_s = 0.5$	8.38	0.00	0.06	1.33	6.34	36.30	115.83
$r_s = 1$	9.57	0.00	0.04	0.88	4.21	33.36	106.03
$r_s = 2$	8.32	0.00	0.02	0.01	1.60	15.40	62.79

Our comparative static analysis of our positive instances further reveals that blocking overall becomes more beneficial when increasing β or increasing the penalty for denying a request. In turn, we observe that blocking becomes less beneficial when increasing the number of slots, the overtime cost, the arrival rate (probability for a potential customer), the expected number of short-term requests or the number of periods. We observe a non-monotonic effect of the revenue for the short-term requests and the weight of the more interesting slots.

Figure 1 (left side) shows an exemplary development of the overall expected profit for an increasing β-value, $0 \leq \beta \leq 8$, considering an appointment system with four slots. The expanded range for the

β-value is solely to additionally show the convergences. Further, Figure 1 (right side) shows the respective development of the no-choice probability. We observe that not blocking any slot (dotted line, left side) leads to a negative gradient of expected profits. Intuitively, if the quality inference effect becomes stronger, the customers' likelihood of not choosing an appointment (no-choice) increases (right side, dotted line). Blocking slots is an effective countermeasure. It keeps expected profits at a steady level (left side, dashed line) by also keeping the no-choice probability at a steady level (right side, dashed line). Overall, we observe an s-shaped effect on expected profits of blocking slots vs. no blocking (left side, solid line).

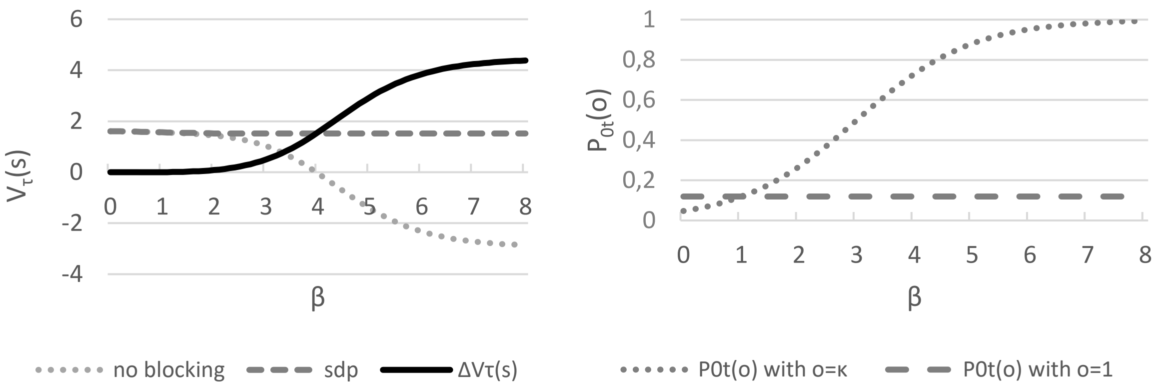

Figure 1: Development of the overall expected profit (left) and no-choice probability (right), with $\kappa = 4, \hat{\tau} = 8, r_p = 1, r_s = 1, c_u = 1, c_o = 0.5, \lambda = 3, w_h = 2, \alpha = 0.8, a = 1$

We conclude that with a noticeable quality inference effect ($\beta > \hat{\beta} > 0$) and without blocking, the willingness of uninformed potential customers to visit the service provider tends to be low if the schedule is (almost) empty. With the lack of demand, uninformed potential customers in later periods also face an empty schedule, again leading to absent demand. If informed customers do not book a slot, the schedule stays empty. A possible countermeasure against the phenomenon is the blocking of specific slots to trigger demand.

To give applicable decision-support and to cope with an upcoming curse of dimensionality with an increasing number of slot types, we developed decision rules.

6.3. Decision Rules

6.3.1. Definition

We present three rules and test their performance in a numerical study. For the comparison, we use the same setup as described in Section 6.1.

We define two simple to implement rules, while the third is more complex. The two simple decision rules both rely on one assumption: When blocking a slot, the multinomial logit framework adds the choice probability of this slot proportionally to the remaining options ("a rising tide lifts all boats", see Kennedy 1963). The more (less) preferred the blocked slots, the higher (lower) the portion added to remaining options, and, most importantly, to the outside-option (no-choice option). Thus, if a preferred slot is blocked, the no-choice probability increases more than if blocking a less preferred slot. Therefore, the rules focus on the blocking of less preferred slots rather than of more interesting slots (in case of several slot types).

(1) $\boldsymbol{R_{low}}$ blocks only less preferred slots of the booking system. In other words, only slots from slot type 2, with a lower slot weight, are blocked. All slots of slot type 1, i.e., all preferred slots (with a higher slot weight), are always offered to the potential customer. If at least one preferred slot is yet unbooked in period t, all less preferred slots are blocked. As soon as all preferred slots are booked, the less preferred slots are unblocked successively, one at a time. **(2) $\boldsymbol{R_{one}}$** offers one slot at a time. With different slot types, the slot weights are decisive for the order. First, the preferred slots of slot type 1 are successively offered. When all preferred slots are booked, the less preferred slots of slot type 2 are offered, again one at a time. Note that we consider offering one slot at a time because $a = \max\{0.25; 1\} = 1$ in all instances. A higher lower bound would increase the number of offered slots proportionally. **(3) $\boldsymbol{R_{myopic}}$** is the more complex decision rule. It checks in each period myopically all possible blocking options for each booking status and chooses the one with the lowest probability for the outside option.

Theorem 1: R_{myopic} is optimal for an appointment system with one single slot type.

We consider systems with one single slot type by setting $w_h = w_l$. For equally weighted slots (i.e., having one single slot type that contains all slots), the problem is reduced to a single dimensional dynamic program. The myopic rule always finds the optimal strategy (see Appendix H for proof of Theorem 1). Thus, for equally weighted slots, we have a static strategy for the blocking action.

6.3.2. Results: Decision Rules

Table 6 reports the performance of the rules compared to the stochastic dynamic program with the optimal state-dependent blocking strategy. Again, we focus on the instances with positive values (11,943) and refer to Table 11 in Appendix I for the results of all instances.[5] -0.0 denotes a deviation from the optimum ($< 0.05\%$) even though the rounded values suggest optimal results. Our base case, i.e., not blocking at all, is denoted by R_{no}.

[5] For equally weighted slots, we do not have any results for R_{low}, as there are no differences between the slot weights.

Table 6: Performance (percentage deviation, Δ%) of the decision rules varying β, slot weight (w_h) and reward for short-term customers (r_s) for all positive instances, with low=R_{low}, One=R_{one}, My.=R_{myopic}, No=R_{no}

	Overall				$w_h = 1$				$w_h = 2$				$w_h = 5$			
Rule	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**
Overall	-2.9	-4.5	-0.0	-8.0	-	-11	0.0	-9.6	-2.6	-3.7	-0.1	-9.7	-0.7	-0.5	-0.0	-5.3
$\beta = 0$	-3.4	-8.7	0.0	0.0	-	-19	0.0	0.0	-2.7	-7.3	0.0	0.0	-0.8	-1.0	0.0	0.0
$\beta = 0.5$	-2.5	-7.1	-0.0	-0.0	-	-16	0.0	0.0	-1.6	-5.6	-0.1	-0.0	-0.7	-0.9	-0.0	-0.0
$\beta = 1$	-1.4	-5.0	-0.1	-0.6	-	-12	0.0	-0.4	-0.8	-4.9	-0.1	-1.2	-0.5	-0.7	-0.0	-0.2
$\beta = 1.5$	-0.8	-3.2	-0.0	-2.8	-	-7.2	0.0	-4.1	-0.5	-2.6	-0.1	-4.2	-0.4	-0.5	-0.0	-0.4
$\beta = 3$	-2.6	-0.3	0.0	-18	-	-0.7	0.0	-31	-2.9	-0.3	-0.0	-25	-0.2	-0.0	-0.0	-2.7
$\beta = 6$	-9.1	0.0	0.0	-42	-	0.0	0.0	-54	-13	0.0	0.0	-56	-1.5	0.0	0.0	-31

	$r_s = 0.5$				$r_s = 1$				$r_s = 2$			
Rule	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**
Overall	-4.7	-11	-0.1	-7.8	-3.5	-6.7	-0.0	-8.7	-2.2	-2.1	-0.0	-7.7
$\beta = 0$	-7.9	-18	0.0	0.0	-4.9	-12	0.0	0.0	-1.5	-4.3	0.0	0.0
$\beta = 0.5$	-6.1	-15	-0.1	-0.0	-3.6	-10	-0.1	-0.0	-1.0	-3.4	-0.0	-0.0
$\beta = 1$	-3.8	-11	-0.2	-1.0	-2.0	-7.2	-0.1	-0.8	-0.5	-2.4	-0.0	-0.4
$\beta = 1.5$	-2.3	-7.6	-0.1	-5.3	-1.1	-4.6	-0.0	-4.0	-0.3	-1.5	-0.0	-1.6
$\beta = 3$	-3.3	-0.6	-0.0	-27	-3.6	-0.4	-0.0	-25	-2.1	-0.2	-0.0	-13
$\beta = 6$	-2.7	0.0	0.0	-54	-8.3	0.0	0.0	-52	-9.9	0.0	0.0	-39

