Forthcoming in *Manufacturing & Service Operations Management*
manuscript MSOM-18-505

Believing in Analytics: Managers' Adherence to Price Recommendations from a DSS

Felipe Caro
University of California, Los Angeles, Anderson School of Management, fcaro@anderson.ucla.edu

Anna Sáez de Tejada Cuenca
IESE Business School, ASaezdetejada@iese.edu

Problem definition: We study the adherence to the recommendations of a decision support system (DSS) for clearance markdowns at Zara, the Spanish fast fashion retailer. Our focus is on behavioral drivers of the decision to deviate from the recommendation, and the magnitude of the deviation when it occurs. *Academic/practical relevance:* A major obstacle in the implementation of prescriptive analytics is users' lack of trust in the tool, which leads to status quo bias. Understanding the behavioral aspects of managers' usage of these tools, as well as the specific biases that affect managers in revenue management contexts, is paramount for a successful rollout. *Methodology:* We use data collected by Zara during seven clearance sales campaigns to analyze the drivers of managers' adherence to the DSS. *Results:* Adherence to the DSS's recommendations was higher, and deviations were smaller, when the products were predicted to run out before the end of the campaign, consistent with the fact that inventory and sales were more salient to managers than revenue. When there was a higher number of prices to set, managers of Zara's own stores were more likely to deviate from the DSS's recommendations, whereas franchise managers did the opposite and showed a weak tendency to adhere more often instead. Two interventions aimed at shifting salience from inventory and sales to revenue helped increase adherence and overall revenue. *Managerial implications:* Our findings provide insights on how to increase voluntary adherence that can be used in any context in which a company wants an analytical tool to be adopted organically by its users. We also shed light on two common biases that can affect managers in a revenue management context, namely salience of inventory and sales, and cognitive workload.

Key words: behavioral operations, empirical operations, revenue management, pricing, retailing

---

1. Introduction

A large portion of the Operations Management literature provides prescriptive solutions to practitioners' complex problems. When these solutions get successfully implemented they lead to higher revenue, lower costs, increased efficiency, or other improvements. Sometimes firms implement these solutions in the form of automated algorithms. However, leaving algorithms to make decisions unsupervised can be risky. For instance, algorithms that set prices automatically by tracking competitors have led to notorious failures, such as a $23 million textbook (Eisen 2011), or thousands of products priced at £0.01, a mistake that brought losses of up to £100,000 to many small vendors (Neate 2014). Other times, operational solutions are implemented in the form of a decision support system (DSS), which are tools that make algorithm-based recommendations to human decision makers, who can choose to override them. The usage of DSSs can greatly improve decision making. For instance, the Franz Edelman Award presented by INFORMS has shown how algorithms implemented in a DSS have created enormous value for organizations.

When managers make operational decisions assisted by a DSS, their choices impact the performance of the firm. If the DSS's algorithm is recommending optimal decisions but managers deviate from them, the potential improvement that the DSS was supposed to cause may never be realized. Therefore, it is important to study how humans interact with these tools and what drives their adherence to its recommendations. Understanding the behavioral aspects of managers' usage of DSSs can provide insights on how to design their interfaces to ensure a successful implementation. Many academics and practitioners have called for more research in this area. For instance, Bendoly et al. (2006) write: "When it comes to implementation, the success of operations management tools [...] relies heavily on our understanding of human behavior". This is also mentioned in Fisher and Raman (2010), who cite the example of a consumer electronics company that implemented an inventory DSS for company experts, but after a year they were overriding its recommendations more than 80% of the time. Fisher and Raman conclude: "This retailer, like many others, focused its effort on the technical aspects of improving forecasting and inventory planning while under-investing in the more anecdotal evidence of the importance of understanding managers' behavior when implementing DSSs in retail revenue management".

In the context of pricing and revenue management, an important research goal is to develop implementable algorithms. According to Talluri and van Ryzin (2005), what defines modern revenue management is making demand decisions with science- and technology-based systems which are overseen by human analysts. The final step, supervision and potential intervention by humans, is still vital. For instance, analysts can tune the DSS's parameters to align its objectives with the company's strategic goals. Or, if there are demand shocks given by unexpected events, they can adjust forecasts manually. However, such analysts can also have difficulties adjusting to the new system because it does not rely on their own intuition and experience.

The lack of adjustment to a new system — an instance of status quo bias — can lead managers to deviate from optimal pricing decisions with the associated losses or unrealized profits. In addition to resistance to change, other biases can interfere with the process of setting optimal prices. On the one hand, inventory and speed of sales may be more salient to managers than revenue, their actual goal. On the other hand, when there are many prices to set — an increasingly common situation as retailers' assortment size keeps growing — the larger number of decisions and information to examine may lead managers to suboptimal choices due to the cognitive workload involved.

In this paper, we study how managers interact with analytics-based tools to make markdown decisions during clearance sales. Our overarching question is the following: *Could status quo bias be hindering the implementation of a pricing DSS?* Status quo bias in our setting manifests itself in managers' attachment to a legacy heuristic, which motivates two sub-questions: *Do managers overestimate the importance of inventory and speed of sales when making (revenue-maximizing) pricing decisions? When there are many prices to set, are managers more or less likely to follow the DSS's recommendations?* Our last question deals with remedies: *Can providing interpretable feedback of a revenue metric in the DSS's interface mitigate the effect of inventory salience bias?*

To answer these questions, we study the aftermath of the implementation of a particular DSS for clearance sales revenue optimization at Zara. A pilot test of a prototype in 2008 showed that the DSS increased revenue by almost 6%; see Caro and Gallien (2012). However, when the DSS was rolled out in 2010, managers ignored the DSS's recommendations too often.

Before the implementation of the DSS, pricing managers at Zara followed a heuristic that was prone to behavioral biases and focused on the inventory run-out time (salience bias) and on consolidating price categories to reduce the number of pricing decisions (cognitive workload). Using a Heckman regression, we find that, even after the revenue-maximizing DSS was implemented, the heuristic pricing behavior persisted, which in turn is evidence of status quo bias. In particular, an incremental week to deplete the inventory was associated with a 2.3% increase in the probability of deviating from the DSS's recommendations, and a 10.8% increase in the magnitude of deviations for countries in which Zara owned the stores. Similarly, an extra pricing decision in countries where Zara owned the stores was associated with a 6.8% increase in the probability of deviating, and with a 22.6% increase in the magnitude of deviations. These results are statistically significant and continue to hold when all countries are considered (own stores together with franchises). An analysis of only franchises showed support for salience bias of inventory, but more pricing decisions was (weakly) associated with less deviations, i.e., franchise managers delegated the larger workload on the algorithmic tool.

In the three years that followed the initial implementation, Zara performed two sequential changes (interventions) to the DSS's interface to address inventory salience. The first intervention displayed the revenue metric that the DSS was maximizing to give managers feedback about their pricing decisions. The second intervention consisted in showing a reference point for the revenue metric to make it more interpretable. Following a difference-in-differences analysis, we find that the first intervention had no significant effect on managers' adherence, but combined with the second one, franchises' adherence increased by 9.5 percentage points from a 46.4% baseline.

Our results illustrate how some human biases can be an obstacle in the usage of pricing and revenue management DSSs, and provide recommendations on how to better design algorithmic tools so their implementation is successful. Arguably, the biases that we document may be pervasive beyond Zara, in many other firms that count on DSSs to solve complex revenue management problems. In such contexts, where the optimal decisions often run contrary to human intuition, analytics-based tools can help achieve the best outcomes, but only if managers believe in them.

We contribute to the academic literature in several ways. First, we contribute to the literature on behavioral issues in pricing and revenue management by showing that managers may not entirely understand the revenue-inventory tradeoff and may be making suboptimal decisions when inventory is too salient. Second, we contribute to the study of human-algorithm interactions and deviations from algorithmic recommendations by providing empirical evidence, using field data, of resistance to change from old heuristics to analytics-based decision-making (consistent with status quo bias).

Some of the data presented in this paper has been disguised to protect its confidentiality, and we emphasize that the views presented here do not necessarily represent those of the Inditex Group.

2. Literature Review

Our work is relevant to three streams of literature: behavioral issues in pricing and revenue management, human-algorithm interaction, and adherence to fixed policies and algorithmic recommendations. We review them separately below. For general reviews of behavioral OM, see Bendoly et al. (2006), Donohue et al. (2019) and Fahimnia et al. (2019).

One distinguishing feature of our paper is that we study managers' decision-making using real data collected by a firm. The need to understand managers' behavior using field data, to confirm laboratory results, is emphasized in Donohue and Schultz (2018), Goldfarb et al. (2012), and Gino and Pisano (2008). We focus on pricing tasks, which are short-term decisions; we complement the studies on the consequences of managerial judgment on firms' longer-term strategy (Bertrand and Schoar 2003, Gallino et al. 2019, Hardcopf et al. 2017).

