

## ORIGINAL ARTICLE

# Decision biases in revenue management revisited: Dynamic decision-making under stationary and nonstationary demand

Catherine Cleophas<sup>1</sup>  | Claudia Schüetze<sup>2</sup>

<sup>1</sup>Christian-Albrechts-Universität zu Kiel, Kiel, Germany

<sup>2</sup>RWTH Aachen University, Aachen, Germany

## Correspondence

Catherine Cleophas,  
Christian-Albrechts-Universität zu Kiel,  
Olshausenstraße 40, 24098 Kiel, Germany.  
Email: [cleophas@bwl.uni-kiel.de](mailto:cleophas@bwl.uni-kiel.de)

[The copyright line for this article was changed on June 21, 2022 after original publication].

## Abstract

State-of-the-art revenue management systems combine forecasting and optimization algorithms with human decision-making. However, only a few existing contributions consider the behavioral aspects of revenue management. To extend the related research, we examine the impact of nonstationary demand and two dynamic decision tasks. We examine human decision-making strategies and biases by implementing a related experimental design in a laboratory study and comparing participant decisions to systematic heuristics. Our results highlight that participants struggle to accommodate a nonstationary willingness to pay. In that, they exhibit a combination of optimism and loss aversion biases. We further find that participants anchor their decisions on customers' willingness to pay. We draw implications and further research opportunities to behaviorally inform the design of symbiotic analytics systems from these results.

## KEYWORDS

analytics, behavioral operations research, pricing, revenue management

## 1 | INTRODUCTION

Since the 1980s, revenue management (RM) has been a crucial success factor for service industries, such as hotels, car rentals, and airlines. It lets firms maximize revenue by predicting demand and optimizing the offered price or the offered set of fare classes. A large and continuously growing body of research focuses on extending models and algorithms to improve predictive and prescriptive analytics for RM—recent research surveys include Gönsch (2017) and Klein et al. (2020).

Industry RM systems let analysts adjust model parameters, demand forecasts, and offers. In other words, they let humans override both inputs and outcomes. This idea became clear to the authors in discussions and demonstrations with experts from RM departments and software suppliers and has been considered in some recent contributions on organizational RM aspects (compare Fürstenau et al., 2020; Schütze et al., 2020). Our research is mainly motivated by research-industry cooperation involving a major European airline that had implemented and funded RM research for several years at the time of the project. One of the coau-

thors served as an airline IT consultant prior to the start of the project; another coauthor took part in an extended internship with the operational RM department. In this cooperation, we had the opportunity to observe and support projects to overhaul both RM analytics systems and organizational structures.

In these projects, we took part in stakeholder workshops collecting requirements, moderated simulation-based war games, and supported training on analytics methodology. Throughout these areas, discussions on the need for analysts to continue to be able to intervene in all process steps motivated us to investigate the existing state of research on behavioral RM. Yet, few existing contributions examine human decision-making for RM from a behavioral perspective. Existing contributions tend to consider demand as stationary (compare designs based on Bearden et al., 2008) or let decision-makers implement static controls without dynamic adaptations over the booking horizon (compare Kocabiyikoglu et al., 2015). Nevertheless, discussions on particularly challenging aspects of RM caused us to consider the implications of nonstationary demand and alternative decision tasks. Theoretical RM literature broadly accepts the idea

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Decision Sciences* published by Wiley Periodicals LLC on behalf of *Decision Sciences* Institute.

that customer demand is *nonstationary* over the booking horizon (Kimes, 1989) and that RM decisions are *dynamic*. To accommodate these aspects, practical RM often relies on dynamic price-based controls, such as bid-prices (Talluri & van Ryzin, 2004, chapter 3.2). Finally, our participation in industry projects on designing RM interfaces offered insights into analyst strategies when pursuing secondary objectives, such as a minimum load. Practitioners' descriptions of such strategies inspired us to compare human decision-making to the patterns created by automated solution approaches.

To contribute to behavioral research on RM analytics, we propose considering human decision-making when customers' arrival rate, willingness to pay, or both vary over the booking horizon. When analysts set controls rather than deciding on individual acceptance, they do not know the customers' actual willingness to pay. Therefore, we further compare dynamic acceptance and pricing tasks. Different decision tasks result in divergent optimal results, as price-setting limits the revenue gain to the price, whereas acceptance decisions can yield the customers' complete willingness to pay.

As empirical contributions, we replicate and cross-validate findings from Bearden et al. (2008) in an experimental study, following the principles outlined in Pagell (2021). Specifically, we show that a nonstationary willingness to pay negatively impacts participants' decision-making, while a nonstationary customer arrival rate has no significant effect. At the same time, nonstationary demand appears to offer more potential for participants to benefit from repeated experience. Furthermore, we find that, under stationary demand, the pricing task also reduces participants' success compared to the acceptance task. We have extensively discussed the experimental design and results with airline RM experts. In particular, these experts also suggested implementing the minimum-sale heuristics to represent a practical loss-averse RM strategy. By comparing participants' decisions to systematic solution approaches, we analyze underlying biases and find evidence for anchoring in pricing decisions. The paper concludes with managerial insights and opportunities for future research to inform the design of symbiotic analytics systems in general and RM systems in particular.

## 2 | LITERATURE REVIEW

Existing behavioral research on RM frequently focuses on demand forecasting. For example, Zeni (2003) evaluates analysts' contribution through a field study. In a later study, Mukhopadhyay et al. (2007) focus on forecast adjustments, analyzing data on revenue and bookings at a U.S. airline. Both contributions find that a complementary system of analysts and algorithms can outperform a fully automated system.

Research on judgmental forecasting is not limited to RM, as exemplified by the pertinent overview in Petropoulos et al. (2016). Related contributions may inform related research on

decision-making. For example, Eroglu and Croxton (2010) highlight human forecasters' susceptibility to systematic optimism, anchoring, and overreaction biases.

Beyond revenue maximization, RM is frequently understood to include revenue sharing and contract design problems. Examples for related behavioral research include Katok and Wu (2009), Niederhoff and Kouvelis (2019), and Tokar et al. (2016). However, when considering the underlying decision problems from a behavioral perspective, there is no direct connection, as revenue sharing focuses more on questions related to cooperation and negotiation, whereas short-term revenue maximization focuses more on individual decisions.

Here, we build on a set of related efforts examining the specific task of deciding at what price and when customers get to buy units of capacity. Table 1 summarizes the research objectives and settings of closely related contributions and delineates the presented research.

Bearden et al. (2008) propose a first and impactful experimental design to analyze RM decision biases under stationary demand. The authors contrast *accept errors*, where participants wrongly accept bookings, and *reject errors*, where participants wrongly reject bookings. They conclude that decision-makers rely on a sophisticated policy, accounting for the left-over capacity and the time remaining within the booking horizon.

Bendoly (2011) combines the experimental design from Bearden et al. (2008) with a real-world decision support system to measure stress levels via physiological markers. This study finds that high capacity levels left at the end of the booking horizon and the number of simultaneous tasks increase stress and induce decision errors. Bendoly (2013) conducts a similar experiment with hotel employees and finds that different levels of feedback influence revenue performance.

Kocabiyyikoglu et al. (2015) implement a more static setting, where participants have to allocate capacity among two customer segments before the start of the booking horizon. The authors use this setting to compare the RM decision to a normatively equivalent version of the newsvendor problem. Kocabiyyikoglu et al. (2018) rely on the same static setting to investigate different RM decision tasks. Their design asks participants to set static booking limits or protection levels, given stationary demand for two fare classes. The authors analyze whether decision-makers resort to anchoring on mean expected values, terming this the *pull-to-center effect*. They conclude that this effect does not reliably occur in RM. We follow up on this idea by confronting participants with dynamic decision tasks and finding evidence for anchoring when participants set prices.

The newsvendor problem is frequently regarded as strongly related to a static version of RM and has triggered a significant body of behavioral research. Its central question, how much capacity to reserve for uncertain demand, is closely related to the quantity-based RM problem when booking limits are set before the booking horizon and are not updated dynamically within the booking hori-

TABLE 1 Related work on behavioral RM

	Demand arrival	Willingness to pay	Decision
<b>Bearden et al. (2008):</b> systematic decision errors	constant	constant	dynamic acceptance or static threshold
<b>Bendoly (2011):</b> effects from concurrent tasks and stress	constant	constant	dynamic acceptance or static threshold
<b>Bendoly (2013):</b> effects from feedback	constant	constant	dynamic acceptance or static threshold
<b>Kocabiyikoglu et al. (2015):</b> revenue management versus newsvendor decision	constant	constant per segment	static booking limits
<b>Kocabiyikoglu et al. (2018):</b> booking limit versus protection-level decisions	constant	constant per segment	static booking limits or static protection levels
<b>This paper:</b> effects from nonstationary demand and dynamic decisions	constant or triangular	constant or increasing	dynamic acceptance or dynamic threshold

zon. However, when Kocabiyikoglu et al. (2015) compare newsvendor decisions to mathematically equivalent RM decisions, they conclude that the related insights are not readily transferable. For the newsvendor problem, Bolton and Katok (2008) investigate learning environments; they find that hands-on experience, for example, through simulation, can improve learning. Benzion et al. (2008) also analyze newsvendor learning, focusing on different profit levels and demand distributions. In this paper, we observe that decision-makers do improve over repeated booking horizons under nonstationary demand.

Several existing studies analyze decision-making under normatively equivalent but behaviorally diverging settings: for example, Ho and Zhang (2008) find that participants' decisions differ when pricing a fixed fee as a discount or a two-part tariff; Katok and Wu (2009) show that buyback and revenue sharing contracts change participants' behavior; Tokar et al. (2016) propose ways how framing can help to reduce participants' decision biases when supply chain managers make inventory control decisions. However, the RM decision tasks of dynamic acceptance or pricing are not fully normatively equivalent, as participants can earn the customers' entire willingness to pay for acceptance while only earning the asking price for pricing tasks.

In summary, existing experimental designs for RM assume stationary demand and ask participants to either dynamically accept individual customers or to set static thresholds or limits. We extend this research by introducing nonstationary demand and a dynamic pricing decision. In replicating and extending results from Bearden et al. (2008), we follow the call to action given in Pagell (2021). In evaluating the implications of the extended design as compared to results from the original design, we implement what Schoenherr and Swink (2012) term cross-validation. In an experimental study, we evaluate decision biases, such as anchoring, which have not previously been examined for RM as dependent on the demand setting and decision task.

### 3 | THEORY

This section first formalizes the considered RM problem and two alternative decision tasks. We expect two aspects of the RM problem to increase its complexity for human decision-makers: nonstationary demand and a dynamic pricing decision. Accordingly, we consider customer arrivals and willingness to pay as stochastic parameters,  $p_t$  and  $w_t$ , which depend on time  $t$ .