The myopic rule performs very well, with an overall percentage difference of 0.02% to the optimum. It finds the optimal strategy for homogenous slots ($w_h = w_l = 1$, see Theorem 1) and for the extreme cases $\beta = 0$ and $\beta = 6$. In some instances with unequal slot weights, the myopic rule does not detect the optimal (time-dependent) strategy. Intuitively, the trade-off for blocking a slot is (a) decreasing the no-choice probability in a given period vs. (b) decreasing the likelihood that a potentially blocked slot is offered during the remaining time. The myopic rule disregards the latter part and tends to block slots too early when there are sufficient periods left until the day of service. In Appendix J, we present an example with two slot types and a system size of three slots for which the myopic rule does not find the optimal solution. In general, the myopic rule starts blocking (less preferred) slots from the very beginning if this minimizes the no-choice probability myopically. However, by blocking a less preferred slot, we eliminate the potential occasion to book that slot by chance and face a relatively high no-choice probability when having the less preferred slot in the offer set during the remaining periods. This reasoning carries only a little weight close to the day of service (because there are only a few periods left with relatively high no-choice probabilities). It neither holds if slots are weighted equally (because there are no relatively high or low no-choice probabilities in later periods, see also proof of Theorem 1).

In settings with a very strong effect, offering one slot at a time while blocking the remaining unbooked slots (R_{one}) also reaches, as the myopic strategy, the optimal expected profit. Note that for differently weighted slots, this means that first successively offering the more preferred slots followed by the less preferred slots is recommended, with a $\Delta\%$ of 0.

For very small β-values, it is worth considering the decision rule that only blocks less preferred slots (R_{low}) because of its easy implementation, even if other rules are dominant. At some point, it is better to block slots instead of offering all unbooked slots. If this threshold is not perfectly known (here around $1 \leq \tilde{\beta} \leq 1.5$), blocking only some slots as in R_{low} may be a good alternative to not blocking at all. In our numerical study, this is the case for $\beta = 1$ in combination with slightly different slot weights ($w_h = 2$; $w_l = 1$) and for all considered instances with $\beta = 1.5$. It should also be noted that R_{one} does not perform well for low and moderate β-values. Following R_{one}, the provider only offers one slot at a time. This scarce offer set is only advisable when the quality inference effect is very strong.

Besides the quality inference effect (β), another important driver is the weight of the slots. To derive recommendations for action, see Figure 2. Whenever decision rules lead to the same result or marginally different results, we chose the rule that is easier to implement. Alternative strategies that are easier to implement but have a slightly lower performance are added in brackets.

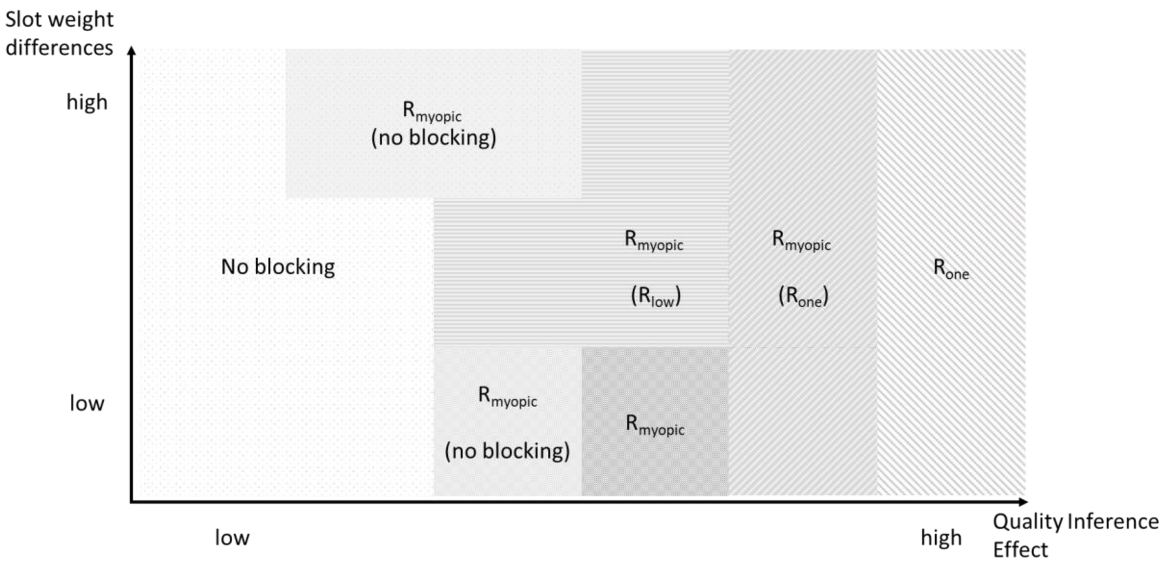

Figure 2: Action recommendation regarding the strength of the effect (QIE) and the difference of the slot preferences

For a very small impact of the effect ($\beta \leq 0.5$), blocking cannot be recommended in general, regardless of the slot weight difference (see Table 6). Instead, it is better to not block at all. For a small impact of the effect ($\beta = 1$ and $\beta = 1.5$), as in our survey, the myopic rule performs notably better than the baseline without blocking and the R_{low}-rule. If a simple decision rule is preferred, it is better to block the low weighted slots instead of not blocking at all. For a stronger impact ($\beta = 3$), the myopic rule is recommended, whereas, for a high impact, it is beneficial to use the R_{one}-rule.

We finally note that a decision rule that primarily blocks prioritized slots (reversed order in R_{one} regarding the order of slot types that are to be offered) is dominated in our numerical study by a strategy that first blocks less preferred slots. Intuitively, the main purpose of blocking slots is to trigger demand. Yet, if preferred slots are blocked, the no-choice probability increases more than if a less preferred slot is blocked.

7. Discussion and Limitations

Our survey results indicate that customers infer quality from the booking status of a dentist's online appointment system. This was, to our knowledge, the first survey on the quality inference effect in an online appointment context. We considered three different occupancy rates within one context. Further research is useful to (1) investigate more variations of number of offered slots and to (2) understand whether the effect also exists for other service areas apart from this specific context and how strongly the effect differs among other contexts. We conjecture that the quality inference effect may not be seen for standard services where the output does not depend on the provider (e.g., certificate of registration) but may be observed for non-standard services (e.g., coaching).

An additional survey may reveal further details about the threshold where the quality inference effect outweighs the scarcity effect. Further, we assumed a linear effect with respect to the occupancy rate of the appointment system. Non-linear effects could be further investigated, however, we expect our main insight to hold under other formulations of this utility component. Time horizons were excluded in the survey but showed an impact in the numerical analysis. The greater the time horizon, the lower the positive impact of the blocking possibility. In an additional survey, the impact of the time horizon on the interest in the providers may be investigated.

We also find indications of a scarcity effect, i.e., it might be more attractive to offer more slots than available. While the service provider has typically limited leeway to increase capacity on an operational level, it might be worth considering how the exact design of the appointment system's user interface (e.g., greying out booked slots vs. not showing them at all) impacts behavior.

We modeled a Markov decision process for one exclusive provider but empirically investigated the choice between two competitors. Our modeling is based on the assumption that potential customers learn from previous observations during the search process. The customer observes an offer set of one provider and decides whether to book one of the appointments or search for another provider. In our model, we assume that customers are at one point within this process but abstract from the comparison between several providers. Including this search process (i.e., equation (8) in Appendix D) in the model is left for future research. The chosen survey design enabled us to control many influencing factors. The same design with only one provider is also conceivable but seems very isolating, e.g., the participant may not know how much effort it would take to find another provider. However, further surveys focusing on one provider asking how likely it is that the participant chooses the provider may be interesting to see whether the quality inference effect can still be observed.

Similar to Gupta and Wang (2008b), the Markov decision process could further be extended to several providers in one system. A company with several providers, e.g., a clinic with several doctors, can still use our approach by having one system per provider. We leave the extension of our approach to include several providers for future research.

We modeled a homogenous group of customers who do not know the provider's service for the requests during the booking horizon ($t \geq 1$) and chose the multinomial logit model. Different discrete choice approaches are conceivable, such as a nested logit model, to model several heterogeneous groups of customers. We conjecture that within the homogenous subgroups, other approaches do not lead to significant changes regarding the blocking recommendation because assumptions remain unchanged within the subgroup. It would be interesting to see whether and to what extend blocking recommendations change when considering a heterogeneous customer group. Further, it may be more realistic to include customers who know the provider's service. This would increase the request probability and may weaken the blocking recommendation. Note that recommendations may further change when including other (overlapping) effects, e.g., very strong scarcity effect, as this may eliminate or significantly reduce the quality inference effect. It seems an interesting avenue for further research investigating if other available personal information could be used to tailor appointment systems (and the time slots being offered) to the individual customer.