2.1. Behavioral Issues in Pricing and Revenue Management

A large fraction of the revenue management literature consists of papers proposing advanced optimization systems, some of which get implemented in real-life companies. Most often, human managers are in charge of supervising and deciding the final prices to be set, to be able to make corrections or account for unexpected events that affect demand or new company policies. Fisher and Raman (2010) and Talluri and van Ryzin (2005) both mention the importance of human analysts in revenue management and retail analytics systems, as well as the real possibility that their beliefs and behavioral biases will undermine the implementation of those systems. One of the biases that pose a barrier to the development of retail analytics tools is resistance to change (Fisher and Raman 2010), also called inertia (Rooderkerk et al. 2021). Some revenue management papers describe the aftermath of implementations, and how they are not always smooth due to managers' behavior (Achabal et al. 2000), but the evidence so far is mostly anecdotal. We help fill the gap with empirical results from a revenue management implementation (DSS) at Zara.

The behavioral aspects of dynamic pricing have been studied in depth when managers are assumed to be setting optimal prices to which consumers react exhibiting different behavioral traits (Aflaki et al. 2019, Aviv and Pazgal 2008, van den Boer 2015, Kazerouni and van Roy 2017, Kremer et al. 2017, Li and Jain 2015, Liu et al. 2014, Mak et al. 2014, Nasiry and Popescu 2011, Osadchiy and Bendoly 2015, Özer and Zheng 2012, 2015, Qiu and Whinston 2017, Zhang et al. 2019). Only a few works study why price decision-makers deviate from seemingly optimal prices, with reasons such as loss of future consumers (Anderson and Simester 2010), firm's myopia (Che et al. 2007), lack of professionalization of the price-setter (Li et al. 2019), and brand perception (DellaVigna and Gentzkow 2019). Our focus are managers' operational drivers in a dynamic setting, as opposed to the more strategic variables that these papers study. More particularly, we study two biases — salience of inventory and speed of sales metrics, and rational inattention — that could potentially affect decision makers in any modern pricing and revenue management setting.

Managers' behavioral biases in revenue management have been studied mostly in theoretical or laboratory papers; see Özer and Zheng (2012) for a review. Some of the topics these experimental papers study are adjustment to inventory levels (Bearden et al. 2008); framing effects (Kocabıyıkoğlu et al. 2018); managers' physiological conditions (Bendoly 2011, 2013); and how pricing interacts with other considerations, such as inventory (Kocabıyıkoğlu et al. 2015), production quantity (Ramachandran et al. 2018), or buyer/seller role (Mak et al. 2018).

Two empirical papers closely related to ours are Elmaghraby et al. (2012) and Elmaghraby et al. (2015), which study how B2B salespeople set prices for groceries assisted by a DSS. Their outcome of interest are the price changes, driven by cost changes and moderated by the DSS's recommendations, while ours is adherence to the DSS's recommendations, and we focus on managers' cognitive biases as its drivers. In addition, we study the effect of two interventions on adherence.

2.2. Human-Algorithm Interaction

Another related stream of literature is the study of interactions between humans and algorithms. Collaboration between humans and algorithmic tools can outperform humans or algorithms alone (Luong et al. 2020a,b). When humans have private information, their insight combined with the algorithm's can outperform the algorithm alone. Examples of such an approach are Ball and Ghysels (2017), Blattberg and Hoch (1990), Flicker (2018), Hoch and Schkade (1996), Ibrahim and Kim (2019), Ibrahim et al. (2021), Oh and Oliva (2020), and van der Staak et al. (2020). Arvan et al. (2019) reviews the literature on human-algorithm forecast integration, noting that all the work on adherence is theoretical or in the laboratory.

Humans exhibit algorithm aversion, which can be mitigated by allowing them to make small changes to the algorithm's output (Dietvorst et al. 2016), and increases in more uncertain environments (Dietvorst and Bharti 2019) or in subjective tasks (Castelo et al. 2019). Consumers are algorithm averse due to concerns about the algorithm's lack of personalization (Leung et al. 2018, Longoni et al. 2019, Luo et al. 2019). In other contexts, however, humans tend to prefer advice generated by an algorithm to that from a human (Bai et al. 2020, Lin et al. 2021, Logg et al. 2019), even if the algorithm's recommendations are wrong (Dijkstra et al. 1998, Dijkstra 1999). In this paper, managers are reluctant to use an algorithmic tool but, as opposed to the previous papers, we do not study whether that is driven by the algorithm's errors or the type of task, but whether that is related to attachment to a legacy heuristic (status quo bias), to the very large number of decisions to make, and to salience of the wrong metrics (inventory and speed of sales).

Another key aspect of human-algorithm interaction is interpretability (Bastani et al. 2018, Mišić and Perakis 2019). We study an intervention that increased adherence to an algorithmic tool's recommendations by helping managers interpret and understand them.

2.3. Adherence to Fixed Policies and Algorithmic Recommendations

A stream of literature related to our work studies adherence to processes and prescribed policies. Sometimes, deviating is a way of incorporating additional information into the decision-making process (Cui et al. 2015). Sometimes, it is good only if deviations are small (Tan and Staats 2016) or large (Fildes et al. 2009). In some contexts, deviations are linked to worse outcomes (Feng and Gao 2020, Ibanez et al. 2017, Kesavan and Kushwaha 2019, Phillips et al. 2015). These papers focus on the consequences of deviating. However, our goal is to study why managers deviated and, by doing so, to understand how to build better DSSs and processes.

With a similar goal, McLaughlin and Spiess (2022) study how algorithmic recommendations may shift a decision-maker's preferences for in risky decisions. They propose that the algorithm strategically withholds information to mitigate that inefficiency. We, instead, focus on drivers of deviations specific to the revenue management context, and the interventions we study deal with interpretability of the DSS's recommendation, as opposed to strategically withholding information.

Staats et al. (2016) analyze how monitoring improved healthcare providers' process compliance. Although our goal is similar, our setting is rather different: healthcare professionals must adhere to safety protocols, while for Zara it was a strategic decision to allow managers to deviate from price recommendations. Therefore, measures to increase adherence should focus on helping managers understand how the DSS can help them, but not on imposing adherence or monitoring them.

Deviations could be due to the fact that the algorithm is not capturing all the relevant costs, incentives, etc. Examples are Käki et al. (2019), Karlinsky-Shichor and Netzer (2019), Sun et al. (2019), van Beuningen (2018), and van Donselaar et al. (2010). In all these papers, the algorithms were improved after understanding why their users deviated from them. In our study, managers' deviations were not driven by a misalignment between the DSS and managers' costs or incentives. As we will show, the path to higher adherence did not consist in structural changes to the underlying optimization, but in small changes on the information presented to managers.

In our study, we show managers' reluctance to adhere to algorithmic recommendations may be driven by status quo bias, which has been well-studied in other decision-making contexts (Kahneman et al. 1991, Samuelson and Zeckhauser 1988). In the more specific setting of retailing, status quo bias has been listed several times, under names like "resistance to change" and "inertia", as a central challenges when it comes to successfully implementing analytics (Fisher and Raman 2010, Rooderkerk et al. 2021, Talluri and van Ryzin 2005).

3. Empirical Setting

3.1. Clearance Sales at Zara

Spanish retailer Zara, like many other fashion retailers, sets a clearance sales period at the end of every season (fall-winter, or W, and spring-summer, or S). This period usually lasts around 10-12 weeks, set by the company' CFO and the specific regulations of every country. Once the sales campaign begins, the prices of products change regularly (usually weekly) and these markdowns occur for all products in a country at the same time. The pricing decisions are the responsibility of one person per country, called the *country manager*, with the support of a small pricing committee.

These markdowns are constrained by some company rules. To list them, we first need to define some company-specific terminology. A *group* is a set of products of the same kind targeted towards a customer type. For instance, young women's knitwear, women's basic shirts, etc. There are a fixed number of groups, which are consistent over the years and across countries, whereas the product assortment within every group changes every new season. Within a group, a *cluster* is a set of SKUs which had the same price during the regular season. For instance, all the basic tops which were 19.95€ during the regular season form a cluster, even though they were not the exact same SKU, and all the basic tops which were 14.95€ are part of a different cluster.