Because we expect human decision-making to rely on systematic biases and heuristics, we secondly outline systematic approaches to solve the RM problem. Third, we list measurement variables to describe RM decision-making, where some of these variables rely on comparisons to outcomes from systematic approaches. Finally, we formulate the hypotheses that guide our research based on the RM model. These hypotheses consider human decision making with particular regard for the RM problem introduced here and describe the expected development of the measurement variables as based on existing behavioral theories.

#### 3.1 | The RM problem with nonstationary demand

Following Talluri and van Ryzin (2004, p. 59), we consider a limited booking horizon defined by a set of time slices  $t \in \{T, \dots, 0\}$ , where at most one customer arrives and decides to book per time slice. This model also underlies Bearden et al. (2008) and Bendoly (2011, 2013). Arrival rate  $p_t \in [0, 1]$  indicates the probability of a customer arriving in time slice  $t$ .

We model the dynamic RM decision generally as valuing the next unit of capacity to be sold by setting a threshold  $R_{t,s}$ , which depends on time  $t$  and left-over capacity  $s \in \{0, \dots, S\}$ . When the booking horizon ends, at  $t = 0$ , the value of capacity perishes.

The *expected willingness to pay* of the customer arriving at  $t$  is  $\hat{w}_t$  and is uniformly distributed in the interval  $[\hat{w}_t^l, \hat{w}_t^u]$ . We denote the *actual willingness to pay* of a customer arriving at  $t$  by  $w_t$ . A customer that arrives at time  $t$ , given left-over capacity  $s$ , books one unit of capacity if their willingness to pay at least equals  $R_{t,s}$ . Given threshold  $R_{t,s}$ , a booking yields revenue  $r_t(R_{t,s})$ . When the threshold is at most equal to the expected willingness to pay, we expect the customer to book with certainty:  $R_{t,s} \leq \hat{w}_t^l \rightarrow b_t(R_{t,s}) = 1$ . When the threshold exceeds the upper bound of the expected willingness to pay, we do not expect the customer to book:  $R_{t,s} > \hat{w}_t^u \rightarrow b_t(R_{t,s}) = 0$ . If the threshold falls within the expected willingness to pay interval, the booking probability,  $b_t(R_{t,s})$  is given by

$$b_t(R_{t,s}) = \begin{cases} 1 & \text{if } R_{t,s} < \hat{w}_t^l, \\ \frac{\hat{w}_t^u - R_{t,s}}{\hat{w}_t^u - \hat{w}_t^l} & \text{if } \hat{w}_t^l \leq R_{t,s} \leq \hat{w}_t^u, \\ 0 & \text{if } R_{t,s} > \hat{w}_t^u. \end{cases} \quad \forall t \in \{T, \dots, 0\}, \quad s \in \{0, \dots, S\} \quad (1)$$

### 3.2 | RM decision tasks

When making acceptance decisions, decision-makers know the customer's willingness to pay. If a customer is accepted, the resulting revenue  $r_t(R_{t,s})$  corresponds to their willingness to pay. Thus, the expected revenue for the *acceptance decision*,  $\hat{r}_t^{acc}(R_{t,s})$ , is at least the threshold  $R_{t,s}$  and at most the upper bound of the willingness to pay interval  $\hat{w}_t^u$ :

$$\hat{r}_t^{acc}(R_{t,s}) = \frac{\hat{w}_t^u + R_{t,s}}{2}. \quad (2)$$

When setting prices ahead of customer arrivals, decision-makers must rely on the expected willingness to pay. The resulting price corresponds to the threshold  $R_{t,s}$  from the acceptance task. Customers book if their willingness to pay exceeds this price, such that the revenue for the *price decision*,  $r_t^{pr}(R_{t,s})$ , corresponds to the threshold:

$$r_t^{pr}(R_{t,s}) = R_{t,s}. \quad (3)$$

Acceptance and pricing decisions do not earn the same revenue even when applying the same solution approach: The price decision forfeits the difference between the optimal threshold and the customer's actual willingness to pay.

### 3.3 | Systematic solution approaches

We formalize several systematic RM solution approaches to evaluate human decision-making and explore the conse-

quences of systematic biases. Applying any of these to a stream of customer arrivals yields a pattern of bookings and earned revenue.

#### Optimal dynamic solution (OPTIMAL)

To optimally exploit demand over the booking horizon, this approach calculates ex-ante optimal thresholds  $R_{t,s}$  per combination of left-over capacity  $s$  and time  $t$  via a dynamic program as described in Talluri and van Ryzin (2004, chapter 2.2).

#### Optimal static solution (STATIC)

The STATIC approach applies a single revenue-optimal threshold across the entire booking horizon,  $R^{stat}$ .

#### Anchoring (ANCHOR- $\hat{w}_t$ )

This heuristic anchors on the expected willingness to pay per time slice,  $\hat{w}_t$ , to set the acceptance threshold or the price. In this, it implements the pull-to-center strategy discussed in Kocabiyikoglu et al. (2018).

#### Demand chasing (CHASING)

This heuristic first sets the threshold or the price to the expected willingness to pay,  $R_{t,s} = \hat{w}_t$ . After each customer arrival at time  $t'$ , the heuristic updates  $R_{t,s}$  to the newly observed willingness to pay,  $R_{t,s} = w_{t'}$ . In this, it implements "demand chasing" as discussed in Lau and Bearden (2013).

#### First-come-first-serve (FCFS)

This heuristic indiscriminately sells the given capacity. In the *pricing task*, FCFS sets the price to the lowest expected willingness to pay,  $R_{t,s} = \min_{\forall t} \hat{w}_t^l$ . Thereby, it implements the ultimate loss-averse strategy.

#### Minimum-sale (MIN- $\phi$ -ANCH, MIN- $\phi$ -OPT)

This heuristic aims to sell at least  $\phi$  units of capacity via FCFS before switching to ANCHOR- $\hat{w}_t$  (MIN- $\phi$ -ANCH) or to OPTIMAL (MIN- $\phi$ -OPT). It models a loss-averse strategy and is inspired by our discussions with practitioners outlined in the Introduction.

#### Random acceptance (RANDOM)

We model random *acceptance* by computing the gain expected from accepting any arriving customer with 0.5 probability for as long as there is left-over capacity,  $r_t = 0.5 * w_t$  if  $s > 0$ , else  $r_t = 0$ . There is no equivalent *pricing* strategy.

### 3.4 | Measurement variables

#### Benchmark gaps

We compute percentage gaps in revenue from any observed RM strategy to the optimal dynamic strategy (*optimality gap*) and the optimal static strategy (*static benchmark gap*).

### Error types

We count the number of accept and reject errors by comparing observed decisions to those prescribed by OPTIMAL for the current combination of expected demand and left-over capacity. If a decision deviates from the current optimal strategy by accepting a customer or offering an acceptable price, there results an *accept error*. If a decision erroneously rejects a customer or offers an excessive price, this is a *reject error*. Both early accept and reject errors affect the left-over capacity, which may cause additional errors to arise later in the booking horizon. Accept errors limit the capacity and thereby increase the later tendency for reject errors; reject errors leave more capacity unsold and increase the later tendency for accept errors.

### Implicit thresholds

While price decisions directly implement the underlying threshold, we estimate the implicit thresholds that underlie acceptance decisions following Bearden et al. (2008). These thresholds embody decision-makers' assumptions about the future value of capacity. For each observed acceptance decision, we calculate the threshold  $R_{t,s}^{part}$  that maximizes the proportion of correctly predicted decisions per left-over capacity and day of the booking horizon. To eliminate outliers, we only calculate  $R_{t,s}^{part}$  for days in the booking horizon when we recorded at least 2% of the decisions for that capacity. When we exclusively observe accepted or rejected customers for one combination of day and left-over capacity,  $R_{t,s}^{part}$  is either an upper bound for accept decisions or a lower bound for reject decisions. In this case, we rely on a best-case estimate based on the optimal strategy  $R_{t,s}$ . If  $R_{t,s}$  correctly predicts the decision, we set  $R_{t,s}^{part} = R_{t,s}$ . Otherwise, we set  $R_{t,s}^{part}$  to a value that is as close as possible to  $R_{t,s}$  while still maximizing the proportion of correctly predicted decisions.

### Threshold descriptors

We distinguish two subsets of observed decisions:  $R_t^{part} \geq \hat{w}_t$ , where implicit thresholds or prices equal or exceed the willingness to pay, and  $R_t^{part} < \hat{w}_t$ , where they undercut it. We further quantify the difference between thresholds  $R_{t,s}^{part}$  and the expected willingness to pay  $\hat{w}_t$  via the mean absolute percentage deviation (MAPD):

$$MAPD^{\hat{w}_t} = \frac{100}{n} \sum_{t=1}^n \left| \frac{R_{t,s}^{part} - \hat{w}_t}{R_{t,s}^{part}} \right|. \quad (4)$$

### Decision bias

To quantify the deviation of a set of observed decisions from the optimal strategy  $R_{t,s}$ , we calculate the BIAS. In the following equation,  $n$  denotes the number of days of the booking horizon when we recorded decisions for that capacity:

$$BIAS = \frac{1}{n} \sum_{t=1}^n (R_{t,s}^{part} - R_{t,s}). \quad (5)$$

### Revenue loss

For acceptance decisions, the implied revenue loss  $E_t^A$  is the absolute difference between the optimal threshold  $R_{t,s}$  and the actual willingness to pay  $w_t$ :  $E_t^A = |R_{t,s} - w_t|$ . For price-setting, the implied revenue loss  $E_t^P$  is the absolute difference between the optimal price  $R_{t,s}$  and the price  $R_t^{part}$  as set by the participant:  $E_t^P = |R_{t,s} - R_t^{part}|$ .  $\bar{E}$  describes the average implied revenue loss over  $t$  and all rounds.

### Revenue distance

To compare sets of decisions across a booking horizon, we evaluate the *revenue pattern*, which shows the earned revenue  $r_t$  for any time slice where a customer booked. To compute the absolute distance between any two patterns  $P$  and  $P'$  and normalize it per booking horizon, we divide the sum of absolute distances by the sum of the willingness to pay of all arriving customers in that booking horizon:

$$D_{P,P'}^r = \frac{\sum_{t=1}^n |r_t^P - r_t^{P'}|}{\sum_{t=1}^n w_t}. \quad (6)$$

## 3.5 | Hypotheses

First, we derive hypotheses on the implications of non-stationary demand and dynamic price-setting. Second, we hypothesize that human decision-makers employ systematic heuristics in line with Katsikopoulos and Gigerenzer (2013), causing related decision biases. Note that this subsection assumes the RM problem to align with the model described in Section 3.1 and to be solvable using the approaches described in Section 3.3. Furthermore, we formulate hypotheses in terms of the measurement variables described in Section 3.4.