For the choice probabilities of customer group 1, we only considered the quality inference effect besides the slot weights. By eliminating all other possible impacts, we could focus on the impact of the booking status on the choice behavior. Further aspects, such as ratings on portals, the distance to the service provider, individual preferences regarding the day or time of service, sex or age of the service provider can be incorporated into the random utility framework. Empirical studies on the time preferences may give further insights on realistic utilities and realistic assignments of appointments to slot types. Different set-ups for the slot weights (w_l and w_h) or more slot types are also possible and may lead to different recommendations in detail, but we assume that the overall action recommendation for the blocking strategy is not impacted significantly. Moreover, we assumed slot weights to be exogenous and time-independent. Further research might relax this assumption and provide further empirical evidence on the heterogeneity of customers and their time slot preferences.

The arrival rate of customers during the booking horizon was held constant for simplicity, even though it may be more realistic to let it be increased towards the planned workday. This would change the probability of a potential customer but not necessarily the slot preferences of that customer. Thus, dependent on the static arrival rate (high or low), considering an increasing probability may lead to less/more periods to be needed to reach the desired booking status.

Same-day customers are assumed to be Poisson-distributed. We assume that other distributions may only affect the number of slots that are to be left available for that customer group but not the blocking strategy to reach the desired booking status.

The development of the performance following the optimal strategy with an increasing strength of the quality inference effect is shown in Figure 1 in Section 6.2. When theoretically considering the boundary conditions, for $\beta \to +\infty$, the optimal strategy would lead to a blocking of as many slots as necessary to reach the lower bound (a) and thereby eliminating the effect. Without blocking, though, all customers would choose the no-choice option. The investigation of realistic, context specific, boundary conditions regarding the quality inference effect is left for future research.

We did not consider no-shows, cancellations and delays. Instead, we assumed that a customer who requests a slot arrives in time for the service. Including those factors would increase the complexity of the model and is left for future research. Considering no-shows and cancellations may lead to a higher booking limit or overbooking. More slot requests may be accepted as the service provider expects not all of the accepted customers to show up. This may also help in case the remaining offer set leads to the scarcity effect. Then, more slots than available may be offered. Delays may retard the service provision and may thus evoke overtime. This may lead to a lower booking limit. Zacharias and Pinedo (2014) for example include no-shows in their overbooking model for appointment planning, Kong et al. (2020) consider time-dependent no-shows in their distributionally robust model. See for example Hall (2006) for the consideration of delays.

The myopic rule leaves room for investigation. We found that the myopic view is optimal when having one single slot type (homogenous slots). However, we also found some critical instances, for which it is time-dependent whether slots should be blocked or not, independent from the fact whether blocking was needed to minimize the no-choice probability of the respective period. As long as enough periods are left until the day of the appointment, no slots are recommended to be blocked. This is no longer in line with the myopic view, even though the myopic rule is close to the optimal expected profit. Instead of the myopic rule as a good approximation of the sdp, an approximate dynamic program may be developed. The insights from the myopic rule may be helpful in that process.

It is important to note that service providers do not have the right to deceive customers online in, e.g., Germany by pretending scarce supply (§ 5 I UWG in conjunction with § 3 UWG)[6]. However, this law does not oblige service providers to announce all conceivable appointments at each point in time.

8. Conclusion

We analyze an online appointment system as a Markov decision process that aims at maximizing the service provider's expected profit. In a survey, we identify that an extensive set of available appointments leads to significantly less demand because customers infer a lower quality of the service (observational learning from previous customer choices). We capture this quality inference effect in a discrete choice model and provide quantitative decision support and qualitative insights on which time slots should be offered during the booking horizon.

We analyze the benefits of blocking and releasing time slots in a numerical study. Intuitively, the benefits are larger the stronger the customer reacts to underutilized service providers. We present three decision rules, since solving the problem to optimality (in realistic dimensions) is computationally expensive. The myopic rule, which chases the minimum no-choice probability in each period, performs very well and is optimal for equally weighted slots. Simpler decision rules perform reasonably if the quality inference effect is sufficiently strong.

Publication bibliography

Abelson, Reed (2020): Doctors and Patients Turn to Telemedicine in the Coronavirus Outbreak. Edited by The New York Times. Available online at https://www.nytimes.com/2020/03/11/health/telemedicine-coronavirus.html, checked on 6/23/2020.

Ahmadi-Javid, Amir; Jalali, Zahra; Klassen, Kenneth J. (2017): Outpatient appointment systems in healthcare: A review of optimization studies. In *European Journal of Operational Research* 258 (1), pp. 3–34.

Banerjee, Abhijit V. (1992): A simple model of herd behavior. In *The Quarterly Journal of Economics* 107 (3), pp. 797–817.

Becker, Gary S. (1991): A Note on Restaurant Pricing and Other Examples of Social Influences on Price. In *Journal of Political Economy* 99 (5), pp. 1109–1116.

Belobaba, Peter P. (1989): Application of a Probabilistic Decision Model to Airline Seat Inventory Control. In *Operational Research* 37 (2), pp. 183–197.

[6] Thus, it is not allowed to indicate that only a specific number of appointments are left unbooked, e.g., "only three (of ten) slots are left".

Ben-Akiva, Moshe; Lerman, Steven R. (1985): Discrete choice analysis: theory and application to travel demand: MIT Press.

Bikhchandani, Sushil; Hirshleifer, David; Welch, Ivo (1992): A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. In *Journal of Political Economy* 100 (5), pp. 992–1026.

Byrt, Ted; Bishop, Janet; Carlin, John B. (1993): Bias, prevalence and kappa. In *Journal of clinical epidemiology* 46 (5), pp. 423–429.

Cardew, Bruce (1967): An Appointment System Service for General Practitioners: Its Growth and Present Usage. In *British Medical Journal* 4, pp. 542–543.

Cayirli, Tugba; Veral, Emre (2003): OUTPATIENT SCHEDULING IN HEALTH CARE: A REVIEW OF LITERATURE. In *Production and operations management* 12 (4), pp. 519–549.

Debo, Laurens G.; Parlour, Christine; Rajan, Uday (2012): Signaling Quality via Queues. In *Management Science* 58 (5), pp. 876–891.

DeCroix, Gregory; Long, Xiaoyang; Tong, Jordan (2020): How Service Quality Variability Hurts Revenue When Customers Learn: Implications for Dynamic Personalized Pricing. In *Operations Research (AiA)*.

Denier, Yvonne (2008): Mind the gap! Three approaches to scarcity in health care. In *Medicine, health care, and philosophy* 11 (1), pp. 73–87.

DeVries, Jelle; Roy, Debjit; DeKoster, René (2018): Worth the wait? How restaurant waiting time influences customer behavior and revenue. In *Journal of Operations Management* 63, pp. 59–78.

Eppler, Martin J.; Mengis, Jeanne (2004): The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. In *The Information Society* 20 (5), pp. 325–344.

Feldman, Jacob; Liu, Nan; Topaloglu, Huseyin; Ziya, Serhan (2014): Appointment Scheduling Under Patient Preference and No-Show Behavior. In *Operations Research* 62 (4), pp. 794–811.

Gerchak, Yigal; Gupta, Diwakar; Henig, Mordechai (1996): Reservation Planning for Elective Surgery under Uncertain Demand for Emergency Surgery. In *Management Science* 42 (3), pp. 321–334.

Giebelhausen, Michael D.; Robinson, Stacey G.; Cronin, J. Joseph (2011): Worth waiting for: increasing satisfaction by making consumers wait. In *Journal of the Academy of Marketing Science* 39 (6), pp. 889–905.

Green, Linda V. (2006): Queueing analysis in healthcare. In R. W. Hall (Ed.): Patient flow: Reducing delay in healthcare delivery: Springer, pp. 281–307.

Green, Linda V.; Savin, Sergei; Wang, Ben (2006): Managing Patient Service in a Diagnostic Medical Facility. In *Operations Research* 54 (1), pp. 11–25.

Gupta, Diwakar; Wang, Lei (2008a): Electronic Companion—"Revenue Management for a Primary-Care Clinic in the Presence of Patient Choice" by Diwakar Gupta and Lei Wang, Operations Research, doi 10.1287/opre.1080.0542., ec1-ec8.

Gupta, Diwakar; Wang, Lei (2008b): Revenue Management for a Primary-Care Clinic in the Presence of Patient Choice. In *Operations Research* 56 (3), pp. 576–592.

Haase, Knut; Müller, Sven (2013): Management of school locations allowing for free school choice. In *Omega* 41 (5), pp. 847–855.