One of the firm's pricing rules is that, during the clearance sales campaign, the price of each cluster cannot increase. Another rule is that clusters cannot be split, i.e., the 19.95€ basic tops cluster may experience several markdowns during the sales campaign, but all products within that cluster will always be marked down by the same amount at the same time. In addition, different clusters can converge to the same price, but never cross: the 19.95€ basic tops cluster, even after several markdowns, must be at least as expensive as than the 14.95€ basic tops cluster, because this was their order during the regular season. Finally, when the price of two or more clusters converge, these clusters will be coupled for the rest of the campaign and they become what is called a (clearance) *category*. For instance, if the 19.95€ cluster was marked down to 14.95€, and the 14.95€ cluster was unchanged, then both clusters become a category and would need to be marked down by the same amount every week until the end of clearance sales. These rules are driven by legal and practical reasons, detailed in Caro and Gallien (2012).

A particular characteristic of Zara's business model is that, in the countries where it is present, it either owns all the stores in the country (or a large majority) or all the stores there belong to a franchise. Some of the franchisee firms manage more than one country, while some others do so for only one. This distinction between country types will be important in our empirical analysis. More specifically, for own-store countries, pricing decisions were all made at Zara's headquarters in Spain, so the managers from those countries were arguably subject to peer effects regarding usage of the DSS. In contrast, pricing decisions for franchise countries were made at the franchisee company (with the support of a country representative located at Zara's Spain headquarters), so the final decision makers that managed those countries were separate from each other.

3.2. The Legacy Markdown Process

Prior to the DSS, country managers made their markdown decisions based on weekly reports. The report for a given country and group contained the most recent sales and inventory data for each clearance category (a screenshot of the report can be seen in the Appendix, Figure 5). The report also contained a metric known as the *rotation*, which was the predicted inventory run-out time for a category. The rotation was computed as the remaining inventory divided by the current sales rate. In principle, each category required a pricing decision every week: to markdown or not, and if yes, by how much. However, the number of pricing decisions could be reduced by merging two categories because they would no longer be allowed to split (see Section 3.1).

Zara did not have a mandatory markdown policy, but the prevailing practice was to introduce a markdown when either (i) the rotation — or inventory run-out time — exceeded the end of clearance sales, or (ii) the country manager wanted to merge two categories, which would reduce the number of categories and pricing decisions to be made. We refer to this policy as the managers' *heuristic*.

The sellthrough of a category, measured as the ratio of units sold to initial inventory, was known at Zara as *success* and was included in the last column of the weekly report. Though there was no set target, managers would usually aim for 90% success or more by the end of clearance sales.

3.3. Implementation of the DSS

In 2008, Zara adopted a sales pricing optimization system (Caro and Gallien 2012). This system built a demand forecast and used it in a dynamic program to find revenue-maximizing prices. The algorithm was run on a rolling-horizon basis and included the company's pricing rules as constraints (see Section 3.1). Zara considered human oversight to be important and gave the final decision on markdowns to the country managers. For this reason, the algorithm was embedded in a DSS that worked in the following way: every week, in addition to the legacy reports, the DSS provided a price recommendation for every category in a group and the corresponding revenue forecast. Managers could overrule and modify price, and the DSS would recompute its revenue forecast.

In S08 a pilot test of the DSS in two countries resulted in a 5.8% increase in clearance revenue. After that, the DSS was implemented gradually (until 2012) in all other countries; a timeline is shown in Figure 1. It was compulsory for managers from own-store countries to use this tool but they had the discretion to override its recommendations. Franchise countries had total freedom to set markdowns, although the DSS was also available to them. Managers' incentive scheme was proportional to the revenue they generated. The latter and the success of the S08 pilot made Zara believe that the country managers would organically follow the recommendations of the DSS. However, the adherence was lower than expected, as shown next.

3.4. Data Description

We use a dataset collected by Zara, which spans seven sales campaigns over 3.5 years, from W10 to W13. Our data contains 20 women's apparel groups. Due to Zara's constant expansion, the number of countries in our data increases from 56 in W10 to 84 in W13. Two countries were dismissed because Zara discontinued its presence there due to social and political conflicts. For each country,

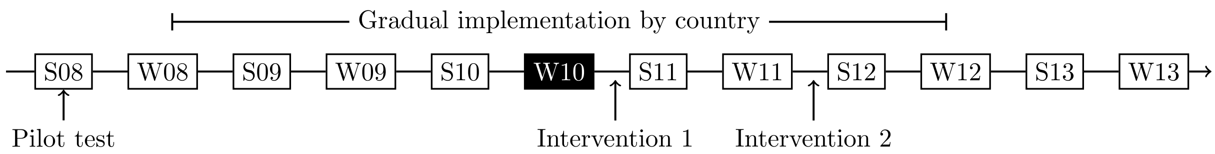


Figure 1 **Timeline of the DSS's deployment. Our data spans fall-winter 2010 (W10) to fall-winter 2013 (W13), and our main analysis is performed using the W10 campaign, before the interventions took place.**

campaign, and group, the dataset contains all pricing decisions made by the country manager for all clusters in that group and all weeks in that campaign. We also gathered information from Zara's annual reports regarding the number of stores in each country and year, as well as the type of country (own-stores or franchise). Table 1 shows a sample observation in the original dataset. The cluster-level data was aggregated to obtain 374,366 observations at the category level, which was the level at which pricing decision were made.

The W10 campaign deserves special attention because it is the data used in our main analysis (see Section 5). In W10, Zara operated in 56 countries, of which 34 were own-stores and 22 were franchises. The own-stores subsample represented 71.4% of the W10 observations, and in this subsample the country managers adhered to the DSS' price recommendation in 61.0% of the cases, and deviated by setting a lower (higher) price in 22.6% (16.4%) of the cases. In contrast, the franchise subsample represented 28.4% of the observations, DSS adherence of franchise managers was only 46.4%, and they deviated by setting a lower (higher) price in 38.5% (15.1%) of the cases. Additional summary statistics for W10 are given in Table 2. A common way in which deviations occurred was that the DSS recommended keeping the price of a category unchanged, but the manager marked it down anyway. In fact, as seen in Table 2, the median markdown recommended by the DSS was 0.0% (none) versus 19.9% chosen by the country managers.[1]

Country	**Campaign**	**Type**	**Num. stores**	**Group**	**Cluster**	**Week**
France	W10	Own stores	113	Women's outerwear	C	7
Initial inventory	**Quantity sold**	**Regular price**	**Previous price**	**DSS price**	**Confirmed price**	**Salvage value**
183	93	X€	Y€	Z€	T€	S€

Table 1 **A sample observation in the cluster-level dataset. The price information has been disguised.**

4. Hypotheses

As noted in Section 3.4, the adherence to the DSS price recommendations was low, especially for franchises. We argue that the low adherence can be (partially) explained by the country managers' resistance to change or inertia.

[1] In this paper, the weekly markdown is understood as the % change in price with respect to the previous week for the same category.

	Mean	Median	St. dev.	Min.	Max.
Summary of countries					
Number of stores	23.34	5	48.89	1	333
Experience using the DSS (number of campaigns)	1.27	1	0.84	0	5
Summary of categories					
Inventory at the beginning of the sales campaign	2,749.14	228	11,761.70	1	459,403
Weekly sales (units)	883.14	101	3,666.59	0	148,568
Markdown recommended by the DSS (%)	14.22	0	16.75	0	91.64
Markdown chosen by the country manager (%)	17.65	19.94	17.66	0	98.00
Absolute deviation from the DSS's recommendation (%)	13.18	0	21.31	0	632.65

Table 2 Summary statistics of the countries and categories in our data.

Note. Generated using the data corresponding to the fall-winter 2010 (W10) clearance sales campaign.

In the legacy markdown pricing process, country managers at Zara typically used the markdown heuristic described in Section 3.2. We posit that, after the DSS was implemented, managers actually continued to set prices with that heuristic in mind — trying to strike a balance between selling all inventory out before the end of clearance sales, and consolidating price categories to progressively simplify their task. If that was their goal, this would naturally lead them to deviate from the DSS's recommendations, which were aimed at maximizing revenue regardless of the inventory levels or the pricing categories. In other words, our overarching hypothesis is that country managers at Zara were reluctant to change from their old heuristic to the new system (the DSS), and also reluctant to change from their old mindset (minimizing inventory levels while reducing price categories) to the new system's goal (maximizing revenue), and that reluctance to change translated into deviations from the DSS's price recommendations.

In the literature, resistance to changing the way decisions are made is well-studied under the name *status quo bias* (Kahneman et al. 1991, Samuelson and Zeckhauser 1988). For managers, policy leaders, and other decision makers, the implementation of algorithmic solutions and DSSs is a departure from the status quo, in which decision-making relies on their expertise and discretion.

In the context of retail analytics and revenue management, status quo bias (or resistance to change) has often been reported as a hurdle in adopting algorithmic tools and, more generally, in becoming a data-driven organization (Bean 2022). As stated in Fisher and Raman (2010), implementing advanced analytics-based solutions in companies "requires people to change their thinking and behavior, and doing that is always challenging". Moreover, "[analysts are] suspicious that an automated system could replace their own intuition and experience" (Talluri and van Ryzin 2005). Similarly, the review by Rooderkerk et al. (2021) mentions inertia tied to legacy systems as one of the main organizational barriers to the successful implementation of analytics tools in retailing.