Some behavioral research, for example, Kremer et al. (2011), finds that humans cope better with evolving rather than stationary environments when adjusting time-series forecasts. However, the RM problem asks decision-makers to dynamically set controls while anticipating future developments in customers' arrival rates and willingness to pay. This problem corresponds to making intertemporal choices following Hardisty et al. (2013). Following research on human error in gauging probabilistic developments (Kahneman and Frederick, 2002; Tversky and Kahneman, 1992) and intertemporal choices (Hardisty et al., 2013), we expect that human decision-makers struggle in this situation. Following Hilbert (2012), there results in a greater gap between the input evidence (the forecast) and the output estimate (decisions). In terms of the measurement variables, we expect the optimality gap from human decisions to increase when customers' willingness to pay  $w_t$  or arrival rate  $p_t$  vary over  $t$  compared to when these parameters are stationary. This causes us to formulate hypothesis 1a:

**Hypothesis 1a)** *Nonstationary demand reduces the success of human decision-making for RM.*

While the general idea may be intuitive, this hypothesis further justifies the idea of evaluating RM decisions under nonstationary demand: If the resulting complexity did not change the success of human decision-making, a more straightforward design (e.g., based on Bearden et al. (2008)) could yield the same insights.

As outlined comprehensively in Hinterhuber (2015), price-setting is notoriously susceptible to biases. Comparing it to the acceptance task highlights some underlying causes: As pointed out in Section 3.2, the pricing decision has to rely on the demand forecast without customer-specific information entirely. While this is true for both automated solution approaches and human decision-makers, related research shows that having to decide under uncertainty induces stress in humans (De Berker et al., 2016). As Bendoly (2011) highlights, stress negatively affects human RM decisions.

Furthermore, as opposed to the *binary* decision task of accepting or rejecting customers, both pricing and the setting of booking limits constitute a *multiary* decision (Hilbert, 2012). Multiary decisions make it harder for decision-makers to link input evidence and output estimate. There emerges a “wicked learning environment” (Hogarth et al., 2015), as feedback from a pricing decision does not include how much more the customer might have been willing to pay and as a multitude of alternative pricing decisions (as opposed to the single alternative to acceptance) are possible.

Last, the pricing decision also includes an element of self-fulfilling prophecies (Kollwitz and Papathanassis, 2011): Setting a lower price is more likely to induce positive feedback in the form of earned revenue. Therefore, we expect this decision task to cause larger optimality gaps, as stated by hypothesis 1b:

**Hypothesis 1b)** *Setting prices reduces the success of human decision-making for RM.*

Intuitively, we might expect positive effects from learning as decision-makers confront the RM problem over multiple booking horizons and can develop a learning curve (Glock et al., 2019). However, Bearden et al. (2008) do not find positive effects from experience for acceptance decisions. This may be due to each booking horizon being unique due to stochastic customer arrivals, making it difficult for decision-makers to assess a new decision’s similarity to a known situation, as also described in Hogarth et al. (2015). Furthermore, when relying on demand forecasts detailing arrival rates and expected willingness to pay, human decision-makers may suffer from the description-experience gap described, for example, in Hertwig et al. (2018, 2019). Accordingly, we also do not expect to see positive effects when participants face the problem repeatedly without explicit reflection or deliberate feedback. Specifically, we expect the optimality gap to remain stable even when a decision-maker repeatedly attempts to solve the RM problem in the same setting, stating hypothesis 1c:

**Hypothesis 1c)** *The more complex the situation, the more difficult it becomes for humans to get better at solving the RM problem through mere repetition.*

Following Bearden et al. (2008), we expect participants to be overly optimistic about future demand, thereby overestimating the value of left-over capacity. When decision-makers fall prey to optimism bias, they overshoot optimal thresholds and prices as described for forecast adjustments in Eroglu and Croxton (2010) and transport planning in Flyvbjerg (2008).

Flyvbjerg (2008) claims that relying on reference classes from historical data, such as given by demand forecasts in RM, can alleviate this tendency. However, following the previously stated hypotheses, RM decisions under nonstationary demand represent intertemporal choices with a gap between input and output, creating wicked learning environments (Hogarth et al., 2015). Furthermore, deciding on what to offer rather than setting a forecast may increase decision-makers’ perception of control, which also increases optimism bias following Tyebjee (1987). Therefore, we expect optimism bias to persist even given accurate forecasts. Accordingly, we expect implicit thresholds or prices set by human decision-makers to exceed those from the optimal solution, causing a positive BIAS and reject errors. This constitutes hypothesis 2a:

**Hypothesis 2a)** *Human RM decisions are susceptible to optimism bias.*

Somewhat contradicting the idea of optimism bias, Schweitzer and Cachon (2000) find that humans underestimate opportunity costs for the newsvendor problem, possibly due to loss aversion. Hinterhuber (2015) highlights the resulting observations of underpricing, particularly concerning new products, where sellers are trying to explore the market.

In our setting, loss aversion would induce decision-makers to set thresholds or prices that undercut the optimal solution, causing a negative BIAS and accept errors. As Bearden et al. (2008) note, loss aversion may combine with optimism bias for larger values of left-over capacity; the authors term this phenomenon *inventory mis-sensitivity*. We expect loss aversion to manifest through implicit thresholds or prices that undercut the optimal solution, causing a negative BIAS and accept errors. This leads us to formulate hypothesis 2b:

**Hypothesis 2b)** *Human RM decisions are susceptible to loss aversion.*

The previously stated hypotheses indicate that RM decision-makers follow a systematic but flawed strategy. This idea goes with the idea of price-setting following overly simplistic heuristics described in Hinterhuber (2015). Following Epley and Gilovich (2006), we expect that such heuristics include relying on *anchors* to decide which customers to accept or what prices to set. Furnham and Boo (2011) summarize relevant literature describing the anchoring bias and its implications for decision-makers. As the authors state, “people are influenced by specific information given before a judgment.”

For newsvendor decisions, Kocabiyikoglu et al. (2018) expect to find a related pull-to-center effect but cannot confirm this in their static setting. In the dynamic RM problem, we expect that decision-makers are overly influenced by information on the expected or the most recently observed willingness to pay, neglecting other crucial param-

eters like arrival probabilities, capacity, and time left. Thus, we conclude with hypothesis 2c:

**Hypothesis 2c)** *Human RM decisions rely on anchors.*

When relying on the most recently observed willingness to pay as their anchor, decision-makers further exhibit demand chasing behaviors (Kirshner and Moritz, 2021). Demand chasing represents a particular type of anchoring bias that research frequently examines for inventory decisions and the related newsvendor problems (Kremer et al., 2010; Lau and Bearden, 2013). From the perspective of the systematic solution approaches formulated in Section 3.3, anchoring creates revenue patterns that resemble those from ANCHOR- $\hat{w}_t$  or those from CHASING.

## 4 | RESEARCH METHODS

Researchers may gain insight into RM analysts' decision-making via surveys, field observations, or examining RM decisions in a laboratory setting. In line with the references listed in Table 1, the research presented in the remainder of this paper relies on laboratory experiments.

This choice of research method is motivated by drawbacks of field surveys and observations when aiming to analyze decision biases and by one main advantage of laboratory experiments. As emphasized, for example, by Furnham and Boo (2011), humans are rarely aware of their decision biases. Accordingly, we do not expect survey answers to provide clear insights. Nevertheless, as outlined in the Introduction, we rely on talks with RM experts to inform the hypotheses and systematic solution approaches.

Second, the opportunity to observe operational RM experts in situ, as described in Schütze et al. (2020), is severely limited by company policy and labor councils. These seem to effectively prevent much empirical research on human analysts overruling analytics systems, with a few recent notable exemptions, such as Sun et al. (2021).

Third, when aiming to analyze decision-making for RM, field experiments suffer from a severe limitation even when overcoming these barriers: The real-world market environment is intransparent and dynamic. Accordingly, judging the quality of empirical RM decisions is difficult, as described in the literature on RM performance evaluation (Vock et al., 2022).

The laboratory setting lets us implement a controlled RM challenge by confronting both participants and systematic solutions with the same sets of arriving customers. This setting enables observations under *ceteris paribus* conditions: All participants in the same treatment confront the same information and the same customer arrivals. By varying the demand generation parameters, we can systematically vary customer arrival probability  $p_t$  and willingness to pay  $w_t$  to cross-validate hypotheses.

As participants in the related experiments, we recruit university students. This recruitment strategy creates a large pool of subjects, ensuring the statistical significance of experimental results. On a related note, Thomas (2011) emphasizes the

opportunity for homogeneous sampling to maximize precision and control in logistics experiments. Accordingly, we sample participants with a homogeneous degree of education and experience. Furthermore, student participants enable us to organize the experiments over a short period, excluding any long-term changes in the participant pool. Finally, we assume that findings from Bolton et al. (2012) for the newsvendor problem also apply to RM so that behavioral effects for students and managers are comparable.

We implement all treatments in the software z-Tree from the University of Zurich (Fischbacher, 2007). As emphasized in Schram and Ule (2019, chapter 17), z-Tree is widely used because it is free of charge and requires only little coding effort, as opposed to, for example, implementing a custom front- and back-end through PHP, MySQL, and Javascript. In addition, the client-server system enables closed-system-data-handling, as opposed to, for example, commercial and cloud-based offers, such as ChoiceFlow. Last, the existing infrastructure and support of the experimental lab were another point in favor of z-Tree.

### 4.1 | Treatment design

Following the same general settings as Bearden et al. (2008) and Bendoly (2011, 2013), we ask subjects to manage capacity over a booking horizon of 40 days. Participants have to handle the same RM task in 40 consecutive rounds, where each round constitutes one booking horizon. Within each booking horizon, they have to sell 5 units of capacity. Participants do not have to process every day of the booking horizon individually but only handle customer arrivals. Thus, their experience of the booking horizon jumps from, for example,  $t = 39$  to  $t = 24$  if no customers are arriving on the days in-between. Because of that, the number of decision stages within a booking horizon varies between rounds. In each booking horizon, the number of decisions that participants have to make, for example, rejecting and accepting, is limited by the capacity and the number of customer arrivals. For example, if a participant accepts the first five customers that arrive, they have no capacity left and will not need to make another decision for the remainder of the horizon. Note that all participants experience the same customer arrivals, as these are drawn randomly per round but not per participant.

The system announces the current time in the booking horizon and the left-over capacity whenever a customer arrives. Subjects have 30 s to decide; if they indicate nothing, the customer is rejected or the previously set price continues to apply. After each customer arrival, the system informs participants about the accumulated revenue and left-over capacity. Each round ends when the booking horizon ends or when there is no capacity left. An entire session takes at most 2 h.