Hall, Randolph W. (Ed.) (2006): Patient Flow: Reducing Delay in Healthcare Delivery. Boston: Springer (International Series in Operations Research & Management Science, 91).

Heath, David; Greenblatt, Mark; Bagchi, Aysha (2020): Dentists under pressure to drill 'healthy teeth' for profit, former insiders allege. Edited by USA Today. Available online at https://eu.usatoday.com/in-depth/news/investigations/2020/03/19/dental-chain-private-equity-drills-healthy-teeth-profit/4536783002/, checked on 7/15/2020.

Hole, Arne R. (2008): Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. In *Journal of health economics* 27 (4), pp. 1078–1094.

Hole, Arne R. (2011): A discrete choice model with endogenous attribute attendance. In *Economics Letters* 110 (3), pp. 203–205.

Jin, Chen; Debo, Laurens; Iravani, Seyed; Kremer, Mirko (2015): Observational Learning in Congested Environments with Multiple Choice Options: The Wisdom of Majorities and Minorities. Chicago Booth Research Paper 8. (Working paper).

Kennedy, John F. (1963): Remarks in Heber Springs, Arkansas, at the Dedication of Greers Ferry Dam., 10/3/1963. Available online at https://www.presidency.ucsb.edu/node/236260, checked on 10/1/2019.

Kong, Qingxia; Li, Shan; Liu, Nan; Teo, Chung-Piaw; Yan, Zhenzhen (2020): Appointment Scheduling Under Time-Dependent Patient No-Show Behavior. In *Management Science* 66 (8), pp. 3480–3500.

Koo, Minjung; Fishbach, Ayelet (2010): A Silver Lining of Standing in Line: Queuing Increases Value of Products. In *Journal of Marketing Research* 47 (4), pp. 713–724.

Kremer, Mirko; Debo, Laurens (2016): Inferring quality from wait time. In *Management Science* 62 (10), pp. 3023–3038.

Liu, Nan; Finkelstein, Stacey R.; Kruk, Margaret E.; Rosenthal, David (2018): When Waiting to See a Doctor Is Less Irritating. Understanding Patient Preferences and Choice Behavior in Appointment Scheduling. In *Management Science* 64 (5), pp. 1975–1996.

Liu, Nan; van de Ven, Peter M.; Zhang, Bo (2019): Managing Appointment Booking under Customer Choices. In *Management Science* 65 (9), pp. 4280–4298.

Luce, R. Duncan (1959): Individual choice behavior; a theoretical analysis: Wiley, New York.

Mackert, Jochen (2019): Choice-based dynamic time slot management in attended home delivery. In *Computers & Industrial Engineering* 129, pp. 333–345.

McFadden, Daniel (1974): Conditional logit analysis of qualitative choice behavior. In Paul Zarembka (Ed.): Frontiers in Economics: Academic Press, pp. 105–142.

McGill, Jeffrey I.; van Ryzin, Garrett J. (1999): Revenue Management. Research Overview and Prospects. In *Transportation Science* 33 (2), pp. 233–256.

Mushtaque, Uzma; Pazour, Jennifer A. (2020): Random Utility Models with Cardinality Context Effects for Online Subscription Service Platforms. In *Journal of Revenue and Pricing Management* 19 (4), pp. 276–290.

Özer, Özalp; Zheng, Yanchong (2016): Markdown or Everyday Low Price? The Role of Behavioral Motives. In *Management Science* 62 (2), pp. 326–346.

Rubin, Greg; Bate, Angela; George, Ajay; Shackley, Phil; Hall, Nicola (2006): Preferences for access to the GP: a discrete choice experiment. In *British Journal of General Practice* 56 (531), pp. 743–748.

Scheibehenne, Benjamin; Greifeneder, Rainer; Todd, Peter M. (2010): Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload. In *Journal of Consumer Research* 37 (3), pp. 409–425.

Shukla, Aishwarya Deep; Gao, Guodong; Agarwal, Ritu (2020): How Digital Word-of-Mouth Affects Consumer Decision Making: Evidence from Doctor Appointment Booking. In *Management Science* 67 (3), pp. 1546–1568.

Span, Paula (2020): With Red Tape Lifted, Dr. Zoom Will See You Now. Edited by The New York Times. Available online at https://www.nytimes.com/2020/05/08/health/coronavirus-telemedicine-seniors.html, checked on 6/23/2020.

Talluri, Kalyan; van Ryzin, Garrett (2004a): Revenue Management Under a General Discrete Choice Model of Consumer Behavior. In *Management Science* 50 (1), pp. 15–33.

Talluri, Kalyan; van Ryzin, Garrett (2004b): The Theory and Practice of Revenue Management. Kluwer Academic Publishers, Boston.

Teraji, Shinji (2003): Herd behavior and the quality of opinions. In *The Journal of Socio-Economics* 32 (6), pp. 661–673.

Train, Kenneth E. (2009): Discrete Choice Methods with Simulation. 2nd ed. Cambrigde: Cambrigde University Press.

Truong, Van-Anh (2015): Optimal advance scheduling. In *Management Science* 61 (7), pp. 1584–1597.

U.S. Government Accountability Office (2017): HEALTH CARE: Telehealth and Remote Patient Monitoring Use in Medicare and Selected Federal Programs. GAO 17-365. Available online at http://www.gao.gov.

Veeraraghavan, Senthil; Debo, Laurens (2009): Joining Longer Queues: Information Externalities in Queue Choice. In *Manufacturing & Service Operations Management* 11 (4), pp. 543–562.

Veeraraghavan, Senthil K.; Debo, Laurens G. (2011): Herding in Queues with Waiting Costs: Rationality and Regret. In *Manufacturing & Service Operations Management* 13 (3), pp. 329–346.

Wiesche, Lara; Schacht, Matthias; Werners, Brigitte (2017): Strategies for interday appointment scheduling in primary care. In *Health care management science* 20 (3), pp. 403–418.

Zacharias, Christos; Liu, Nan; Begen, Mehmet A. (2020): Dynamic interday and intraday scheduling working paper. SSRN. Available online at https://ssrn.com/abstract=3728077 or http://dx.doi.org/10.2139/ssrn.3728077, updated on 11/10/2020.

Zacharias, Christos; Pinedo, Michael (2014): Appointment Scheduling with No-Shows and Overbooking. In *Production and operations management* 23 (5), pp. 788–801.

Zhao, Peng; Yoo, Illhoi; Lavoie, Jaie; Lavoie, Beau James; Simoes, Eduardo (2017): Web-Based Medical Appointment Systems: A Systematic Review. In *Journal of medical Internet research* 19 (4), e134.

Zhou, Minglong; Loke, Gar Goei; Bandi, Chaithanya; Liau, Zi Qiang Glen; Wang, Wilson (2021): Intraday Scheduling with Patient Re-entries and Variability in Behaviours. In *Manufacturing & Service Operations Management (AiA)*.

Appendix

A. Survey Supplement

In Figure 3, we present an exemplary scenario given to, e.g., treatment T1a (first choice) in the survey.

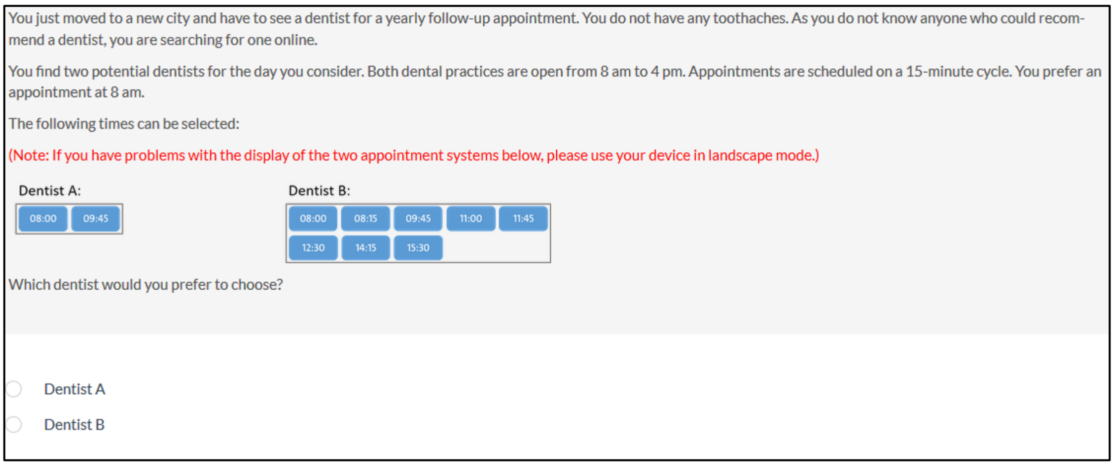

Figure 3: Exemplary scenario in the survey

B. Motives for Choices in the Survey

In Table 7, we present the motives for the choice behavior of all respondents in the survey. The two independent raters also coded the answers of all the remaining respondents, see Section 3.3 for details. Additionally to the frequency of mentions, again the Cohen's kappa is given for the inter-rater reliability.