In the specific setting of clearance sales at Zara, the status quo was represented by the old heuristic, which was not revenue-maximizing. Instead, the heuristic tried to achieve two other goals, namely: (i) minimize inventory levels, and (ii) consolidate price categories. These two suboptimal goals inform our two hypotheses and, as we discuss next, are not unique to managers at Zara, and could affect users of any other revenue management DSS.

4.1. Salience of Inventory and Speed of Sales

Before the DSS was implemented, all the metrics in the managers' weekly report were related to inventory and speed of sales. The interface of the DSS was designed following the legacy reports and, therefore, the legacy metrics remained prominently displayed; see see Figure 2 in the appendix of Caro and Gallien (2012). Note that inventory is a count of something that is material and visible. Sales deplete inventory, making it visible as well. In contrast, revenue is an intangible metric that must be computed or forecasted, making it more abstract. For these reasons, inventory and speed of sales were likely more salient to country managers than revenue.

Salience is a well-documented bias (Tversky and Kahneman 1974). For instance, consumers make product choices differently when some attributes are more salient than others (Bordalo et al. 2013), or change their consumption completely when taxes are made salient (Chetty et al. 2009). Making consumption salient by providing real-time feedback helps consumers preserve resources, such as energy or water (Tiefenbeck et al. 2016). In warehouse workers, shifting salience from some KPIs to different ones greatly affects their productivity (Weerasinghe et al. 2021).

One of the key insights of pricing and revenue management is that it can be optimal to set prices in a way that leaves part of the inventory unsold at the end of the season as long as the overall revenue is maximized (Lazear 1986). However, when inventory is salient, not stocking out is counterintuitive. A similar tradeoff is observed in inventory management settings, in which managers often err on the side of understocking because excess inventory feels much worse to them (Hammond and Raman 1995). Newsvendor experiments have confirmed this observation. Making stockout costs more salient than leftover costs, or viceversa, changes managers' ordering behavior (Ho et al. 2010), and managers tend to deviate from profit-maximizing inventory decisions to minimize ex-post inventory errors (Schweitzer and Cachon 2000).

Zara's legacy markdown heuristic was based on the rotation metric (described in Section 3.2), which predicted the inventory run-out time, i.e., how long it would take to run out of inventory for a given category. The DSS displayed the inventory run-out time, which again, made inventory and speed of sales more salient than revenue. Hence, if managers exhibit salience bias, we would expect them to deviate from revenue-maximizing recommendations when the inventory run-out time (rotation) is greater. Based on this, we formulate our hypothesis on salience bias as follows:

Hypothesis 1. *Managers' probability of deviating from the DSS's revenue-maximizing prices, and the magnitude of these deviations, were higher when the inventory run-out time was larger.*

4.2. Cognitive Worlkoad

During clearance sales, managers at Zara set prices for a very large number of products. For instance, at the beginning of winter 2010's clearance sales there were over 200 price categories in our data for 20 groups. Hence, managers had hundreds of prices to set every week. Moreover, the number of possible price scenarios to consider grows dramatically when there are more categories within a group. In other words, the complexity of the pricing problem faced by a country manager increased with the number of categories. We posit two competing hypotheses based on how managers behaved when they faced a more complex problem given by a larger number of price categories.

The prevailing practice by Zara's country managers to deal with the complexity associated with a larger number of categories was to make the prices converge, i.e., merge the categories, very quickly. The pilot test in 2008 showed that merging the categories too fast was suboptimal, and that from a revenue-maximization standpoint it was better to leverage the range of prices and possible combinations available to its fullest extent. In fact, Table 4 in Caro and Gallien (2012) shows that revenue increased by using a wider set of prices within a group. Note that there is no additional benefit in consolidating prices (or merging categories) faster because Zara's official clearance sales duration is set, usually in line with country regulations, and cannot be altered unilaterally by the country manager. Hence, Caro and Gallien attributed the faster price consolidation to the country managers' "desire to simplify the problem structure because of time constraints and perhaps cognitive limitations".

Humans have indeed limited cognitive capacity in complex decision-making settings (Kahneman 1973). When information acquisition is costly it can be rational to deviate from optimality as a tradeoff with the effort of gathering and analyzing new data. This deviation from optimality is called rational inattention (Caplin and Dean 2015, Cheremukhin et al. 2015, Sims 2003). Even experienced decision-makers, such as investors, exhibit it (Akepanidtaworn et al. 2019).

In the context of Zara's markdown pricing, it could be rational for managers to refrain from analyzing all the possible price scenarios because processing all that information is costly, a form of rational inattention. An immediate implication under status quo bias is that, when the number of categories increased, managers would deviate from the DSS's revenue-maximizing prices by following the legacy heuristic. Put differently, if after the DSS implementation managers were still reducing their cognitive workload by merging categories, then we should see more deviations from its recommendations when the number of categories was larger.

Under the previous argument, country managers revert to the legacy heuristic when the problem is more complex as given by the number of price categories. A countervailing argument can be made in the absence of status quo bias. For instance, Snyder et al. (2022) find that workers are more likely to rely on algorithmic recommendations when their workload is higher. Similarly, country managers at Zara might want to avoid dealing with a more complex problem altogether. In that case, managers' adherence to the price recommendations of the DSS would increase when the number of categories is larger, since that strategy amounts to offloading their work entirely. In other words, when there are more categories, managers could be inclined to "let the algorithm do the thinking" and adopt the DSS's recommendations without trying to interfere. Note that in doing so, they effectively remove themselves from the decision-making process.

In sum, we argue that country managers might exhibit two types of behavior when faced with a larger number of price categories. They either follow the legacy heuristic and deviate more from the DSS's prices (compatible with status quo bias and rational inattention), or they give up control and deviate less (similar to the workers in Snyder et al. 2022). This reasoning leads us to formulate the following competing hypotheses:

Hypothesis 2. *Managers' probability of deviating from the DSS's revenue-maximizing prices, and the magnitude of these deviations, were higher when...*

(a) ...there were more categories in the group, i.e., more price decisions to make.

(b) ...there were fewer categories in the group, i.e., fewer price decisions to make.

5. Methods and Results

5.1. Methods

We use the two-step Heckman regression, which models a decision process like the one followed by managers at Zara when faced with the DSS. First, the manager decides whether to deviate or adhere to the DSS's recommendation. This is the *selection step* and it is modeled like a probit regression. Second, if they have decided to deviate, they decide by how much. This is the *deviation step* and is estimated as a linear regression with a term that corrects for the selection decision of the first step. The standard errors are robust, following Heckman (1979). Additional details of this method can be found in Cameron and Trivedi (2005), Greene (2012), and Wooldridge (2010).

We include the same variables in both the selection and deviation parts of the Heckman model. In the absence of exclusion restrictions, identification relies on the nonlinearity of the inverse Mills ratio, i.e., the term that corrects for the selection decision of the first step. This raises concerns of potential collinearity, but they are less grave if there is enough variation across observations in the argument of the Mills function (Cameron and Trivedi 2005, Wooldridge 2010); we show that this holds in the results discussion below (Section 5.2). We test alternative models in the E-Companion.

The Heckman regression model is traditionally used to correct for sample selection problems, i.e., when the dependent variable is only observable if the selection variable in the first step takes a value of 1. In our case, we always observe the recommended and confirmed prices, but we do not observe the price that the managers would have set if they had not adhered. In particular, we cannot know if the recommended price coincides with the one the manager already had in mind, or if, upon seeing the recommendation, they decided to implement the recommended prices and discard their previously preferred price. Hence, the selection step can be interpreted as the manager's propensity to "(not) believe in analytics", i.e., when they are not persuaded by the tool's recommendation. Note also that, when the same independent variables are used in the selection and the deviation steps, the Heckman model coincides with the Type II Tobit model, which accommodates situations in which the data does not suffer from a selection issue, but the continuous dependent variable has a large mass at zero, also called a corner solution; see Greene (2012) and Wooldridge (2010).

To focus on the drivers of managers' low adherence when the DSS was implemented, we estimate the Heckman regression using data from the W10 campaign only, before any of the interventions mentioned in Section 1 took place. We study those interventions in depth in Section 6.