For all participants, an accurate demand forecast indicates the mean customer arrival rate  $p_t$  and the interval of the expected willingness to pay  $\hat{w}_t$  per time slice of the booking horizon  $t \in \{T, \dots, 0\}$ . This information follows the practice

**TABLE 2** Treatment overview

Treatment	Decision task	Arrival rate	Willingness to pay
T1	acceptance	–	–
T2	acceptance	X	–
T3	acceptance	–	X
T4	acceptance	X	X
T5	price	–	–
T6	price	X	X

Note: Overview of treatments. “X” indicates that the parameter evolves over the booking horizon, whereas “–” indicates that it is stationary.

of considering expected booking curves in real-world RM systems. As no information on the presentation of information and interface design from Bearden et al. (2008) was available upon inquiry, we provide our own design. However, for transparency, we include related screen-shots in the online Appendix.

Table 2 lists all treatments; the remainder of this section describes the underlying parameters.

Treatment T1 seeks to replicate the setting of Bearden et al. (2008) by combining stationary demand and the acceptance decision task. The arrival rate is  $p_t = 0.3$  for all  $t \in \{40, \dots, 1\}$ , resulting in an average of 12 customers arriving per round. Across the booking horizon, customers’ willingness to pay is uniformly distributed in the interval  $[5;10]$  with mean  $\hat{w}_t = 7.5$ . The interval differs from Bearden et al. (2008) to make the treatment comparable to those where the willingness to pay evolves.

In treatments T2 and T4, the customer arrival rate evolves over the booking horizon. The arrival rate peaks 30 days before the end of the booking horizon. It decreases linearly to zero at  $t = 0$  and  $t = 40$ . While the number of days is arbitrary, we match the arrival rate to that of empirical airline RM data.

In treatments T3 and T4, customers’ willingness to pay increases over the booking horizon, so late arrivals are more valuable than early arrivals. Thus, at 40 days to departure,  $w_{40}$  is €2.5, increasing to  $w_1 = €12.5$ . Again, the absolute numbers are arbitrary, whereas the span of revenue differences is based on empirical airline RM data.

Treatments T5 and T6 ask participants to set prices. While T5 features stationary demand, T6 combines nonstationary arrival rates and willingness to pay. For parsimony, we do not implement a completely symmetrical treatment design but instead test the pricing decision only on the extreme cases of stationary and nonstationary demand.

## 4.2 | Experimental procedure

We recruit participants via the AixLab database, implemented by the School of Business and Economics at RWTH Aachen University Laboratory. The database lists anonymous potential participants with their program and year of studies, age,

gender, and flags indicating prior participation. We randomly invite business, engineering, or mathematics students who have studied for at least a year. We pick a unique set of participants for every treatment, such that none can participate more than once.

The sample size per treatment is 30 participants, yielding 180 participants overall. According to our survey, the age of the 180 participants ranges from 18 to 45, with the median at 23 years. In addition, 39% of subjects identify as female, while 61% identify as male. We do not survey participants’ ethnic backgrounds. To accommodate limited room sizes and reduce external effects, we divide the participants into groups of 15 and schedule sessions at different times and on different dates. We conducted all experiments within 6 weeks.

In line with the recommendations stated in Eckerdt et al. (2021), we design the experimental procedure as transparently as possible, clearly informing subjects about the research objectives and intended analysis and avoiding deception. Furthermore, as the research objective is to observe decision-making when participants are motivated to achieve good results, we rely on performance-dependent payments calibrated to the difficulty of the treatments. Specifically, after each experiment, we compute a fixed factor that, when multiplied by each participant’s performance relative to the optimum, results in an average payment of €10 per hour across all participants. These payments approximate the wage of a student assistant at the time. Thus, for an exemplary factor of 13, a participant who earned 90% of the optimal revenue would earn 11.70 Euro. As the duration of the sessions and the relative performance of participants vary, the factor is recalibrated per session.

In the introductory explanation, the researcher reads out extensive written instructions and informs the participants that their payment will depend on their revenue across all rounds. The instructions explain the distributions of arrival rate and willingness to pay graphically and in text. After reading the instructions, the researcher verbally queries the participants’ understanding. The online Appendix includes the instructions as well as screen-shots from the experimental study.

Subsequently, the participants perform an initial training round to familiarize themselves with the setting; their performance in that round does not affect their reward and does not enter the analysis. Before the evaluated part of the experiment begins, the researcher asks the participant to confirm their understanding again or to ask questions. After the experimental task, all participants fill out a survey stating demographic data and their view of the task.

We implement manipulation checks to check whether the participants have understood the task, feel comfortable with handling the interface, and do, in fact, attempt to solve the RM problem. To avoid distracting participants with traps or potentially annoying attention checks (Hauser et al., 2018), we rely on verbal checks and a behavior check. We verbally query participants’ understanding of the task

**TABLE 3** Participants' results

	Optimality gap	Reject errors	$\bar{E}$ % per accept error	$\bar{E}$ % per reject error	Static benchmark gap
T1	**2.69%	75%	1.07%	1.44%	**−9.67%
T2	**2.76%	66%	1.14%	1.60%	**−10.35%
T3	**5.69%	19%	2.05%	1.72%	0.84%
T4	**4.38%	27%	1.87%	1.97%	0.58%
T5	**5.32%	80%	2.80%	3.60%	−1.41%
T6	**3.72%	44%	2.95%	4.22%	*−1.68%

Note: \*\* significantly different from 0 with  $p < 0.01$ .

\*significantly different from 0 with  $p < 0.05$ .

before the experiment and include a related question in the survey. All participants stated that they understood the instructions.

As a behavior-based manipulation check, we compare the results from participants' decision-making to those from RANDOM and FCFS. Participants' revenue patterns strongly differ from those approaches. However, on average, participants' decision-making is more similar to OPTIMAL than to random or indiscriminate strategies. In conclusion, we assume that all participants understood the task and were reasonably motivated to perform well.

## 5 | RESULTS

This section first presents results from replicating the setting of Bearden et al. (2008). Subsequently, it describes results from the extended setting featuring nonstationary demand parameters and the pricing decision. Finally, we discuss the hypotheses stated in Section 3.5 in the light of these results.

Unless otherwise stated, we conduct two-sample nonparametric Wilcoxon tests to check the average benchmark gaps over participants and rounds for significant differences. We check the significance of BIAS indicators using a simple binomial sign test, where the probability of success under the null hypothesis is  $p = 0.5$ , given a confidence level of 95%. The resulting significance values are given in the respective tables.

The optimal revenue varies across treatments: When the arrival rate and the willingness to pay are not stationary, the value randomly deviates across rounds due to the stochastic demand. For the same demand streams (T1 vs. T5, T4 vs. T6), the optimal strategies earn less through pricing, where only the price paid turns into revenue, than through acceptance, where the complete willingness to pay turns into revenue.

### 5.1 | Replication and validation

To validate the experimental design, treatment T1 replicates the findings from Bearden et al. (2008). It yields a significant optimality gap of 2.69% (see Table 3), where Bearden et al.

(2008) observe an optimality gap of 3.23%. Furthermore, we observe a larger share of reject errors, with 75% as opposed to 54% in Bearden et al. (2008).

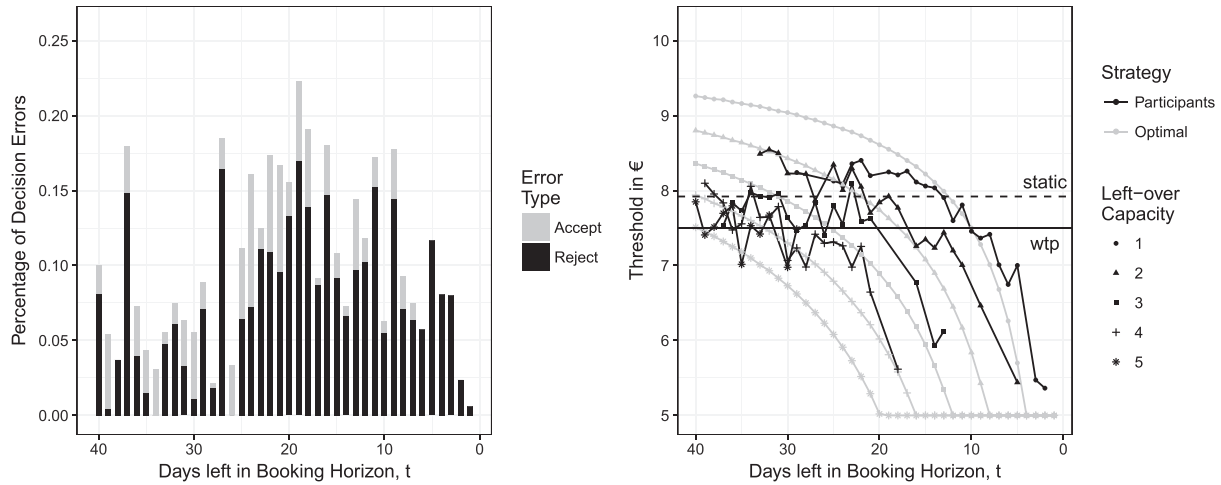
We visualize decision errors for T1 in Figure 1. The left panel shows how the percentage of decision errors varies across the booking horizon: The bars' height indicates the percentage of erroneous decisions per day in the booking horizon, whereas the bars' color indicates the share of accept versus reject errors. Thus, for example, the first bar from the left in Figure 1 shows that of all decisions that participants took on the first day, 10% deviated from the optimal strategy. Specifically, 8% were reject and 2% were accept errors. The right panel compares the participants' implicit thresholds to the optimal thresholds dependent on the left-over capacity. Figures 2–4 follow the same logic. The nature of the RM problem is such that the optimal threshold is lower when there is more capacity left or when there are fewer days left in the booking horizon. Comparing the participants' implicit thresholds to those from OPTIMAL indicates an over- or underestimation of the value of left-over capacity.

In T1, most reject errors occur during the second half of the booking horizon when optimal thresholds per left-over capacity decrease as the accumulated future customer arrival probability decreases. Participants' implicit thresholds exceed the optimal thresholds for left-over capacities of 2–5 units. For left-over capacities of 4 and 5 units, we observe a significant positive BIAS: Participants seem to have an exaggerated notion of the future value of capacity and, therefore, accept fewer bookings than optimal. However, the BIAS for 1 and 2 units of left-over capacity is not significantly negative.

Implied revenue loss from reject errors significantly exceeds that from accept errors. This finding contradicts Bearden et al. (2008), but seems intuitive: For accept errors, the implied revenue loss describes the difference between the optimal threshold and customers' willingness to pay. This value increases when the optimal threshold decreases throughout the booking horizon so that a late reject error is more costly than an early accept error. Since the implicit thresholds are not systematically lower than the optimal thresholds, the higher number of late reject errors explains the higher implied revenue loss for reject errors.