Table 7: Frequency of the motives for the choice behavior of all 248 respondents

	Quality	Choice overload	Flexibility	Scarcity	Less wait time	Other
Frequency in %	42.94	4.23	38.10	3.70	5.82	23.80
(Cohen's kappa)	(0.9589)	(0.7514)	(0.7350)	(0.5088)	(0.8902)	(0.5270)

Note. Multiple answers possible, which leads to a sum of the frequencies >100%.

C. Results of All Three Scenarios of the Survey

We summarize the choice frequency of all observations in Table 8.

Table 8: Descriptive statistics of all choices

	2 vs. 8 slots (T1)	2 vs. 32 slots (T2)	8 vs 32 slots (T3)
Choice frequency	40% vs. 60%	58% vs. 42%	73% vs. 27%
# observations	248	248	248

In Table 9, we show the estimates of the $\hat{\beta}$-coefficients for all observations. For all observations, we use a random-effects model in R (version 4.0.2.) to account for multiple decisions per respondent. This individual specific error term is included in the (higher) robust standard error. We further add a (fixed) treatment effect (one $\hat{\beta}$-coefficient per treatment, $\hat{\beta}_{T1a}, \hat{\beta}_{T1b}, \hat{\beta}_{T2a}, \hat{\beta}_{T2b}, \hat{\beta}_{T3a}, \hat{\beta}_{T3b}$). The estimates are given in table form for the sake of clarity.

Table 9: Estimates of $\hat{\beta}$-coefficients for all observations, $\hat{\beta}_{T1a}$ = base for fixed effect

	Treatment	$\hat{\beta}$	Std. error	t-ratio	Rob. Std. err.	Rob. t-ratio
All observations	size	0.6052	0.1033	5.86	0.1282	4.72
	T1b	-0.3013	0.1624	-1.86	0.2532	-1.19
	T2a	-0.0493	0.1681	-0.29	0.2832	-0.17
	T2b	-0.1323	0.1686	-0.79	0.2840	-0.47
	T3a	-0.0296	0.1696	-0.17	0.2908	-0.10
	T3b	0.6786	0.1963	3.46	0.4361	1.56
2 vs 8	size	-2.3451	0.6748	-3.48	0.6656	-3.52
	T1b	-0.3187	0.2821	-1.13	0.2580	-1.24
	T2a	-0.0896	0.2919	-0.31	0.2879	-0.31
	T2b	-0.1356	0.2926	-0.46	0.2887	-0.47
	T3a	0.0668	0.2950	0.23	0.2947	0.23
	T3b	0.8509	0.3416	2.49	0.4325	1.97
2 vs 32	size	0.3365	0.1307	2.58	0.1305	2.58
	T1b	-0.2899	0.2776	-1.04	0.2526	-1.15
	T2a	-0.0519	0.2881	-0.18	0.2835	-0.18
	T2b	-0.1173	0.2885	-0.41	0.2842	-0.41
	T3a	-0.0448	0.2906	-0.15	0.2914	-0.15
	T3b	0.7020	0.3388	2.07	0.4424	1.59
8 vs 32	size	1.3013	0.1870	6.96	0.1883	6.91
	T1b	-0.4164	0.3206	-1.30	0.3098	-1.34
	T2a	0.0088	0.3219	0.03	0.3172	0.03
	T2b	-0.3535	0.3334	-1.06	0.3437	-1.03
	T3a	0.1011	0.3226	0.31	0.3180	0.32
	T3b	0.6083	0.3564	1.71	0.4233	1.44

D. Multiple Service Providers

Let I be the number of considered service providers (e.g., $I = 2$ in our survey, see Section 3). Each service provider i has a fixed capacity of κ_i slots on the workday. Assuming that the providers have the same capacity because the providers offer similar services, enables us to drop the index. This leads to a capacity of κ per provider. For simplicity and without loss of generality, we assume that all considered service providers have identical slot types. Each slot belongs to one unique slot type n and can be booked for at most one customer per provider. The booking status of the workday per provider $i \in I$ is denoted by $\vec{s}_i = (d_{i1}, \dots, d_{iN}, q_{i1}, \dots, q_{iN})$, whereby $d_{in} \in \{0,1, \dots, \kappa_n\}$ states how many slots of slot type n offered by provider i are already booked for a customer, with $\vec{d}_i = (d_{i1}, \dots, d_{iN})$. It further states by $q_{in} \in \{0,1, \dots, \kappa_n - d_{in}\}$ how many slots of slot type n of provider i are temporarily blocked, with $\vec{q}_i = (q_{i1}, \dots, q_{iN})$. If an offered slot of slot type n and provider i is booked for a customer, d_{in} is increased by 1. $\vec{o}_i = (o_{i1}, \dots, o_{iN})$ states the resulting number of offered slots per slot type of provider i, with $\sigma_i = \sum_{n\in N} o_{in} = \sum_{n\in N}(\kappa_n - d_{in} - q_{in})$ the size of the overall offer set of provider i. $\vec{o}_J = (\vec{o}_1, \dots, \vec{o}_I)$ gives the overall set of offer sets the customer observes.

A potential customer chooses either an appointment offered by one of the providers $i \in I$ or no appointment at all. $P_{in}(\vec{o}_J)$ denotes the request probability of any slot of type n of provider i observing the offer sets of all providers $i \in I$, see equation (7). The weight of the no-choice option (not booking any service) is denoted by $\overline{w}_0$.

$$\mathcal{P}_{in}(\vec{o}_J) = \begin{cases} \dfrac{o_{in} \cdot e^{w_n - \beta\cdot\left(\frac{\sigma_i - a}{b-a}\right)^+}}{\sum_{i\in I}\sum_{m\in N} o_{im} \cdot e^{w_m - \beta\cdot\left(\frac{\sigma_i - a}{b-a}\right)^+} + e^{\overline{w}_0}} & \text{if } o_{in} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Simplifying equation (7) by accounting for the quality inference effect in the no-choice option instead of subtracting it from each appointment option, gives equation (8).

$$\mathcal{P}_{in}(\vec{o}_J) = \begin{cases} \dfrac{o_{in} \cdot e^{w_n}}{\sum_{i\in I}\sum_{m\in N}\left(o_{im} \cdot e^{w_m} + e^{\overline{w}_0 + \beta\cdot\left(\frac{\sigma_i - a}{b-a}\right)^+}\right)} & \text{if } o_{in} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

As we focus on one service provider, we let z be the choice possibilities of all other providers $\tilde{j} \in I\backslash i$, with the constant $z = \sum_{\tilde{j}\in I\backslash i} o_{\tilde{j}m} \cdot e^{w_m}\, e^{\overline{w}_0 + \beta\cdot\left(\frac{\sigma_{\tilde{j}} - a}{b-a}\right)^+}$. Rearranging equation (8) gives equation (9).

$$\mathcal{P}_{in}(\vec{o}_i) = \begin{cases} \dfrac{o_{in} \cdot e^{w_n}}{\sum_{m\in N} o_{im} \cdot e^{w_m} + z + e^{\overline{w}_0 + \beta\cdot\left(\frac{\sigma_i - a}{b-a}\right)^+}} & \text{if } o_{in} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

From the provider i's perspective, it makes no difference whether the customer chooses another provider $\tilde{j} \in I\backslash i$ or the no-choice option. For the sake of brevity, we thus aggregate those options to the no-choice option with weight w_0 by defining $z + e^{\overline{w}_0 + \beta\cdot\left(\frac{\sigma_i - a}{b-a}\right)^+} = e^{w_0 + \beta\cdot\left(\frac{\sigma_i - a}{b-a}\right)^+}$ while normalizing w_0 to zero. Dropping the index i results in equation (2b) in Section 4.

E. Booking Limit

We follow Gupta and Wang (2008a, 2008b) for showing that the booking limit remains unchanged to the booking limit presented by Gupta and Wang (2008b).

Gupta and Wang show the existence of b^* on an intuitive level and by calculating (2008b) and proof the booking-limit policy (2008a). The booking limit b^* saves capacity for same-day requests and thus depends on those customers. Consider the expected benefit ($(r_s + c_o - r_p - \pi)P(X > \kappa - b)$) and expected loss ($(r_p + \pi + c_u)P(X \leq \kappa - b)$) of saving one more slot for same-day requests. b^* is determined by the maximum value of b for which $(r_s + c_o - r_p - \pi)P(X > \kappa - b) < (r_p + \pi + c_u)P(X \leq \kappa - b)$. The critical ratio is thus set by

$$\frac{(r_s+c_o-r_p-\pi)}{(r_s+c_o-r_p-\pi)+(r_p+\pi+c_u)} = \frac{r_s+c_o-r_p-\pi}{r_s+c_o+c_u}. \qquad (10)$$

As our blocking option neither affects the expected number of same-day requests nor the parameters $r_p, r_s, \pi, c_o, c_u, \kappa$, the booking limit b^* is not affected by our additional blocking action.