5.1.1. Dependent Variables.

Selection step. The Heckman first step models the probability of deviating from the DSS's recommended price. Its dependent variable is binary, and takes value 1 if manager i's confirmed price for category c of group g in week w was different from the recommended one:

$$ProbDeviation_{igwc} = \mathbb{I}\left(Price_{igwc} \neq PriceDSS_{igwc}\right).$$

Deviation step. The second step models the amount of deviation. We compute the difference between the manager's confirmed price and the DSS's recommendation and, to make it comparable across products, we divide it by the DSS's recommended price. We take the absolute value of this normalized price difference to interpret our results in terms of price deviations, and not in terms of higher or lower prices. Hence, the dependent variable in the Heckman's deviation part is

$$AbsDeviation_{igwc} = \left|\frac{Price_{igwc} - PriceDSS_{igwc}}{PriceDSS_{igwc}}\right|.$$

5.1.2. Independent Variables. To test Hypothesis 1, we build the variable $StockoutWeek$ that emulates Zara's rotation metric. Specifically, we compute the period in which each category would be sold out if the speed of sales is the same as in the previous week. For manager i, week w, and category c of group g, it is defined as $StockoutWeek_{igwc} = w + \frac{Inventory_{igwc}}{Sales_{ig,w-1,c}}$. A large value of

$StockoutWeek$ means that at the current sales rate there will be leftover inventory at the end of clearance sales; in contrast, low values of this variable should reassure managers that the category will be sold out[2] As a robustness check, in the E-Companion we also test Hypothesis 1 using Zara's sellthrough (success) metric that was also displayed on the interface of the DSS.

To test Hypothesis 2, we use the independent variable $NumCategsDT_{igw}$, defined as follows. The number of product categories in a week and group is $NumCategs_{igw}$. Given the price consolidation rules set by the company (see Section 3.1), the number of categories within a group will always be decreasing — or at least nonincreasing — over time, which means that $NumCategs_{igw}$ reflects a monotonous time trend in addition to capturing pricing complexity. This could lead to inconclusive results due to collinearity with the week dummies. To correct for this issue, we detrend $NumCategs_{igw}$ by subtracting its weekly average $\widehat{NumCategs}_w$ across all managers and groups. The independent variable we ultimately use to test Hypothesis 2 is $NumCategsDT_{igw} = NumCategs_{igw} - \widehat{NumCategs}_w$.

As control variables, we include, in both parts of the Heckman estimation, a full set of dummy variables: week, product group, and country. These dummy variables account for differences across managers in characteristics such as experience using the DSS, overall experience at Zara, idiosyncratic tendency to "believe in analytics" and adhere to algorithmic recommendations, etc. They also account for differences across product groups (e.g., some groups may have a more predictable demand, therefore managers may trust the DSS's recommendations more or less for those groups), and differences over time (as the clearance sales campaign progresses, all prices and inventory levels decrease, which may affect adherence and magnitude of deviations).

5.2. Results

Table 3 shows the average partial effects (APEs) of the Heckman estimation. Note that, for the deviation step, they correspond to the coefficients of the linear model; for the selection step, the coefficients are available upon request. Robustness checks are reported in the E-Companion.

As shown in Table 3, the variable $StockoutWeek$ has a positive effect on both the probability of deviating as well as on the magnitude of deviations, and this effect is significant at the 0.001 level. For own-store countries, the average probability of deviating is 39.0% and the (unconditional) $AbsDeviation$ average is equal to 11.6%. Hence, the APEs in Table 3 mean that an additional week to stock out in a given category is associated with a 2.3% (0.91 percentage points) increase in the probability of deviating, and with price deviations that are 10.8% (1.25 percentage points) higher.

[2] When sales tend to zero, $StockoutWeek$ will tend to infinity. In that case, we cap $StockoutWeek$ at 24, which is twice the typical length of a clearance sales campaign..

	Heckman APEs					
	All countries (1)		Own stores (2)		Franchises (3)	
AbsDeviation						
StockoutWeek	0.0104***	(0.00076)	0.0125***	(0.00101)	0.0090***	(0.00121)
NumCategsDT	0.0075**	(0.00285)	0.0262***	(0.00418)	-0.0087	(0.00451)
Week dummies	Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes	
ProbDeviation						
StockoutWeek	0.0088***	(0.00099)	0.0091***	(0.00117)	0.0098***	(0.00184)
NumCategsDT	0.0127***	(0.00379)	0.0264***	(0.00479)	-0.0150*	(0.00670)
Week dummies	Yes		Yes		Yes	
Group dummies	Yes		Yes		Yes	
Country dummies	Yes		Yes		Yes	
Lambda	0.291***	(0.0193)	0.327***	(0.0223)	0.268***	(0.0258)
Rho	0.990		1.000		0.982	
Sigma	0.294		0.327		0.273	
N	22180		16534		5646	
N (deviated)	9775		6584		3191	

Table 3 Estimated average partial effects of the Heckman regression.

Note. Top half of the table: APEs (coefficients) of the deviation step with dependent variable *AbsDeviation*; bottom half: APEs of the selection step with dependent variable *ProbDeviation*. (1) includes all countries; (2), only own-store countries; (3), franchises only. Lambda, Rho and Sigma correspond to the Heckman correction. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In the case of franchises, with baseline probability of deviating of 53.6% and average deviation of 17.2%, an additional week to stock out is associated with a 1.8% (0.98 percentage points) increase in the probability of deviating, and with a 5.2% (0.90 percentage points) increase in the magnitude of deviations. These results provide strong support for Hypothesis 1.

The coefficients of the variable $NumCategsDT$ have opposite signs for own-store countries and franchises. For the former, the APEs of $NumCategsDT$ are positive and significant at the 0.01 level. An additional category within a product group is associated with a 6.8% (2.64 percentage point) increase in the probability of deviating, and with a 22.6% (2.62 percentage point) increase in the magnitude of deviations. This provides support for Hypothesis 2(a) in own-store countries.

In the case of franchises, the APEs of $NumCategsDT$ are negative but only weakly significant ($p < 0.05$) in the selection step, and insignificant in the deviation step. In other words, for every additional category within a group, franchise managers were 2.8% (1.5 percentage points) less likely to deviate from the DSS's recommendations, but the magnitude of their deviations was not affected. This provides partial support for Hypothesis 2(b). A possible explanation for this contrasting behavior is that franchise managers may have understood that the DSS was meant to be used when the markdown problem was more complex, so they were more likely to follow its recommendations when the number of pricing decisions was higher (although their overall adherence was lower than own-store countries). Also, franchise managers did not face peer comparisons and did not have to "prove themselves worthy" when the cognitive workload increased, which possibly facilitated delegating control to the DSS in those situations. In contrast, the same effects might have reinforced status quo bias for own-store managers.

As mentioned in Section 5.1 above, identification in the Heckman model when both steps contain the same variables can be threatened by collinearity, but this concern is mitigated if the argument of the lambda function, or Mills selection ratio, has enough variation across observations. To test whether that holds, we recover the argument of the lambda function as the sum of each variable times its corresponding coefficient in the Heckman selection step, and compute several measures of variation (or spread). For own-stores countries, the standard deviation of the lambda function argument is 1.29 times greater than its mean, and its inter-quartile range is 1.81 times greater than its median; this means that the argument of the lambda function has a fair amount of variation. For franchises, the argument's standard deviation is 2.83 times larger than its mean, and its inter-quartile range is 4.23 times larger than its median, which is a great amount of variation across observations. Therefore, it is safe to state that the collinearity concern is, at most, mild. Additionally, in the E-Companion we show the results of different regression models that are more robust in the presence of collinearity.

In sum, when the inventory run-out time was late in the campaign (or the category was expected to have leftover inventory by the end of clearance sales), managers from both country types were more likely to deviate from the DSS's recommendations and deviated by larger amounts. This evidence suggests that managers were overly concerned with selling everything, which can be explained by inventory levels being more salient to them than revenue. When there were a large number of price categories within a product group, managers in own-store countries were more likely to deviate from the DSS's recommendations, and did so by larger amounts, whereas managers from franchises deviated less. Overall, managers presented resistance to change from their legacy heuristic to the new, analytics-based decision-making that the DSS enabled. We now present further evidence that managers' deviations from the DSS were in line with their old heuristic.

5.3. Persistence of the Legacy Heuristic

Figure 2 shows boxplots of managers' decisions in terms of markdown and number of categories, as compared to the DSS's recommendations, plotted using the observations in which managers deviated from the DSS. In other words, Figure 2 shows what decisions managers ultimately made when they decided to deviate from the revenue-maximizing prices.

As we can see in the left plot of the figure, both managers of own-store countries and managers of franchises set more aggressive (higher) markdowns than what the DSS recommended; however, the difference between recommended and confirmed markdowns is much larger for franchises. This is another clear indication that managers were still following the legacy heuristic as they set higher markdowns to liquidate the remaining inventory as opposed to maximizing revenue.