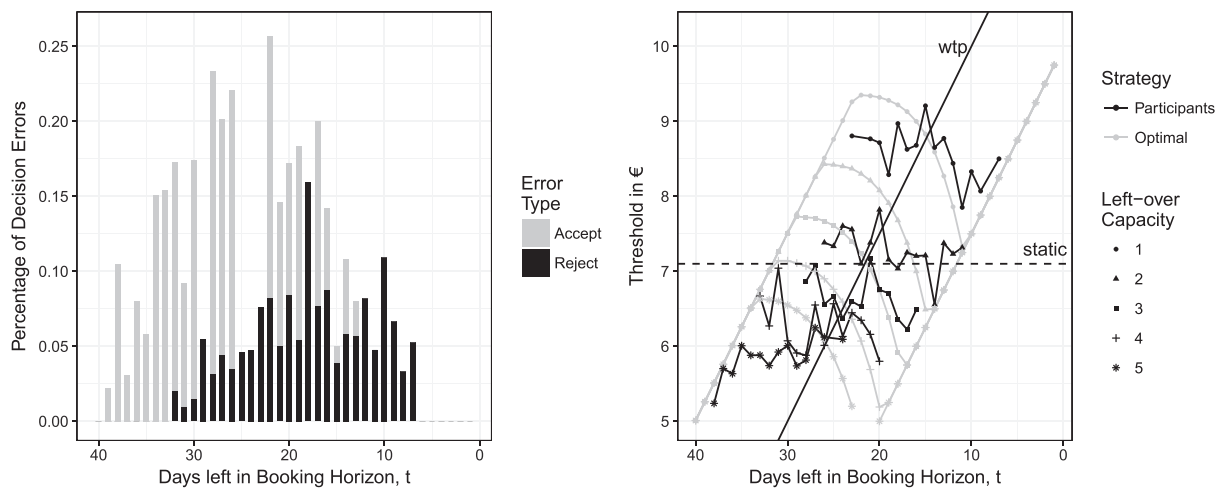
### 5.2 | Results from the extended experimental design

*Contrary to H1a, nonstationary customer arrivals do not significantly worsen participants' acceptance decisions.* There is no significant difference between the optimality gaps from T1 to T2, and subjects outperform the static benchmark in both treatments. In addition, the number of accept and reject errors, estimated thresholds, and biases do not significantly differ. However, comparing participants' revenue patterns to those from systematic approaches (Table 6) shows a slight shift in strategies: The number of participants that earn revenue patterns that most closely resemble OPTIMAL or



**FIGURE 1** T1: Decision errors and estimated thresholds

Notes: Results for acceptance decision with constant demand. The left panel shows the percentage of decision errors over the booking horizon. The right panel shows implicit thresholds in Euro per unit of left-over capacity across the booking horizon. Horizontal lines represent the average willingness to pay ( $wtp$ ) and the optimal static threshold ( $static$ ).



**FIGURE 2** T4: Decision errors and estimated thresholds

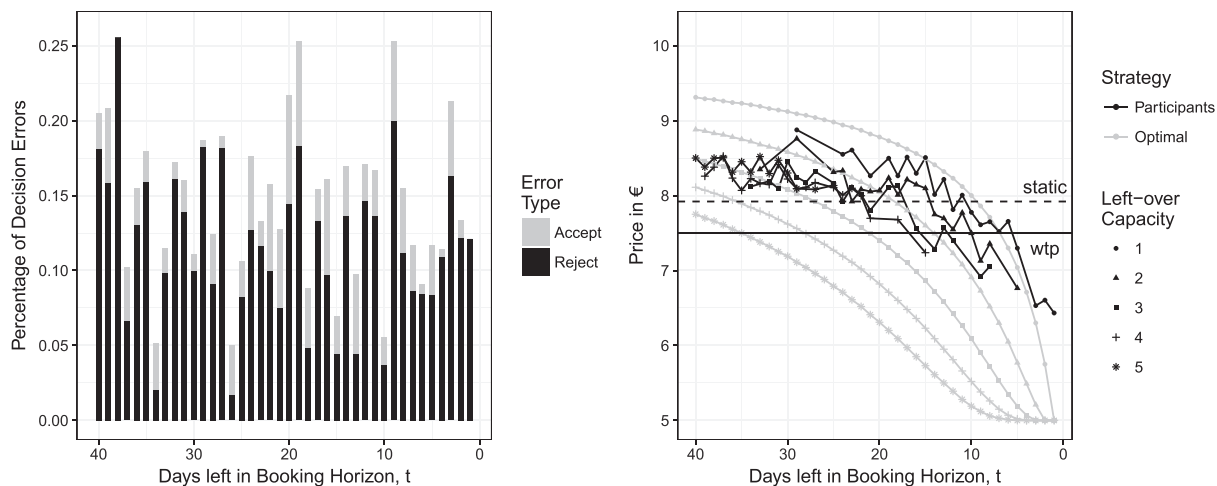
Notes: Results for acceptance with nonstationary arrival rate and willingness to pay ( $wtp$ ). The left panel shows the percentage of decision errors over the booking horizon. The right panel shows implicit thresholds per unit of left-over capacity. Additional lines represent the average willingness to pay ( $wtp$ ) and the optimal static threshold ( $static$ ).

STATIC shrinks from 23 in T1 to 14 in T2, while the number of revenue patterns that most closely resemble ANCHOR- $\hat{w}$  increases from 7 to 16.

In line with H1b, customers' nonstationary willingness to pay does impede participants' decision-making. To analyze this effect, we compare treatments T1 versus T3, T2 versus T4, and T5 versus T6. When  $p_t$  is stationary but  $\hat{w}_t$  is nonstationary, the optimality gap more than doubles, increasing from 2.69% in T1 to 5.69% in T3. This difference is significant at  $p < 0.001$ . Thus, given a stationary arrival rate, a nonstationary willingness to pay negatively affects participants' performance. Remarkably, the number of revenue

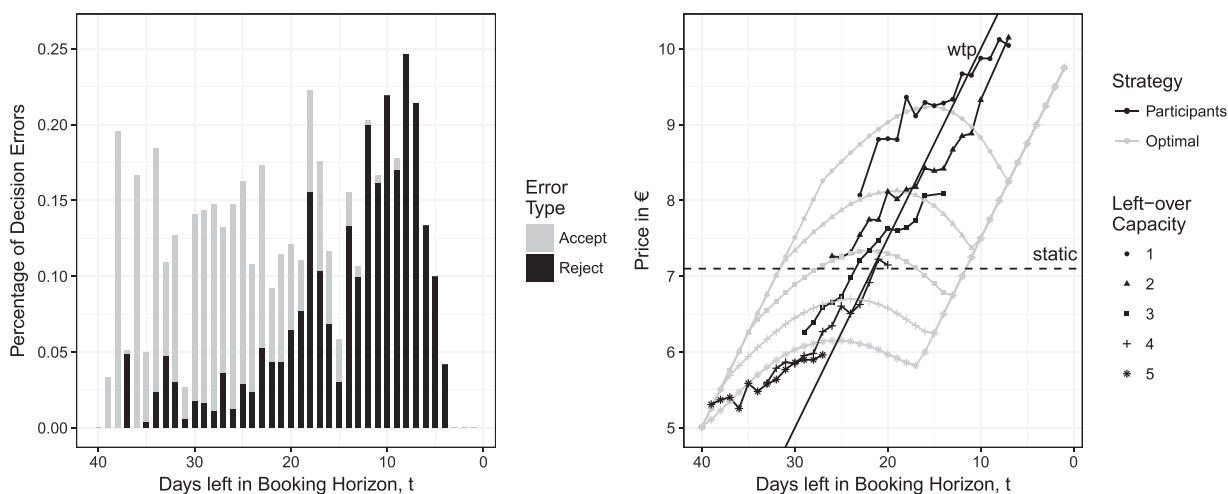
patterns that most closely resemble OPTIMAL is largest in T3 and T4. However, while this strategy is the most similar, its average distance to participants' patterns is wider in T3 and T4 than in T1.

Combining nonstationary arrival rate and willingness to pay does not complicate the problem further. Rendering both demand aspects nonstationary (T4) does not significantly increase the optimality gap compared to T3. Instead, the optimality gap seems to shrink, but applying a one-sided Wilcoxon test shows that this effect is insignificant at  $p < 0.01$ . Comparing T4 to T2, thereby adding a nonstationary willingness to pay  $\hat{w}_t$  to a nonstationary arrival rate  $p_t$ , results



**FIGURE 3** T5: Decision errors and prices

*Notes:* Results for pricing with stationary demand. The left panel shows the percentage of decision errors over the booking horizon. The right panel shows the price per unit of left-over capacity across the booking horizon. Additional lines represent the average willingness to pay and the optimal static threshold (*static*).



**FIGURE 4** T6: Decision errors and prices

*Notes:* Results for pricing with nonstationary arrival rate and willingness to pay (*wtp*). The left panel shows the percentage of decision errors over the booking horizon. The right panel shows the prices per unit of left-over capacity across the booking horizon. Additional lines represent the average willingness to pay (*wtp*) and the optimal static threshold (*static*).

in an increased optimality gap with a significance level of  $p < 0.01$ . Adding this insight to our findings on the insignificant change from T1 to T2, we conclude that a nonstationary arrival rate does not make RM harder for human decision-makers. The opposite is true for a nonstationary willingness to pay.

Contrary to H1b, pricing decisions are only less successful than acceptance decisions when demand is stationary. When participants set prices for stationary demand, the optimality gap grows from 2.69% in T1 to 5.32% in T5. Once more, this increase is significant at  $p < 0.01$ . Additionally, participants no longer outperform the static benchmark (see

Table 3). In contrast, for nonstationary demand, the optimality gaps from acceptance decisions (T4) and pricing decisions (T6) do not significantly differ ( $p > 0.2$ ). Even under difficult circumstances, participants achieve at least 82% of the optimal revenue (see Table 5). In contrast, systematic approaches earn as little as 34% (FCFS in T5, see Table 7).

Contrary to H1c, human RM decisions do benefit from repetition under nonstationary demand. To test this effect, we compare the optimality gap between the first and last 20 rounds of each treatment in Table 4. Under the stationary demand given in T1 and T5, participants' performance slightly but significantly decreases over multiple rounds.

**TABLE 4** Gaps over rounds

Treatment	First 20 rounds	Last 20 rounds	Percentage point difference
T1	**2.26%	**3.12%	*+0.86PP
T2	**3.63%	**1.85%	**−1.78PP
T3	**7.14%	**4.26%	**−2.88PP
T4	**6.77%	**1.91%	**−4.86PP
T5	**4.56%	**6.09%	+1.53PP
T6	**5.14%	**2.31%	**−2.83PP

Note: \*\* significantly different from 0 with  $p < 0.01$ .