F. Properties of $\Delta V(s)$

We show the existence of a threshold value $\tilde{\beta}$ for which blocking is beneficial if $\mathfrak{u} = 2$, $b^* = \kappa$, $c_u = c_o = \lambda = 0$ and $\alpha_1 = \alpha_2 = 1$. With two remaining slots, we consider two slot types, each containing one slot. Slot 1 belongs to slot type 1; slot 2 belongs to slot type 2. Thus, it makes no difference whether we consider the slot types or the specific slots. We compare the expected profit in case slot 1 is blocked in period 1 with the expected profit without blocking. Let Δp_j be the overall probability that slot j is requested within two periods, in period one or in period two given what happened in period one. If $\beta > \tilde{\beta}$, then $\Delta V(\vec{s}) > 0$, with

$$\Delta V(\vec{s}) = (\Delta p_1 + \Delta p_2) \cdot r_p$$
$$= \frac{-2r_p e^{2w_1+w_2+w_0} - 2r_p e^{w_1+2w_2+w_0} + r_p e^{\beta}\left(e^{2w_1+w_2+w_0} + e^{w_1+2w_2+w_0} - e^{w_1+3w_0} - w^{w_2+3w_0} - 3e^{w_1+w_2+2w_0} + e^{\beta}\left(e^{w_2+3w_0} + 2e^{w_1+w_2+2w_0}\right)\right)}{(e^{w_1} + e^{w_2} + e^{w_0+\beta})^2 (e^{w_1} + e^{w_0})(e^{w_2} + e^{w_0})} \qquad (11)$$

A: For $\beta = 0$, the numerator is negative $[-r_p e^{2w_1+w_2+w_0} - r_p e^{w_1+2w_2+w_0} - r_p e^{w_1+w_2+2w_0} - r_p e^{2w_1+2w_0} - r_p e^{w_1+3w_0} < 0]$ while the denominator is positive $[(e^{w_1} + e^{w_2} + e^{w_0})^2 (e^{w_1} + e^{w_0})(e^{w_2} + e^{w_0}) > 0]$, as $w_1, w_2, w_0 \geq 0$, which results in a negative value, $\Delta V(\vec{s}) < 0$.

B: For $\beta \to +\infty$, both the numerator and the denominator approach $+\infty$. The L'Hôpital's rule does not give any insights about the convergence (all derivations have the same characteristic regarding the convergence, $\lim_{\beta \to +\infty} \Delta V(\vec{s}) = \frac{+\infty}{+\infty}$). Both, the numerator and the denominator, have the same highest power of $e^{2\beta}$ (with specific pre-factors). The function thus approaches the pre-factors of the highest power $\lim_{\beta \to +\infty} \Delta V(\vec{s}) = \frac{r_p e^{w_2+3w_0} + 2r_p e^{w_1+w_2+2w_0}}{e^{w_1+w_2+2w_0} + e^{w_1+3w_0} + e^{w_2+3w_0} + e^{4w_0}}$, which results in a positive value.

C: For $\beta = 0$, it starts with a negative value (for $\beta \rightarrow -\infty$, it even converges towards $\lim_{\beta \rightarrow -\infty} \Delta\mathrm{V}(\vec{s}) = \frac{-2r_p e^{2w_1+w_2+w_0} - 2r_p e^{w_1+2w_2+w_0}}{(e^{w_1+w_2+w_0})^2 (e^{w_1+w_0})(e^{w_2+w_0})}$) and ends with a positive value for $\beta \rightarrow +\infty$. Thus, a zero of the function must exist. For $\Delta V(\vec{s}) = 0$, we found several cases, as it is not possible to solve it in closed form. The analysis of the monotony for $\beta \geq 0$ and $w_0 = 0$ (base category) helps. The derivation is determined using the sum rule, factor rule, quotient rule and chain rule and will be handed out on demand. The denominator of the derivation is obviously positive (as it is the squared denominator). The numerator of the derivation (for which the numerator and the denominator of the formula are needed) evidently consists of positive and negative terms. For each negative term, it exists a greater positive term. When cancelling out the negative terms, only positive terms remain which results in a positive numerator. The derivation is thus > 0. This shows a monotonically increasing function and excludes any oscillation.

To conclude, this curve implies a threshold $\tilde{\beta}$, for which $\Delta V(\vec{s}) > 0$ if $\beta > \tilde{\beta}$. Thus, blocking is beneficial if the quality inference effect is sufficiently strong ($> \tilde{\beta}$) for $|\mathcal{U}(\vec{s})| = 2$, $b^* = \kappa$, $c_u = c_o = \lambda = 0$ and $\alpha_1 = \alpha_2 = 1$. □

G. Results of Full Factorial Analysis: SDP vs. Base Without Blocking

Regarding the whole data set (optimal blocking strategy compared to a baseline without blocking), we observe cost savings (instances with negative values) and profit improvements (instances with positive values), see Table 10. To make both blocks comparable, we change the basis if necessary. Assume for example an improvement from -4 to -2, which results in cost savings from 50%. An improvement from 2 to 4 results in a profit increase of 100%. In absolute numbers, both result in the same improvement of 2, but the percentage improvement is significantly different. For profits (overall positive values for both, the base and the sdp), we show the improvement of our sdp compared to the base without blocking and mark this comparison with $^+$. Thus, the basis in those cases is the base without blocking. For the instances where both, the base and the sdp, overall result in costs (negative values), we change the base to the sdp and indicate the percentage decline of the base without blocking compared to the sdp with $^-$. Furthermore, we have instances, where the base without blocking overall results in costs (a negative value) and the sdp achieves an overall profit (positive value). We interpret those cases as cost savings of more than 100% and take the base without blocking as the basis. Those cases are indicated with $^\pm$. Note that 100% is on the one hand an improvement to zero cost and on the other hand a decline to zero profit, which will both result in difficulties regarding the interpretation.

Table 10: Percentage change varying β, w_h and r_s regarding all instances

Parameter	Overall	$\beta = 0$	$\beta = 0.5$	$\beta = 1$	$\beta = 1.5$	$\beta = 3$	$\beta = 6$
Overall	3178.03$^-$	0.00	1.21$^+$	321.80$^-$	130.25$^-$	620.01$^-$	3283.78$^-$
$w_h = w_l = 1$	509.07$^-$	0.00	0.00	5.70$^-$	34.51$^-$	307.58$^-$	1299.33$^-$
$w_h = 2, w_l = 1$	100.47$^\pm$	0.00	1.52$^+$	362.51$^\pm$	640.38$^-$	1243.91$^-$	7737.08$^-$
$w_h = 5, w_l = 1$	458.01$^\pm$	0.00	0.55$^+$	2.17$^+$	5.68$^+$	82.05$^+$	114.50$^+$
$r_s = 0.5$	84.60$^-$	0.00	0.06$^-$	1.35$^-$	6.81$^-$	69.36$^-$	380.00$^-$
$r_s = 1$	207.84$^-$	0.00	0.18$^-$	3.55$^-$	16.05$^-$	149.60$^-$	831.30$^-$
$r_s = 2$	72.85$^+$	0.00	0.04$^+$	0.72$^+$	3.55$^+$	59.43$^+$	181.62$^\pm$

For $\beta = 1$ to $\beta = 1.5$, there seems to be a decrease in improvement/decline. The absolute values show a different interpretation. For $\beta = 1$, the overall cost saving with the sdp amount to around 37, whereas for $\beta = 1.5$, the overall cost saving amount to around 191.

H. Proof Theorem 1 (Myopic rule is optimal for $w_l = w_h$)

Let b_t be the blocking option in period t. Blocking option means to block as many slots as possible up to a given limit (sum of blocked and booked slots). E.g., $b_1 = 2$ means to block up to two slots in period 1 such that the sum of blocked and booked slots is 2. Note that at least one slot is needed unblocked and unbooked in order to offer at least one appointment. We consider homogenous slots regarding the customers' preferences, which is like having only one slot type, and therefore consider the specific slots. From equation (3) we know that the overall expected profit is to be maximized. Revenue is assigned by accepting requests and same-day requests. We focus on the requests before the day of service, as blocking is only conceivable during the booking horizon ($t \geq 1$). Remember that same-day requests must not be denied, thus, exp. revenue for that customer group is given by their expected number. The no-choice probability for a given state is denoted by $A_{\max\{w;b_t\}}$ with w the number of booked slots, $0 < A_{\max\{w;b_t\}} \leq 1$. In each period, the myopic rule chooses the blocking option that myopically minimizes $A_{\max\{w;b_t\}}$.