In the right plot of Figure 2 we can see managers' decisions in terms of number of categories. Managers in both country types had the tendency to consolidate more prices to have fewer categories than recommended (a difference of one category, in median). Interestingly, we see that when managers in franchise countries deviated, they set fewer categories than what was recommended, even though — as we have seen in Section 5.2 — when there were a large number of categories their probability of deviating was lower. In other words, when the number of categories was large, managers of franchise countries tended to deviate less, but when they deviated, they set fewer categories. Again, this evidence supports the claim that the managers were following the legacy heuristic when they deviated from the DSS.

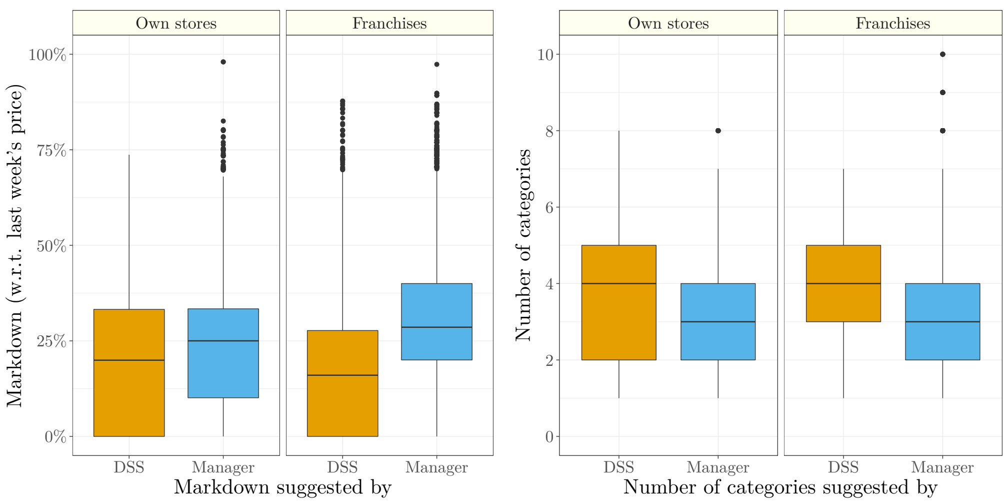


Figure 2 DSS's recommendation compared to managers' decision when there was a deviation.

Note. Plot uses observations in which managers deviated from the DSS during the W10 clearance sales. Left plot: weekly markdown recommended by the DSS (orange box) vs. markdown finally decided by the manager (blue box). Right plot: number of categories within a group, recommended by the DSS (orange box) vs. finally decided by the manager (blue box). Within each plot, the left facet corresponds to own-store countries and the right facet corresponds to franchises.

6. Effect of Two Interventions

6.1. Context

The results from the W10 campaign showed that inventory salience was associated with deviations from the recommended markdowns; see Section 5. Hence, Zara decided to perform two interventions in the form of changes to the interface of the DSS. The objective was to address inventory salience to increase adherence and, ultimately, revenue.

6.1.1. Intervention 1: Feedback. At Zara, the key performance indicator during clearance sales is the metric Y, which is the ratio between clearance sales revenue and the initial inventory valued at regular season prices. The objective of the DSS is to maximize revenue, which is equivalent to maximizing the metric Y because the denominator is a constant.

In S11 the DSS started displaying the revenue metric Y. After this intervention, managers were able to see feedback of their own performance via the realized revenue metric Y up to the current week for every product group. Highlighting the metric Y was expected to counteract inventory salience so managers would be nudged to maximize revenue.

6.1.2. Intervention 2: Interpretability. The Y metric that managers were shown after Intervention 1 took values between 0 and 100. This normalization was meant to make it comparable across product groups and across countries. However, even when a metric is normalized, it may be difficult to interpret. In S12 Zara made a second change to the DSS's interface by showing a reference point for the metric Y. The benchmark was the manager's own Y in the previous year for the same group and week. The goal of the intervention was to make the value of Y easier to interpret. Providing a reference point would allow managers to benchmark their performance at each point in time and for each group. It was expected that facilitating an easy comparison of the Y metric would replace or overshadow the logic of the legacy heuristic in which the inventory run-out time was measured against the end of clearance sales. A screenshot of the DSS's interface after the interventions is shown in the Appendix, Figure 6.

6.2. Methods

The company rolled out the interventions for all country managers at the same time, which prevents us from establishing a causal effect like we could in a randomized experiment. However, we can exploit the country types to study the effect of the interventions. In fact, own-store country managers actively provided feedback on the DSS's interface and helped improve it. Hence, for own-store country managers the interventions were mostly endogenous. In contrast, franchise country managers did not participate in the development of the DSS, so the two interventions were exogenous to them, and mostly targeted to them, as they were the ones showing the lowest adherence. Therefore, we label own-store countries as the control group and franchise countries as the treatment group, and use the difference-in-differences (DiD) estimator to compute the effect of each intervention on franchise countries versus own-store countries.

Given that the assortment of categories within groups changes every season, but groups do not, we need to aggregate our outcome of interest at the product group level to make the units of observation comparable in the pre- and post- intervention periods. We define the group-aggregate adherence $Adherence_{itg}$ as the fraction of all the pricing decisions of manager i and campaign t for group g for which the manager adhered to the DSS's recommended price.

We estimate the following linear regression using OLS:

$$\begin{aligned} Adherence_{itg} = & \alpha + \beta_1 \times Int1_t \times Franchise_i + \beta_2 \times Int2_t \times Franchise_i + \\ & \gamma_1 \times Int1_t + \gamma_2 \times Int2_t + \delta \times Franchise_i + \xi \times X_{it} + \varepsilon_{itg}, \end{aligned} \tag{1}$$

where the estimates of β_1 and β_2 are the DiD coefficients of interest and X_{it} contains the control variables $ExperienceWithDSS_{it}$, how many clearance sales campaigns had country i's manager been using the DSS, not counting t, and $LogNumStores_{it}$, how many physical stores did Zara have in country i during campaign t, in log scale. Standard errors are clustered at the country level.

The model in Equation (1) does not control for the variation in adherence across managers or product groups. In addition, it does not capture changes in overall adherence between campaigns that could be mistakenly lumped into the interventions' dummy variables. Hence, a more refined estimate can be obtained by including group dummies (ζ_g), replacing the intervention indicators with campaign dummies (η_t), and replacing the binary $Franchise_i$ by country dummies (θ_i):

$$\begin{aligned} Adherence_{itg} = & \alpha + \beta_1 \times Int1_t \times Franchise_i + \beta_2 \times Int2_t \times Franchise_i + \\ & \xi \times X_{it} + \zeta_g + \eta_t + \theta_i + \varepsilon_{itg}, \end{aligned} \tag{2}$$

To account for heterogeneity of the interventions' effects, we label all country managers according to their pre-Intervention 1 adherence by creating the dummy variables $AdhPreInt1Qk_i$ such that managers whose adherence before Intervention 1 was in the lowest 25% have $AdhPreInt1Q1_i = 1$; managers in the second quartile, $AdhPreInt1Q2_i = 1$; and so on. Then, we compute the effect of Intervention 1 in each one of these segments by interacting the quartile dummy variables with our main term, $Int1_t \times Franchise_i$. Note that for this regression we can only use countries in which the DSS was implemented before Intervention 1, as $AdhPreInt1Qk_i$ are undefined otherwise. Analogously, for Intervention 2 we compute the variables $AdhPreInt2Qk_i$ and interact them with the corresponding DiD coefficient. The E-Companion contains robustness checks of the DiD approach, a similar analysis looking at the markdowns conditional on deviating from the DSS, and an analysis of the effect of the interventions using the synthetic control method.

6.3. Results

Figure 3 shows managers' adherence per campaign, for own-store countries and franchises separately. The average adherence of these two country types did not change after Intervention 1. However, after Intervention 2, franchises' adherence increased dramatically.

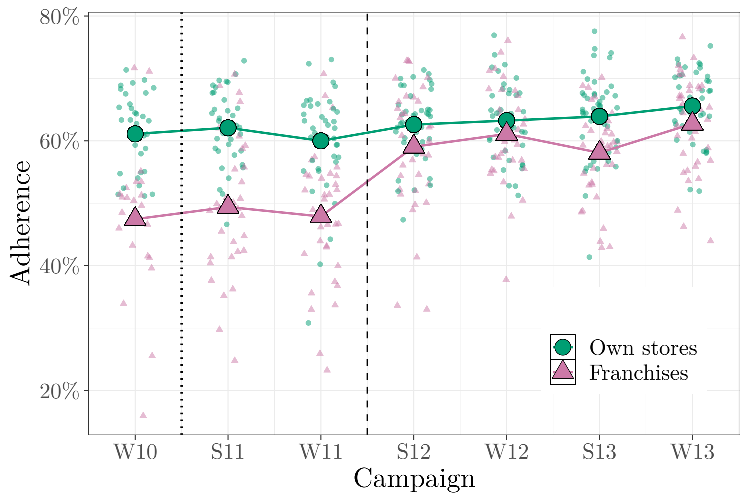


Figure 3 **Managers' adherence to the DSS's recommendations by campaign, for own-store countries (green, circle) and franchises (pink, triangle).**

Note. The small points correspond to $Adherence_{it}$ of a country manager in each campaign on the horizontal axis. The larger points connected by lines correspond to the average $\overline{Adherence}_t$ for own-store country managers (green) and franchises (pink) in each campaign. The dotted line marks when Intervention 1 took place (S11), and the dashed line marks Intervention 2 (S12).