\* significantly different from 0 with  $p < 0.05$ .

**TABLE 5** Variation in participants' individual optimality gaps

Treatment	Avg. opt. gap	Min. opt. gap	Max. opt. gap	Coefficient of variance
T1	2.26%	1.53%	4.71%	0.92%
T2	3.63%	0.37%	5.99%	1.34%
T3	7.14%	1.40%	17.41%	3.65%
T4	6.77%	−0.62%	8.50%	2.38%
T5	4.56%	−1.89%	17.20%	5.13%
T6	5.14%	−1.65%	15.23%	4.05%

**TABLE 6** Percentage revenue distance between participants and systematic approaches versus the absolute number of participants for whom the strategy is most similar

Strategy	T1	T2	T3	T4	T5	T6
OPTIMAL	19% 10 17%	9 24% 28	23% 27 20%	4 27% 20		
STATIC	19% 13 19%	5 31% 0	28% 2 18%	23 28% 8		
CHASING	28% 0 26%	0 45% 1	37% 1 25%	0 33% 1		
ANCHOR- $w_t$	21% 7 16%	16 30% 1	37% 0 21%	3 35% 0		
MIN-2-ANCH	27% 0 27%	0 39% 0	39% 0 27%	0 35% 0		
MIN-3-ANCH	32% 0 34%	0 45% 0	41% 0 32%	0 35% 0		
MIN-2-OPT	25% 0 27%	0 37% 0	36% 0 24%	0 31% 0		
MIN-3-OPT	31% 0 34%	0 43% 0	41% 0 29%	0 34% 0		
RANDOM	45% 0 43%	0 48% 0	45% 0 -	- - -		
FCFS	40% 0 45%	0 57% 0	54% 0 39%	0 42% 0		

Note: All percentages significantly differ from 0 with  $p < 0.01$ .

**TABLE 7** Optimality gaps from heuristics

Heuristic	T1	T2	T3	T4	T5	T6
STATIC	12.36%	11.92%	5.09%	3.82%	8.91%	8.13%
CHASING	6.79%	7.55%	27.43%	19.18%	10.70%	31.21%
ANCHOR- $w_t$	8.13%	3.75%	4.46%	12.86%	6.91%	26.80%
MIN-2-ANCH	7.56%	3.45%	17.74%	12.17%	29.58%	26.68%
MIN-3-ANCH	7.45%	5.83%	26.55%	18.55%	40.38%	34.07%
MIN-2-OPT	3.11%	3.19%	12.73%	12.77%	23.71%	18.64%
MIN-3-OPT	5.37%	5.60%	19.53%	19.45%	36.44%	29.46%
RANDOM	15.39%	13.13%	29.90%	24.02%	-	-
FCFS	9.95%	11.1%	45.61%	37.64%	65.14%	60.09%
Participants	2.69%	2.76%	5.69%	4.38%	7.45%	6.38%

Note: All gaps significantly different from 0 with  $p < 0.01$ .

However, the gap shrinks for decisions under nonstationary demand (T2, T3, T4, and T6), as indicated by bold values in the table. In all treatments except T4, the number of patterns that most resemble OPTIMAL steeply increases with repetition. Across the board, fewer participants earn revenue patterns that most closely resemble ANCHOR- $w_t$  and fewer participants appear to attempt a CHASING strategy.

*Human decision-making for RM is comparatively successful.* Participants' decisions generally outperform systematic heuristics. To underline this, Table 7 lists the optimality gap from all approaches and compares them to the average gap from participants' decisions. Across all treatments, only a few approaches achieve similarly small optimality gaps as the participants' decisions, and besides STATIC, only ANCHOR- $w_t$  can ever undercut it.

*In line with H2a, participants exaggerate the value of capacity when the willingness to pay is stationary.* We take a positive BIAS, where the participants' threshold exceeds the optimum, to show that they overestimate the value of capacity. As Figure 1 illustrates for T1, participants do express a significant positive BIAS when 4 or 5 units of capacity are left for sale. Also, they express a significant positive BIAS regardless of capacity when only the arrival rate varies (T2, not shown). However, this is not the case for the acceptance decision when the willingness to pay varies (T3, not shown, and T4, Figure 2): In T3 and T4, the subjects' thresholds undercut the optimum. For pricing under stationary demand (T5, Figure 3), participants' exaggerated expectations are even more apparent, as 79% of all errors are reject errors, and the average implied revenue loss per reject error exceeds that per accept error. Reject errors occur throughout the booking horizon, as opposed to T1, where they are limited to the second half. However, this is not the case when demand is nonstationary (T6, Figure 4).

*In line with H2b, participants underestimate the value of capacity when the willingness to pay is nonstationary for the acceptance decision.* When participants decide on acceptance under a nonstationary  $w_t$ , BIAS is significantly negative for left-over capacities of 2, 4, and 5 units. However, participants' revenue patterns are rarely closer to the MIN- $\phi$ -ANCH or MIN- $\phi$ -OPT approaches than to, for example, the OPTIMAL strategy. This observation holds when considering individual revenue patterns.

*Conciliating H2a and H2b, for pricing under nonstationary demand, participants overcontrol in both directions.* As the right panel of Figure 4 illustrates, participants' pricing depends on the left-over capacity and increases steeply in the latter half of the booking horizon. Participants' prices first undercut the optimal prices when 3 or more units of capacity are left. Then, they overshoot the optimal prices given less left-over capacity, displaying a significant positive BIAS.

*In line with H2c, participants anchor on the willingness to pay for pricing.* Figure 3 highlights that, during the first half of the booking horizon, participants do not differentiate between 3, 4, and 5 units of left-over capacity when pricing for stationary demand. They seem to statically wait for valuable customers to arrive, potentially based on the stationary

**TABLE 8** Anchoring effect and deviation from  $\hat{w}_t$ 

	% with anchor	% where $R_t^{part} > \hat{w}_t$	MAPD for $R_t^{part} \geq \hat{w}_t$	MAPD for $R_t^{part} < \hat{w}_t$
T1	<b>61%</b>	60%	44.10	55.89
T4	<b>88%</b>	57%	105.06	119.75
T5	37%	<b>87%</b>	64.71	<b>46.09</b>
T6	<b>82%</b>	<b>70%</b>	86.25	<b>39.58</b>

Note: Percentage of participants' decisions that indicate an anchor. Mean average percentage deviation (MAPD) calculated across the booking horizon. Bold type indicates notable differences.

expected willingness to pay. In that setting, participants' revenue patterns are similar to both those from CHASING and ANCHOR- $\hat{w}_t$ . Even though the optimality gaps from acceptance versus pricing under nonstationary demand (T4 vs. T6) do not significantly differ, the prices in T6 are even steeper than the implicit thresholds in T4. When comparing T5 and T6, differences in participants' strategies become apparent once more. In T5, participants set excessively high prices throughout the booking horizon, whereas, in T6, they only express a positive BIAS late in the booking horizon. The slightly smaller revenue distance (see Table 6) indicates that, in pricing for nonstationary demand, participants' strategies more closely resemble CHASING, relying on the previously observed  $w_t$ , than ANCHOR- $\hat{w}_t$ , relying on the expected willingness to pay. However, when considering individual participants' revenue distances to the systematic approaches, this only holds for a few participants and mainly in the first 20 rounds of the experiment.

*Participants rely most on customers' willingness to pay when setting prices and under nonstationary demand.* To further evaluate anchoring, we calculate the percentage of implicit thresholds or prices that fall between the optimal strategy and the average willingness to pay. We assume that participants rely on the expected willingness to pay if this percentage exceeds 50%. Table 8 shows that the number of related decisions is much higher for nonstationary demand in T4 and T6. The revenue distance from participants' revenue patterns to both OPTIMAL and to CHASING and ANCHOR- $\hat{w}_t$  grows as the RM problem gains complexity through nonstationary demand and pricing.

*Participants rely less on anchors for the acceptance decision.* For acceptance decisions, the deviation from the willingness to pay is always greater in the subset  $R_t^{part} < \hat{w}_t$ . In contrast, for the price decision, the deviation from the willingness to pay is always smaller in the subset  $R_t^{part} < \hat{w}_t$ . When setting prices, participants hesitate to adjust strategy to undercut  $\hat{w}_t$ , whereas they are more flexible when deciding on acceptance.

### 5.3 | Discussion

In terms of replicating and validating results from Bearden et al. (2008), T1 was largely successful. Our results confirm

that participants' strategies depend on the left-over capacity and the time left within the booking horizon. Differences in the results observed here and those reported by Bearden et al. (2008) may be due to several factors. As noted in Pagell (2021), it is almost impossible to exactly replicate behavioral laboratory experiments. Our experiments relied on a different pool of participants that also differed in cultural background, they were conducted about a decade later, and we could not perfectly replicate the instructions or screens from Bearden et al. (2008). However, we argue that these differences make it even more impressive that the qualitative direction of most findings remains stable.

In considering the hypotheses from Section 3.5 in the light of experimental results, we extend and cross-validate these findings. We summarize the approach to evaluation and our conclusions in Table 9.

**Hypothesis 1a)** If hypothesis 1a did apply without restrictions and any nonstationary demand rendered human decision-making less successful, both T2 and T3 should yield a wider optimality gap than T1. However, since there is no significant difference in optimality gaps between T1 and T2, we must dismiss this hypothesis when considering only customers' arrival rate. Nevertheless, while participants handle nonstationary arrival rates well, they struggle under a nonstationary willingness to pay.

**Hypothesis 1b)** For stationary demand, the price decision in T5 causes a wider optimality gap than the acceptance decision in T1. However, for nonstationary demand, the pricing decision does not negatively affect participants' outcomes (T4 vs. T6). Thus, we conclude that varying the decision task does not further exacerbate results when RM decision-making is already challenged by nonstationary demand, and we reject the hypothesis in its most general form. However, the pricing decision does not alleviate the adverse effects of nonstationary demand. This observation indicates a floor to human decision-making in RM, possibly due to humans' ability to adapt a strategy that returns unsatisfactory results.

**Hypothesis 1c)** Following hypothesis 1c, the participants' optimality gap should not shrink as they repeatedly engage with the same task. We can confirm this for stationary demand and acceptance decisions in T1, replicating Bearden et al. (2008). However, we find evidence for improvement in T2, T3, T4, and T6, rejecting the hypothesis for those settings: under nonstationary demand, experience improves participants' decision-making. This observation confirms findings from Kremer et al. (2011), who stress humans' ability to learn in evolving environments. Nevertheless, we note that our study's experimental setting does not focus on different types of repetition or on different approaches to providing feedback. These aspects are deserving of further research.