This proof is by induction on t. First, Theorem 1 holds for $t = 1$ (last period before the day of service), as the maximum overall expected profit is reached if the maximum expected profit in $t = 1$ is reached. The greater the probability of a request, the greater the expected profit. The maximized probability of a slot request ($1 - A_{\max\{w;b_t\}}$) is synonymous with the minimized no-choice probability $A_{\max\{w;b_t\}}$. Assume two possible strategies for a period t, namely b_t^1 and b_t^2, with b_t^1 the dominant strategy (if the strategies differ). Then,

$$A_{\max\{w;b_1^1\}} \leq A_{\max\{w;b_1^2\}} \,. \tag{12}$$

The lowest $A_{\max\{w;b_1\}}$ leads to the optimal blocking strategy in $t = 1$.

For $t = 1,2$ (remember that time is counted backwards),

$$A_{\max\{w;b_2\}} + A_{\max\{w;b_2\}} A_{\max\{w;b_1\}} + \left(1 - A_{\max\{w;b_2\}}\right) A_{\max\{w+1;b_1\}} \tag{13}$$

is to be minimized in order to maximize the overall expected profit. We know from (12) that in the last period ($t = 1$), the lowest $A_{\max\{w;b_1\}}$ is optimal. Then,

$$A_{\max\{w;b_2^1\}}\left(1 + A_{\max\{w;b_1\}} - A_{\max\{w+1;b_1\}}\right) \leq A_{\max\{w;b_2^2\}}\left(1 + A_{\max\{w;b_1\}} - A_{\max\{w+1;b_1\}}\right) . \tag{14}$$

Thus, the lowest $A_{\max\{w;b_2\}}$ leads to the optimal blocking strategy in $t = 2$.

For $t = 1,2,3$,

$$\begin{aligned}
&A_{\max\{w;b_3\}} + A_{\max\{w;b_3\}} A_{\max\{w;b_2\}} + \left(1 - A_{\max\{w;b_3\}}\right) A_{\max\{w+1;b_2\}} + A_{\max\{w;b_3\}} A_{\max\{w;b_2\}} A_{\max\{w;b_1\}} \\
&+ A_{\max\{w;b_3\}} \left(1 - A_{\max\{w;b_2\}}\right) A_{\max\{w+1;b_1\}} + \left(1 - A_{\max\{w;b_3\}}\right) A_{\max\{w+1;b_2\}} A_{\max\{w+1;b_1\}} \\
&+ \left(1 - A_{\max\{w;b_3\}}\right)\left(1 - A_{\max\{w+1;b_2\}}\right) A_{\max\{w+2;b_1\}})
\end{aligned} \tag{15}$$

is to be minimized in order to maximize the overall expected profit. From (12) and (14), it is known that b_1^1 and b_2^1 are optimally chosen. Then,

$$\begin{aligned}
&A_{\max\{w;b_3^1\}} (1 + A_{\max\{w;b_2\}} - A_{\max\{w+1;b_2\}} + A_{\max\{w;b_2\}} A_{\max\{w;b_1\}} + A_{\max\{w+1;b_1\}} - A_{\max\{w;b_2\}} A_{\max\{w+1;b_1\}} \\
&- A_{\max\{w+1;b_2\}} A_{\max\{w+1;b_1\}} - A_{\max\{w+2;b_1\}} + A_{\max\{w+1;b_2\}} A_{\max\{w+2;b_1\}}) \\
&\leq A_{\max\{w;b_3^2\}} (1 + A_{\max\{w;b_2\}} - A_{\max\{w+1;b_2\}} + A_{\max\{w;b_2\}} A_{\max\{w;b_1\}} + A_{\max\{w+1;b_1\}} - A_{\max\{w;b_2\}} A_{\max\{w+1;b_1\}} \\
&- A_{\max\{w+1;b_2\}} A_{\max\{w+1;b_1\}} - A_{\max\{w+2;b_1\}} + A_{\max\{w+1;b_2\}} A_{\max\{w+2;b_1\}})
\end{aligned} \tag{16}$$

Thus, the lowest $A_{\max\{w;b_3\}}$ leads to the optimal blocking strategy in $t = 3$.

Assume the Theorem 1 holds for some t. By showing that the results also hold for $t + 1$, the proof is completed. By equation (3), it follows for any $t \geq 1$ that

$$\begin{aligned}
&\sum_{u=1}^{t} \prod_{v=1}^{u} A_{\max\{w;b_{t-v+1}\}} + \sum_{u_1=1}^{t-1} \sum_{u_0=u_1+1}^{t} \prod_{\substack{v_1=1;\\ v_1 \neq u_1}}^{u_1} \prod_{v_0=u_1+1}^{u_0} \left(1 - A_{\max\{w;b_{t-u_1+1}\}}\right) A_{\max\{w;b_{t-v_1+1}\}} A_{max\{w+1;\, b_{t-v_0+1}\}} \\
&+ \sum_{u_2=1}^{t-2} \sum_{u_1=u_2+1}^{t-1} \sum_{u_0=u_1+1}^{t} \prod_{\substack{v_2=1;\\ v_2 \neq u_2}}^{u_2} \prod_{\substack{v_1=u_2+1;\\ v_1 \neq u_1}}^{u_1} \prod_{v_0=u_1+1}^{u_0} \left(1 - A_{\max\{w;b_{t-u_2+1}\}}\right) A_{\max\{w;b_{t-v_2+1}\}} \\
&\left[\left(1 - A_{\max\{w+1;b_{t-u_1+1}\}}\right) A_{\max\{w+1;b_{t-v_1+1}\}} A_{max\{w+2;\, b_{t-v_0+1}\}}\right] + [\dots] \\
&+ \sum_{u_{\kappa-w}=1}^{t-(\kappa-w)} \sum_{u_{\kappa-w-1}=u_{\kappa-w}+1}^{t-(\kappa-w-1)} [\dots] \sum_{u_1=u_2+1}^{t-1} \sum_{u_0=u_1+1}^{t} \prod_{\substack{v_{\kappa-w}=1;\\ v_{\kappa-w} \neq u_{\kappa-w}}}^{u_{\kappa-w}} \prod_{\substack{v_{\kappa-w-1}=u_{\kappa-w}+1;\\ v_{\kappa-w-1} \neq u_{\kappa-w-1}}}^{u_{\kappa-w-1}} [\dots] \prod_{\substack{v_1=u_2+1;\\ v_1 \neq u_1}}^{u_1} \prod_{v_0=u_1+1}^{u_0} \left(1 - A_{\max\{w;b_{t-u_\kappa+1}\}}\right) A_{\max\{w;b_{t-v_\kappa+1}\}} \\
&\left[\left(1 - A_{\max\{w+1;b_{t-u_{\kappa-1}+1}\}}\right) A_{\max\{w+1;b_{t-v_{\kappa-1}+1}\}} \left[[\dots] \cdot \left[\left(1 - A_{\max\{\kappa-1;b_{t-u_1+1}\}}\right) A_{\max\{\kappa-1;b_{t-v_1+1}\}} A_{max\{2;\, b_{t-v_0+1}\}}\right]\right]\right]
\end{aligned} \tag{17}$$

is minimized if b_t minimizes $A_{\max\{w;b_t\}}$ in each period.

For $t+1$, it follows that

$$\begin{aligned}
&\sum_{u=1}^{t+1}\prod_{v=1}^{u} A_{\max\{w;b_{t-v+2}\}} + \sum_{u_1=1}^{t}\sum_{u_0=u_1+1}^{t+1}\prod_{\substack{v_1=1;\\ v_1\neq u_1}}^{u_1}\prod_{v_0=u_1+1}^{u_0}\left(1-A_{\max\{w;b_{t-u_1+2}\}}\right)A_{\max\{w;b_{t-v_1+2}\}}\,A_{max\{w+1;\,b_{t-v_0+2}\}}\\
&+\sum_{u_2=1}^{t-1}\sum_{u_1=u_2+1}^{t}\sum_{u_0=u_1+1}^{t+1}\prod_{\substack{v_2=1;\\ v_2\neq u_2}}^{u_2}\prod_{\substack{v_1=u_2+1;\\ v_1\neq u_1}}^{u_1}\prod_{v_0=u_1+1}^{u_0}\left(1-A_{\max\{w;b_{t-u_2+2}\}}\right)A_{\max\{w;b_{t-v_2+2}\}}\\
&\left[\left(1-A_{\max\{w+1;b_{t-u_1+2}\}}\right)A_{\max\{w+1;b_{t-v_1+2}\}}A_{max\{w+2;\,b_{t-v_0+2}\}}\right]+[\dots]\\
&+\sum_{u_{\kappa-w}=1}^{t+1-(\kappa-w)}\ \sum_{u_{\kappa-w-1}=u_{\kappa-w}+1}^{t+1-(\kappa-w-1)}[\dots]\sum_{u_1=u_2+1}^{t}\sum_{u_0=u_1+1}^{t+1}\prod_{\substack{v_{\kappa-w}=1;\\ v_{\kappa-w}\neq u_{\kappa-w}}}^{u_{\kappa-w}}\ \prod_{\substack{v_{\kappa-w-1}=u_{\kappa-w}+1;\\ v_{\kappa-w-1}\neq u_{\kappa-w-1}}}^{u_{\kappa-w-1}}[\dots]\prod_{\substack{v_1=u_2+1;\\ v_1\neq u_1}}^{u_1}\prod_{v_0=u_1+1}^{u_0}\left(1-A_{\max\{w;b_{t-u_\kappa+2}\}}\right)A_{\max\{w;b_{t-v_\kappa+2}\}}\\
&\left[\left(1-A_{\max\{w+1;b_{t-u_{\kappa-1}+2}\}}\right)A_{\max\{w+1;b_{t-v_{\kappa-1}+2}\}}\left[[\dots]\cdot\left[\left(1-A_{\max\{\kappa-1;b_{t-u_1+2}\}}\right)A_{\max\{\kappa-1;b_{t-v_1+2}\}}A_{max\{2;\,b_{t-v_0+2}\}}\right]\right]\right]
\end{aligned} \tag{18}$$