Table 4 confirms the observation in Figure 3. Intervention 1 did not have a significant effect on franchises ($p \geq 0.05$). In contrast, Intervention 2 had a positive effect ($p < 0.001$). When this intervention took place, franchises' adherence increased between 9.4 and 9.6 percentage points. In regression (4) we can see that this effect was uneven: franchise managers who were in the bottom quartile of pre-intervention adherence increased their post-intervention adherence by 15.3 percentage points; the second quartile's adherence increased by 5.51 percentage points; the third quartile's did not change; and the top quartile's adherence decreased by 4.69 percentage points.

The takeaway from Figure 3 and Table 4 is that Intervention 1 did not have a robust effect on managers' adherence but Intervention 2 did, and this effect was larger for country managers that had the lowest pre-intervention adherence. Intervention 2 leveled up the adherence of franchises on average, reaching a level similar to own-store countries.

	DiD coefficients Adherence							
	(1)		(2)		(3)		(4)	
Int1×Franchise	0.0169	(0.0269)	0.0110	(0.0216)			0.0132	(0.0216)
Int2×Franchise	0.0942***	(0.0168)	0.0963***	(0.0172)	0.0932***	(0.0213)		
Int1×Franchise×AdhPreInt1Q1					0.0525	(0.0304)		
Int1×Franchise×AdhPreInt1Q2					-0.0149	(0.0274)		
Int1×Franchise×AdhPreInt1Q3					0.0104	(0.0150)		
Int1×Franchise×AdhPreInt1Q4					-0.161***	(0.0454)		
Int2×Franchise×AdhPreInt2Q1							0.153***	(0.0228)
Int2×Franchise×AdhPreInt2Q2							0.0554***	(0.0154)
Int2×Franchise×AdhPreInt2Q3							0.000826	(0.0229)
Int2×Franchise×AdhPreInt2Q4							-0.0469***	(0.00759)
Int1	-0.0204*	(0.0101)						
Int2	-0.0146	(0.00959)						
Franchise	-0.136***	(0.0292)						
ExperienceWithDSS	0.0133***	(0.00223)	0.266***	(0.0490)	0.420***	(0.0938)	0.323***	(0.0764)
LogNumStores	0.00145	(0.00454)	0.0188	(0.0210)	-0.0462	(0.0399)	-0.00526	(0.0327)
Constant	0.582***	(0.0158)						
Group dummies	No		Yes		Yes		Yes	
Campaign dummies	No		Yes		Yes		Yes	
Country dummies	No		Yes		Yes		Yes	
N	10431		10431		7812		9991	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4 Difference-in-differences coefficient estimates.

Note. Models (1) and (2) correspond to Equations (1) and (2) respectively. Models (3) and (4) account for pre-intervention adherence heterogeneity. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

6.4. Discussion

Feedback has a positive impact on decision-making (Bolton and Katok 2008, Hogarth 1981, Kayande et al. 2008, Kleinmuntz 1985), and interventions that provide feeedback to shift salience to the metrics that are important have been proven to affect the behavior of many different types of decision makers (Bordalo et al. 2013, Chetty et al. 2009, Gilbert and Graff Zivin 2014, Tiefenbeck et al. 2016, Weerasinghe et al. 2021), in particular by helping them set better goals (Locke et al. 1968, Locke and Latham 2006). However, numeracy research shows that decision makers may not use numbers until they are contrasted with available data or reference points to determine their meaning (Barrio et al. 2016, Mcintosh et al. 1992, Peters et al. 2006, Peters 2012). What our results suggest is that managers did not fully understand the revenue metric Y in isolation when it was introduced as feedback in Intervention 1, but increasing its interpretability in Intervention 2 may have been the key that allowed them to make sense of it.

One could conjecture that the interventions for the franchises served as a reminder that there was a DSS available to take advantage of, and that the higher adherence was just a result of higher usage of the DSS. In that case, the specific changes to the interface would not really matter. However, if a reminder effect is the main driver, then one would expect the first intervention to be associated with an increase in adherence, but it was not. And if Intervention 1 did not produce a significant reminder effect, there is no particular reason to think that Intervention 2 would have been any different. Hence, it is safe to rule out this alternative explanation.

6.5. Aftermath

Ultimately, the interventions were deployed because Zara wanted higher adherence to the DSS to achieve higher revenue. Is this what happened? In Figure 4, we can see the relationship between adherence to the DSS's recommendations and the revenue metric Y at the end of clearance sales, in W10 (before the interventions), and W13 (after the interventions). Compared to W10, the adherence range in W13 not only shifted to the right — meaning higher adherence overall — but it also became narrower, which means that there was more consistency across countries. In particular, the franchises were very disperse in W10, but in W13 they clustered together in the same upper right corner as the own-store countries. Following the higher adherence, the Y metric also improved overall, which was expected based on the 2008 pilot reported in Caro and Gallien (2012).

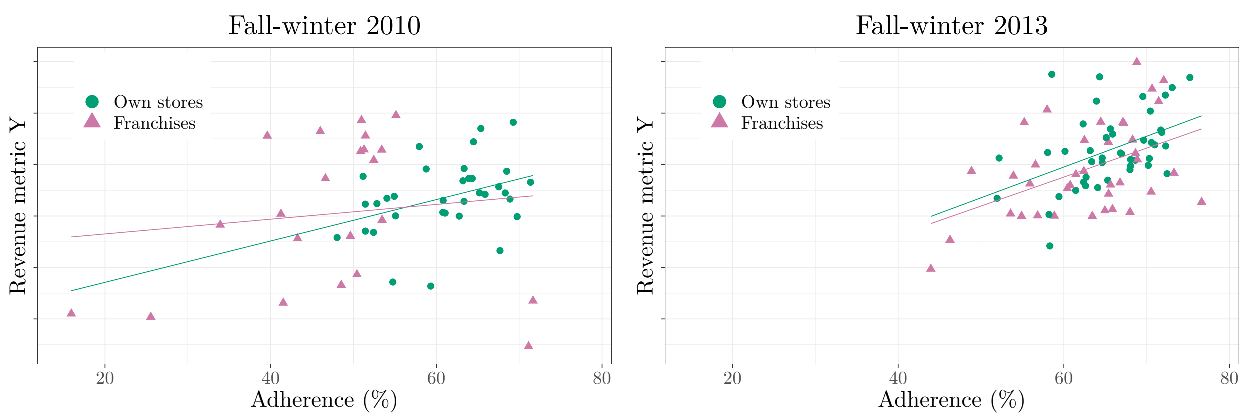


Figure 4 Relationship between average adherence to the DSS's recommendations and revenue as measured by the Y metric in W10, before the interventions (left), and W13, after the interventions (right).

Note. The green dots correspond to managers from countries in which Zara owns the stores, and the pink triangles correspond to managers from franchise countries. The exact values of the revenue metric Y have been disguised.

7. Conclusion and Managerial Insights

A large portion of revenue management and pricing research is aimed at providing advanced analytics tools to assist managers in their price decision-making. Many companies choose not to fully automate said decisions and, instead, implement their analytics tools in the form of DSS, tools that make recommendations but allow managers to deviate from them. In few instances, managers have access to private information that is unknown to the DSS, such as trends, special days, etc., and it is valuable that they deviate from the tool's recommendations. However, a competent DSS on average should make optimal recommendations, and managers overriding them (when they should adhere) translates into foregone revenues. The success of a DSS's implementation, and its effect on the firm's performance, depends on decision makers' adherence to its suggestions. Therefore, it is critical that we understand how human decision makers interact with such tools to design them in a way that entices managers to adhere to their recommendations.

Several biases could affect revenue managers that use a DSS, undermining its successful implementation. First, status quo bias: humans are reluctant to change, and the implementation of an advanced analytics tool represents a big jump from their previous intuition-based pricing process. Second, salience of inventory: it is counterintuitive that, to maximize revenues, prices may need to be high, and that this may lead to unsold inventory; this may be worsened by inventory levels being much more salient than revenue forecasts that are only realized at the end of the season. Third, cognitive limitations: due to companies' increasingly broad assortments, setting prices for each product may be a daunting task for managers, who may choose suboptimal prices if that simplifies their cognitive workload, or they might instead give up control and let the DSS be the decision maker.