**Hypothesis 2a)** This hypothesis states that human RM decisions are susceptible to optimism bias. We observed confirmatory results in T1 and T2 for acceptance decisions but not under a nonstationary willingness to pay in T3 and T4. Accordingly, for the acceptance decision, we can only confirm hypothesis 2a when, at most, the customer arrival rate is nonstationary. A similar pattern emerges for pricing

TABLE 9 Hypotheses reconsidered

	Evaluation	Conclusion
H1a	Compare the optimality gap from T1 to T2 and T3.	Reject, as there is no significant difference between T1 and T2. Specify to nonstationary willingness to pay due to the significant difference in T3.
H1b	Compare the optimality gap from T1 to T5 and the optimality gap from T4 to T6 to evaluate acceptance versus pricing tasks.	Confirm for stationary demand, given a significant difference between T1 and T5, but reject for nonstationary demand, given insignificant differences between T4 and T6.
H1c	Compare the optimality gap in the first 20 rounds to that in the last 20 rounds for all treatments.	Confirm for stationary demand, lacking significant improvements in T1 and T5, but reject for nonstationary demand, given significant improvements in T2, T3, T4, and T6.
H2a	Analyze the share of reject errors, excessive implicit thresholds, and positive BIAS for all treatments.	Confirm optimism bias in T1, T2, and T5, where the willingness to pay is stationary, but reject in T3, T4, and T6, where it is nonstationary.
H2b	Analyze the share of accept errors, low implicit thresholds, and negative BIAS for all treatments. Compare participants' revenue patterns to those from MIN- $\phi$ -ANCH and MIN- $\phi$ -OPT.	Reject for T1 and T2, when participants accept requests under stationary willingness to pay, but confirm in T3 and T4, under nonstationary willingness to pay. Under the pricing task in T5 and T6, confirm for early parts of the booking horizon.
H2c	Analyze distance between implicit thresholds and potential anchors across treatments. Compare participants' revenue patterns to those from CHASING and ANCHOR approaches.	Reject for acceptance decisions (T1-T4), but confirm for pricing (T5 and T6).

ing decisions: participants are more optimistic when setting prices under stationary demand (T5) than when demand is nonstationary (T6).

**Hypothesis 2b)** The loss aversion predicted by hypothesis 2b would cause participants to accept many bookings early in the booking horizon, exhibiting a negative BIAS, to avoid losing revenue from reject errors. We observe a corresponding BIAS and accept errors in T3 and T4, where the willingness to pay is nonstationary, but not in T1 and T2, where the willingness to pay is stationary. This observation also reconciles the apparent contradictory hypotheses 2a and 2b: Both biases apply to the acceptance decision but in different demand settings. For the price decision, participants' bias depends on the time in the booking horizon: They first underestimate the value of capacity, possibly due to loss aversion, and later attempt to catch up by setting excessively high prices.

Accept and reject errors are not independent. By affecting the remaining capacity and thereby the RM decision space, accept errors early in a booking horizon can induce reject errors later: When selling all capacity early, decision-makers have to reject bookings that an optimal strategy would accept later in the booking horizon. Accordingly, when decision-makers express a loss aversion bias early in the booking horizon, their later decisions are likely to appear overly optimistic. Even reverting to an optimal strategy for the reduced left-over capacity would mean rejecting more customers than the initial optimal strategy. In contrast, early reject errors do not affect the capacity and, therefore, enable, but do not enforce, later accept errors. Here, we compare the "currently optimal" dynamic strategy when measuring errors so that

unavoidable deviations from the initial optimum should not obfuscate actual biases from overcorrecting.

**Hypothesis 2c)** Given the results from the extended experimental setting, we can broadly confirm that human decision-making for RM relies on anchors. For pricing under stationary demand, participants do not differentiate their strategy based on the left-over capacity, which we attribute to their reliance on a *stationary anchor*. In addition, differences between the prices observed in T5 and implicit thresholds estimated for T1 indicate that the pricing anchor is the expected willingness to pay. However, this finding can be partially attributed to the optimistic approach to estimating implicit thresholds. Nevertheless, the similarity between participants' revenue patterns and those from ANCHOR- $\hat{w}_t$  supports the idea that participants anchor on the expected willingness to pay. When participants set prices under nonstationary demand, they apparently rely on the most recently observed willingness to pay as a *nonstationary anchor*. The slightly smaller revenue distance to CHASING supports this idea. When we consider the share of implicit thresholds or prices that fall between the optimal strategy and the average willingness to pay as an indicator, anchoring increases when demand varies, as in T4 (acceptance) and T6 (pricing). However, participants appear more open to adjusting their anchor when tasked with acceptance decisions, as the variance observed in acceptance decisions is higher than that observed in pricing decisions.

**Excursus: Individual differences.** While the main focus of the result analysis is on aggregate measurements, we observe considerable variation across subjects. This variation increases under nonstationary demand and for pricing

decisions. Notably, participants can outperform the optimal solution when pricing, as the optimal solution is based on expected demand, and actual demand may deviate from the forecast. This deviation explains the negative optimality gaps in Table 5. However, the average optimality gaps and particularly the gaps induced by the worst performers increase for pricing. While surveying participants' demography and task experience in our study, we could not link either factor to performance. The same applies to the strategy choices apparent from revenue distances (Table 6). To relate participants' strategies to characteristics such as their motivation, as proposed in Eroglu and Knemeyer (2010), we recommend a separate study with little variation in task or demand and a larger pool of participants per treatment.

## 6 | CONCLUSION AND MANAGERIAL IMPLICATIONS

As its first theoretical contribution, this paper extended the experimental setting from Bearden et al. (2008) by adding variables describing nonstationary demand over the booking horizon. In a related empirical contribution, we replicated results from Bearden et al. (2008) under stationary demand. Furthermore, we cross-validated the results in the sense of Schoenherr and Swink (2012) by setting the same task under different degrees of nonstationary demand. A nonstationary willingness to pay significantly increased the subjects' gap to the dynamic optimum and canceled out their advantage over the static optimum. Additionally, we cross-validated results from Bearden et al. (2008) regarding benefits from experience: Participants did not benefit from gaining experience under stationary demand, but their performance improved under nonstationary demand.

As a second theoretical contribution, we extended research on RM decision-making by asking participants to set prices dynamically. In a related empirical contribution, we cross-validated results from Bearden et al. (2008) for the pricing task under stationary and nonstationary demand. We found that the pricing decision complicates human decision-making more than dynamic acceptance.

As a third theoretical contribution, we formalized several systematic RM solution approaches that extend beyond the dynamic and static optima, and that implement decision biases, such as demand chasing (Lau & Bearden, 2013), anchoring (Kocabiyikoglu et al., 2018), and loss aversion (Schweitzer & Cachon, 2000). In a related empirical contribution, we compared the participants' revenue patterns to those that resulted from systematic approaches given the same demand streams. In the experiments, subjects outperformed the systematic approaches. Furthermore, we found evidence for both optimism bias and loss aversion in differentiated demand settings. Last, we concluded that the pricing decision caused participants to anchor on the expected willingness to pay.

## 6.1 | Implications and research opportunities

We conclude this paper by considering implications for the behaviorally informed design of RM systems and pointing out related research opportunities.

**Individual perspective: the analyst.** Insights from the pricing task are relevant for the work of individual analysts, as many RM systems let analysts adjust the opportunity cost of capacity via a bid price. We found that decision-makers tend to anchor on the given information, such as the expected willingness to pay. Analysts may neglect essential aspects, such as the demand variance, left-over time, or remaining capacity. Therefore, RM systems need to emphasize the role of alternative key indicators. For example, contrasting the values of different capacity levels may help to reduce bias. To test this, an experimental study that provides participants with a demand forecast and indicates the resulting opportunity cost of capacity could offer further insights.

Other related approaches are conceivable when examining systematic solution approaches, for example, systematically increasing the implicit threshold or the price by a set increment per time step. Additionally, to investigate implications from early decisions for later parts of the booking horizon and dependencies in error types, future research could implement approaches that revert to systematic rules after an initial period of human decision-making. For example, a "revert to dynamic optimum" strategy could apply controls set by a human decision-maker for the first part of the booking horizon and later dynamically optimize revenue for the remaining capacity. Such a strategy would differ per decision-maker, remaining capacity, and the timing of the revert. To further examine individual differences, confronting more participants with the same RM task and the same demand setting and clustering the resulting revenue patterns could also provide valuable insights.

**Managerial perspective: the team.** Managers may counterbalance anchoring behavior by adapting their performance evaluation. However, in this assessment, they need to be aware of the dangers of communicating simplified strategies, such as achieving a minimum load, as this might worsen loss-averse decision biases. A multiobjective problem emerges when organizational loss aversion calls for robust RM beyond revenue maximization. In this regard, research on human decision-making for multiobjective problems would be promising. Further research assigning merit for good RM decisions could draw on existing work regarding revenue sharing and contract design.

**Design perspective: the system.** In contrast to the classical definition of decision support systems (Kasper, 1996), RM systems do not support infrequent, managerial decisions but instead guide dynamic operational planning. Therefore, we view such systems as symbiotic analytics systems (Schütze et al., 2020). System designers need to consider the implications of biases for the human–algorithm interface of symbiotic analytics systems. For example, existing RM

systems communicate demand forecasts in terms of mean and variance. However, when the system visualizes expected booking curves as deterministic bar charts or does not exemplify the implications of a high variance in bookings, analysts may be tempted to undervalue the variance of demand. Therefore, we argue that system design needs to include careful framing for such information.

As emphasized, for example, in Bendoly (2016), data visualization also plays a significant role in systems design. To test different approaches to visualization, we recommend implementing such experimental interfaces as an interactive app, for example, based on the Shiny package and R or the Dash package and Python.

In the experimental setting described here and in large parts of RM practice, decision-makers struggled to complement demand forecasts with dynamic booking information. Well-designed systems should provide dynamic guidance, alerts, and the opportunity to pursue what-if analyses (Parikh et al., 2001). We recommend following advice found in Kasper (1996) and Fildes et al. (2006) to communicate the relevant information without inspiring overconfidence.

As opposed to the minimalist setting given here, a maximalist experimental setting could evaluate the interaction of experts within the business context and when facing a realistic user interface. Such a setting would let researchers assess many differences in both process and system design. Furthermore, researchers could observe how participants handle interlinking RM subproblems, for example, adjusting the demand forecast versus overriding optimization parameters. However, to yield meaningful results, we expect a maximalist study to require participants with in-depth expertise. This requirement severely limits the pool of potential participants. In addition, the broad diversity of possible RM system and process designs and the complexity of the related models create a wide range of possible system variants. In the light of this challenge, empirical studies, for example, as described in Zeni (2003) or Schütze et al. (2020), might offer a better way of considering RM in its full complexity.

Future research on symbiotic analytics systems is needed to consider the interaction between automated systems and analysts in dynamic, operational decision scenarios. As also emphasized in O'Keefe (2016), such future research must draw on experimental behavioral research from the interface of information systems and operational management.