is to be minimized. From equation (17), we know that the blocking option is optimally chosen until period t. Factoring out $A_{\max\{w;b_{t+1}\}}$ following the same structure as in (12), (14) and (16), the remaining terms for the periods with optimally chosen blocking option in brackets would be equivalent for any blocking strategy in $t+1$. Thus, the lowest $A_{\max\{w;b_{t+1}\}}$ leads to the optimal blocking strategy in $t+1$. These arguments complete the proof that the myopic view is optimal for one single slot type ($w_l = w_h$). □

I. Results of Full Factorial Analysis: Performance of Decision Rules

In Table 11, we show the performance of the decision rules regarding the whole data set. Analyzing the performance of the decision rules by comparing them to the sdp, we keep the sdp as the basis to be able to directly compare the performance of the rules for specific parameter settings. We indicate the case, where both, the considered rules and the sdp, result in costs (negative overall value), with $^{--}$, as in Appendix G and interpret the results as a cost increase. In case, both considered overall values are positive, we indicate the profit decline with $^{++}$. The instance with a negative value (cost) for the decision rule but a positive value (profit) for the sdp, we indicate the profit decline (resulting in cost) with **. Remember that $w_l = 1$.

Table 11: Performance of the decision rules (Δ% decline compared to the sdp) regarding all instances, with low=R_{low}, One=R_{one}, My.=R_{myopic}, No=R_{no}

	Overall			
Rule	**Low**	**One**	**My.**	**No**
Overall	795^{--}	386^{--}	2.25^{--}	3178^{--}
$\beta = 0$	98.88^{++}	234**	0	0
$\beta = 0.5$	137**	351**	2.31^{++}	1.19^{++}
$\beta = 1$	953^{--}	2946^{--}	39.46^{--}	322^{--}
$\beta = 1.5$	43.63^{--}	143^{--}	1.30^{--}	130^{--}
$\beta = 3$	90.18^{--}	7.12^{--}	0.02^{--}	6201^{--}
$\beta = 6$	720^{--}	0	0	3284^{--}

	$w_h = 1$				$w_h = 2$				$w_h = 5$			
Rule	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**
Overall	-	88.54^{--}	0	509^{--}	3544**	1523**	30.59^{++}	21426**	12.46^{++}	9.07^{++}	0.14^{++}	128**
$\beta = 0$	-	4299**	0	0	69.52^{++}	170**	0	0	13.89^{++}	16.89^{++}	0	0
$\beta = 0.5$	-	512^{--}	0	0	74.56^{++}	219**	4.38^{++}	1.50^{++}	11.77^{++}	14.38^{++}	0.54^{++}	0.55^{++}
$\beta = 1$	-	136^{--}	0	5.70^{--}	126**	480**	24.44^{++}	138**	9.43^{++}	11.51^{++}	0.22^{++}	2.12^{++}
$\beta = 1.5$	-	52.06^{--}	0	34.51^{--}	104^{--}	387^{--}	13.15^{--}	640^{--}	6.84^{++}	8.18^{++}	0.08^{++}	5.38^{++}
$\beta = 3$	-	52.71^{--}	0	308^{--}	134^{--}	10.11^{--}	0.11^{--}	1244^{--}	3.59^{++}	0.78^{++}	0.00^{++}	45.07^{++}
$\beta = 6$	-	0	0	1299^{--}	1156^{--}	0	0	7737^{--}	30.02^{++}	0	0	790**

	$r_s = 0.5$				$r_s = 1$				$r_s = 2$			
Rule	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**	**Low**	**One**	**My.**	**No**
Overall	22.03^{--}	11.25^{--}	0.08^{--}	84.60^{--}	52.09^{--}	25.35^{--}	0.15^{--}	208^{--}	9.83^{++}	4.31^{++}	0.01^{++}	42.15^{++}
$\beta = 0$	12.81^{--}	28.38^{--}	0	0	33.97^{--}	80.08^{--}	0	0	3.27^{++}	9.08^{++}	0	0
$\beta = 0.5$	9.16^{--}	21.62^{--}	0.17^{--}	0.06^{--}	21.77^{--}	55.33^{--}	0.36^{--}	0.18^{--}	2.37^{++}	7.36^{++}	0.03^{++}	0.04^{++}
$\beta = 1$	5.09^{--}	14.10^{--}	0.24^{--}	1.35^{--}	10.64^{--}	32.67^{--}	0.43^{--}	3.55^{--}	1.26^{++}	5.17^{++}	0.03^{++}	0.71^{++}
$\beta = 1.5$	2.82^{--}	8.18^{--}	0.10^{--}	6.81^{--}	5.41^{--}	17.58^{--}	0.16^{--}	16.05^{--}	0.71^{++}	3.16^{++}	0.01^{++}	3.43^{++}
$\beta = 3$	10.03^{--}	0.87^{--}	0.00^{--}	69.36^{--}	21.72^{--}	1.71^{--}	0.00^{--}	150^{--}	5.49^{++}	0.37^{++}	0.00^{++}	37.28^{++}
$\beta = 6$	84.29^{--}	0	0	380^{--}	182^{--}	0	0	831^{--}	47.91^{++}	0	0	223**

Note. High percentage differences (>100%) are rounded to the closest integer value.

J. Exemplary Instance: Myopic Rule Suboptimal

We exemplarily consider one instance with a system size of three slots that are distributed to two slot types and a minor impact of the quality inference effect.[7] In that instance, the myopic rule leads to an expected profit of 0.65, whereas the dynamic program yields an expected profit of 0.66. The related no-choice probabilities are presented in Table 12. Note that the booking status $\vec{s} = (d_1, d_2, q_1, q_2)$ denotes the number of booked slots of type n by d_n and the number of blocked slots of type n by q_n. The booking status $\vec{s} = (0,0,0,1)$ thus states that the two prioritized slots of slot type 1 are offered while the less preferred slot of slot type 2 is blocked. With booking status $\vec{s} = (0,0,2,1)$, no slot is

[7] Parameter of the instance: $\kappa = 3; \kappa_1 = 2; \kappa_2 = 1;\ \hat{\tau} = 6;\ \alpha = 0.5;\ r_p = 1;\ r_s = 0.5;\ \pi = 2;\ c_o = 0.5;\ c_u = 1;\ \beta = 0.5; w_1 = 2;\ w_2 = 1;\ \lambda = 2.25$

offered, which automatically leads to the no-choice option. The no-choice probabilities do not change in case a slot of one slot type is blocked $\vec{s} = (0,0,0,1)$ or booked $\vec{s} = (0,1,0,0)$. Thus, the remaining possible states are left out in Table 12 for reasons of redundancy regarding the probabilities.

Table 12: No-choice probabilities for each possible booking status $\vec{s}$ of the exemplary instance, with $\vec{s} = (d_1, d_2, q_1, q_2)$, $\kappa_1 = 2$ and $\kappa_2 = 1$.

Booking status	No-choice probability (%)
(0,0,0,1)	7.99
(0,0,0,0)	8.61
(0,0,1,0)	11.27
(0,0,1,1)	11.92
(0,0,2,0)	26.89
(0,0,2,1)	1

The minimum no-choice probability (7.99%) is myopically achieved when the two slots of slot type 1 are offered and the slot of slot type 2 is either booked or blocked. In case, all slots are available, the myopic rule thus recommends blocking the slot of slot type 2, $\vec{s} = (0,0,0,1)$, regardless of the remaining time. In contrast, the optimal recommendation (of the dynamic program) is time-dependent. As long as the time horizon until the day of service is large enough ($t \geq 3$), no blocking is recommended, $\vec{s} = (0,0,0,0)$. During that time, decreasing the likelihood that a less preferred slot is offered during the remaining time (with a relatively high no-choice probability) outweighs the decrease of the no-choice probability in the given period. This reverses in the last two periods of the booking horizon before the day of service ($t = 1,2$ with the day of service in $t = 0$), as not much time is left to book the less preferred slot by chance. It is then optimal to block the slot of slot type 2 if all slots are available (in line with the myopic rule).