## ORCID

Catherine Cleophas  <https://orcid.org/0000-0002-3676-2782>

## REFERENCES

- Bearden, J.N., Murphy, R.O. & Rapoport, A. (2008) Decision biases in revenue management: some behavioral evidence. *Manufacturing & Service Operations Management*, 10(4), 625–636.
- Bendoly, E. (2011) Linking task conditions to physiology and judgment errors in RM systems. *Production and Operations Management*, 20(6), 860–876.
- Bendoly, E. (2013) Real-time feedback and booking behavior in the hospitality industry: moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management*, 31(1–2), 62–71.
- Bendoly, E. (2016) Fit, bias, and enacted sensemaking in data visualization: frameworks for continuous development in operations and supply chain management analytics. *Journal of Business Logistics*, 37(1), 6–17.
- Benzion, U., Cohen, Y., Peled, R. & Shavit, T. (2008) Decision-making and the newsvendor problem: an experimental study. *Journal of the Operational Research Society*, 59(9), 1281–1287.
- Bolton, G.E. & Katok, E. (2008) Learning by doing in the newsvendor problem: a laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management*, 10(3), 519–538.
- Bolton, G.E., Ockenfels, A. & Thonemann, U.W. (2012) Managers and students as newsvendors. *Management Science*, 58(12), 2225–2233.
- De Berker, A.O., Rutledge, R.B., Mathys, C., Marshall, L., Cross, G.F., Dolan, R.J. & Bestmann, S. (2016) Computations of uncertainty mediate acute stress responses in humans. *Nature Communications*, 7(1), 1–11.
- Eckerd, S., DuHadway, S., Bendoly, E., Carter, C.R. & Kaufmann, L. (2021) On making experimental design choices: discussions on the use and challenges of demand effects, incentives, deception, samples, and vignettes. *Journal of Operations Management*, 67(2), 261–275.
- Epley, N. & Gilovich, T. (2006) The anchoring-and-adjustment heuristic: why the adjustments are insufficient. *Psychological Science*, 17(4), 311–318.
- Eroglu, C. & Croxton, K.L. (2010) Biases in judgmental adjustments of statistical forecasts: the role of individual differences. *International Journal of Forecasting*, 26(1), 116–133.
- Eroglu, C. & Knemeyer, A.M. (2010) Exploring the potential effects of forecaster motivational orientation and gender on judgmental adjustments of statistical forecasts. *Journal of Business Logistics*, 31(1), 179–195.
- Fildes, R., Goodwin, P. & Lawrence, M. (2006) The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, 42(1), 351–361.
- Fischbacher, U. (2007) Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178.
- Flyvbjerg, B. (2008) Curbing optimism bias and strategic misrepresentation in planning: reference class forecasting in practice. *European Planning Studies*, 16(1), 3–21.
- Furnham, A. & Boo, H.C. (2011) A literature review of the anchoring effect. *Journal of Socio-economics*, 40(1), 35–42.
- Fürstenau, D., Cleophas, C. & Kliever, N. (2020) How do market standards inhibit the enactment of digital capabilities? *Business & Information Systems Engineering*, 62(4), 279–287.
- Glock, C.H., Grosse, E.H., Jaber, M.Y. & Smunt, T.L. (2019) Applications of learning curves in production and operations management: a systematic literature review. *Computers & Industrial Engineering*, 131, 422–441.
- Gönsch, J. (2017) A survey on risk-averse and robust revenue management. *European Journal of Operational Research*, 263(2), 337–348.
- Hardisty, D.J., Appelt, K.C. & Weber, E.U. (2013) Good or bad, we want it now: fixed-cost present bias for gains and losses explains magnitude asymmetries in intertemporal choice. *Journal of Behavioral Decision Making*, 26(4), 348–361.
- Hauser, D.J., Ellsworth, P.C. & Gonzalez, R. (2018) Are manipulation checks necessary? *Frontiers in Psychology*, 9, 998.
- Hertwig, R., Hogarth, R.M. & Lejarraga, T. (2018) Experience and description: exploring two paths to knowledge. *Current Directions in Psychological Science*, 27(2), 123–128.
- Hertwig, R., Wulff, D.U. & Mata, R. (2019) Three gaps and what they may mean for risk preference. *Philosophical Transactions of the Royal Society B*, 374(1766), 20180140.
- Hilbert, M. (2012) Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychological Bulletin*, 138(2), 211.
- Hinterhuber, A. (2015) Violations of rational choice principles in pricing decisions. *Industrial Marketing Management*, 47, 65–74.

- Ho, T. & Zhang, J. (2008) Designing pricing contracts for boundedly rational customers: does the framing of the fixed fee matter? *Management Science*, 54(4), 686–700.
- Hogarth, R.M., Lejarraga, T. & Soyer, E. (2015) The two settings of kind and wicked learning environments. *Current Directions in Psychological Science*, 24(5), 379–385.
- Kahneman, D. & Frederick, S. (2002) Representativeness revisited: attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. S. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81).
- Kasper, G.M. (1996) A theory of decision support system design for user calibration. *Information Systems Research*, 7(2), 215–232.
- Katok, E. & Wu, D.Y. (2009) Contracting in supply chains: a laboratory investigation. *Management Science*, 55(12), 1953–1968.
- Katsikopoulos, K.V. & Gigerenzer, G. (2013) Behavioral operations management: a blind spot and a research program. *Journal of Supply Chain Management*, 49(1), 3–7.
- Kimes, S. (1989) Yield management: a tool for capacity-considered service firms. *Journal of Operations Management*, 8(4), 348–363.
- Kirshner, S.N. & Moritz, B.B. (2021) Measuring demand chasing behavior. *Decision Sciences*.
- Klein, R., Koch, S., Steinhardt, C. & Strauss, A.K. (2020) A review of revenue management: recent generalizations and advances in industry applications. *European Journal of Operational Research*, 284(2), 397–412.
- Kocabiyikoglu, A., Gogues, C.I. & Hekimoglu, M.H. (2018) The impact of decision types on revenue management decisions: an experimental study. *Decision Sciences*, 49(2), 225–249.
- Kocabiyikoglu, A., Gogus, C.I. & Gonul, M.S. (2015) Revenue management vs. newsvendor decisions: does behavioral response mirror normative equivalence? *Production and Operations Management*, 24(5), 750–761.
- Kollwitz, H. & Papathanassis, A. (2011) Evaluating cruise demand forecasting practices: a Delphi approach. In G. Philip, P. Alexis & M. Petra (Eds.), *Cruise sector challenges* (pp. 39–55). Springer.
- Kremer, M., Minner, S. & Van Wassenhove, L.N. (2010) Do random errors explain newsvendor behavior? *Manufacturing & Service Operations Management*, 12(4), 673–681.
- Kremer, M., Moritz, B. & Siemsen, E. (2011) Demand forecasting behavior: system neglect and change detection. *Management Science*, 57(10), 1827–1843.
- Lau, N. & Bearden, J.N. (2013) Newsvendor demand chasing revisited. *Management Science*, 59(5), 1245–1249.
- Mukhopadhyay, S., Samaddar, S. & Colville, G. (2007) Improving revenue management decision making for airlines by evaluating analyst-adjusted passenger demand forecasts. *Decision Sciences*, 38(2), 309–327.
- Niederhoff, J.A. & Kouvelis, P. (2019) Effective and necessary: individual supplier behavior in revenue sharing and wholesale contracts. *European Journal of Operational Research*, 277(3), 1060–1071.
- O’Keefe, R.M. (2016) Experimental behavioural research in operational research: what we know and what we might come to know. *European Journal of Operational Research*, 249(3), 899–907.
- Pagell, M. (2021) Replication without repeating ourselves: addressing the replication crisis in operations and supply chain management research. *Journal of Operations Management*, 67(1), 105–115.
- Parikh, M., Fazlollahi, B. & Verma, S. (2001) The effectiveness of decisional guidance: an empirical evaluation. *Decision Sciences*, 32(2), 303–332.
- Petropoulos, F., Fildes, R. & Goodwin, P. (2016) Do ‘big losses’ in judgmental adjustments to statistical forecasts affect experts’ behaviour? *European Journal of Operational Research*, 249(3), 842–852.
- Schoenherr, T. & Swink, M. (2012) Revisiting the arcs of integration: cross-validations and extensions. *Journal of Operations Management*, 30(1–2), 99–115.
- Schram, A. & Ule, A. (2019) *Handbook of research methods and applications in experimental economics*. Edward Elgar Publishing.
- Schütze, C., Cleophas, C. & Tarafdar, M. (2020) Revenue management systems as symbiotic analytics systems: insights from a field study. *Business Research*, 13(3), 1007–1031.
- Schweitzer, M.E. & Cachon, G.P. (2000) Decision bias in the newsvendor problem with a known demand distribution: experimental evidence. *Management Science*, 46(3), 404–420.
- Sun, J., Zhang, D.J., Hu, H. & Van Mieghem, J.A. (2021) Predicting human discretion to adjust algorithmic prescription: a large-scale field experiment in warehouse operations. *Management Science*.
- Talluri, K.T. & van Ryzin, G.J. (2004) *The theory and practice of revenue management*. Springer.
- Thomas, R.W. (2011) When student samples make sense in logistics research. *Journal of Business Logistics*, 32(3), 287–290.
- Tokar, T., Aloysius, J., Waller, M. & Hawkins, D.L. (2016) Exploring framing effects in inventory control decisions: violations of procedure invariance. *Production and Operations Management*, 25(2), 306–329.
- Tversky, A. & Kahneman, D. (1992) Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 323, 297–323.
- Tyebjee, T.T. (1987) Behavioral biases in new product forecasting. *International Journal of Forecasting*, 3(3–4), 393–404.
- Vock, S., Garrow, L.A. & Cleophas, C. (2022) Clustering as an approach for creating data-driven perspectives on air travel itineraries. *Journal of Revenue and Pricing Management*, 21(2), 212–227.
- Zeni, R. (2003) The value of analyst interaction with RM systems. *Journal of Revenue and Pricing Management*, 2, 37–46.

## AUTHOR BIOGRAPHIES

**Catherine Cleophas** holds the professorship Service Analytics at CAU Kiel University. Her research interests are matching supply and demand through anticipative optimization, evaluating the success of planning approaches through simulation modeling, and developing decision support systems that combine automated analytics and analyst expertise.

**Claudia Schütze** conducted research as a doctoral student at RWTH Aachen University. Her research interests include revenue management and behavioral operations research.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Cleophas C., & Schütze C. (2024) Decision biases in revenue management revisited: Dynamic decision-making under stationary and nonstationary demand. *Decision Sciences*. 55, 159–175. <https://doi.org/10.1111/deci.12573>