In this paper, we used data collected by Zara during seven clearance sales campaigns to study managers' adherence to price recommendations from a DSS that maximized revenue. We show that lower levels of adherence are consistent with inventory salience bias: when inventory levels were high or speed of sales was low, managers were more likely to deviate from the DSS's optimal prices. When there was a very large number of prices to set (a measure of complexity), managers of countries where Zara owns the stores deviated more to simplify their workload, but franchise managers deviated less and delegated the larger workload to the DSS. To tackle the salience of inventory and increase adherence to the DSS's price recommendations, Zara performed two sequential interventions in the form of changes in the DSS's interface. We studied the effect of the interventions using a DiD analysis. We find that the second intervention, i.e., providing not only feedback on managers' realized revenue but also making it more interpretable, had a strong and positive effect on the DSS adherence of franchise managers.

Our findings provide a few managerial insights for any company that wants to implement a pricing DSS to be adopted voluntarily by its users. First, significant attention should be placed on designing the DSS's interface. The interface of Zara's DSS replicated the legacy reports, which fomented status quo bias. Similarly, to mitigate inventory salience, the DSS's interface should emphasize metrics that are aligned with DSS's goal, e.g., maximize revenues or profits. Second, managers should understand that the DSS's purpose is to help them make better decisions. Firms that want to implement a DSS but leave the final decision to managers should be mindful of humans' cognitive capacity, and minimize the amount of information that managers need to process.

Finally, it is necessary to educate managers to accept the changes when moving from purely manual processes to processes that are embedded with algorithms, so that they will build trust in analytics.

Our work has, of course, numerous limitations. First, while we can hypothesize that salience bias and large cognitive workload were the mechanisms driving the deviations from optimality that we have reported empirically, it is very hard to prove that a specific behavioral bias is driving an observed phenomenon using secondary data — in our study, we cannot manipulate the specific conditions in which managers were setting prices, as they were driven by customer and company dynamics. Control variables help us rule out as many confounding effects as possible, but that is the most we can control for in this setting. As such, the attribution of certain observed phenomena to specific behavioral biases cannot be proven, just hypothesized like we did in Section 4. A similar approach to documenting real-life deviations from rationality and arguing that those are compatible with certain behavioral biases is also used in Camerer (1998). Despite these limitations, there is value in showing that field data is consistent with well-known behavioral biases: we provide empirical evidence that status quo bias can hinder a DSS's implementation, and that salience bias and high cognitive workload might interfere with revenue management goals. These are facts that until now had mostly been reported anecdotally in the literature. We expect that future research will continue to examine the interplay between these biases and the usage of analytics tools.

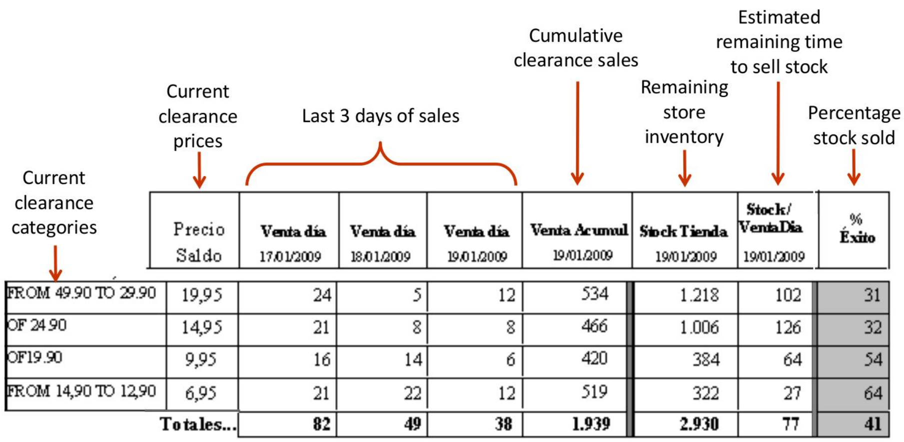
Second, regarding the effect of the interventions, the lack of randomization and a true control group prevents us from obtaining a clean, causal estimate of the effect of the interventions. Despite the interventions being endogenous to own-store country managers, they could also have an effect on those managers — in the end, they asked for those changes. This means that our DiD coefficient estimates may be downward biased. Our results regarding the interventions are only preliminary in nature, but they can inform studies on interventions using controlled experiments to obtain a more conclusive answer.

In closing, our work opens up questions that could be explored in future research. More biases in pricing could be studied. For instance, overconfidence, or managers' tendency to overestimate the goodness of their own decisions compared to those recommended by someone else (Galasso and Simcoe 2011). Another topic could be understanding when does a high cognitive workload induce a manager to simplify the problem, e.g., by using a heuristic, or give up control, as shown by the two competing hypotheses in Section 4.2. Finally, adherence to the optimal prices remained below 70% on average after the interventions. Hence, our study raises the question of what other types of changes to a DSS's interface, and in the information it presents to its users, have an effect on the way managers interact with it and increase adherence to its recommendations.

Acknowledgments

We are grateful to the editors and anonymous reviewers for their great comments and suggestions to the previous three versions of this paper.

Appendix. Screenshots of the Legacy Reports and the DSS



	Precio Saldo	Venta día 17/01/2009	Venta día 18/01/2009	Venta día 19/01/2009	Venta Acumul 19/01/2009	Stock Tienda 19/01/2009	Stock/ VentaDía 19/01/2009	% Éxito
FROM 49.90 TO 29.90	19,95	24	5	12	534	1.218	102	31
OF 24.90	14,95	21	8	8	466	1.006	126	32
OF19.90	9,95	16	14	6	420	384	64	54
FROM 14,90 TO 12,90	6,95	21	22	12	519	322	27	64
	Totales...	**82**	**49**	**38**	**1.939**	**2.930**	**77**	**41**

Figure 5 **One of the weekly inventory and sales reports that helped managers set markdowns before the DSS's implementation. The last two columns correspond to the rotation and success (see Section 3.3).**

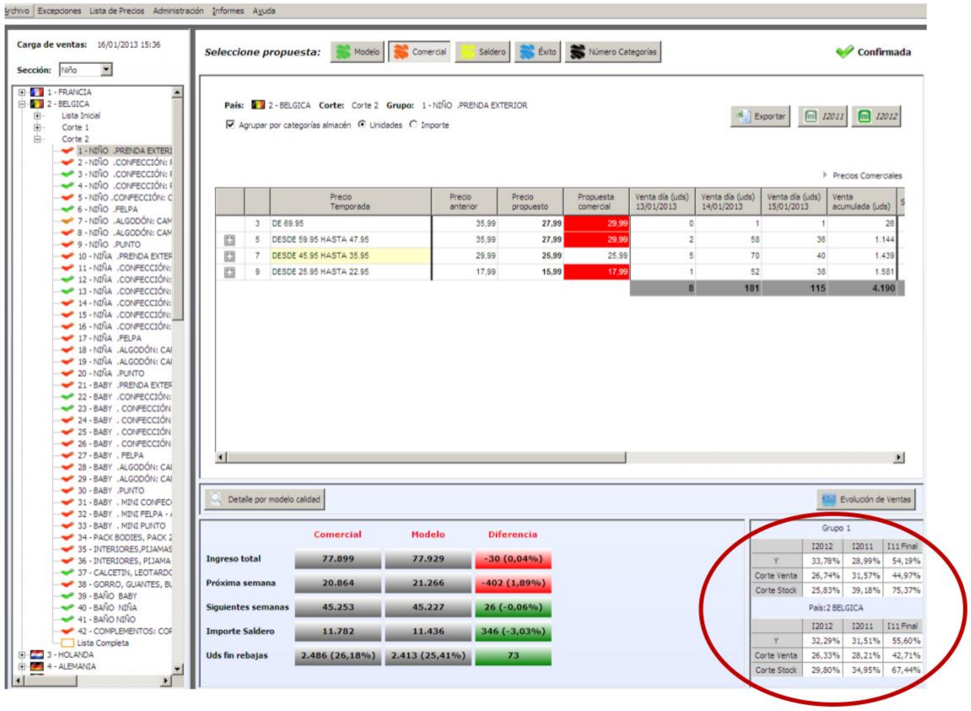


Figure 6 **The DSS's interface after the interventions. The top area contains the inventory and sales reports. The bottom left area contains the price recommendation and the managers' confirmed price, and their respective revenue and sales forecasts (for different horizons). The bottom right area contains what was added during the interventions: the Y metric for that group and country (first column of the tables), plus the Y metric for that group and country in the same week of the previous year as a reference point (second column), and at the end of the season corresponding to the reference point (third column).**

References

Achabal DD, McIntyre SH, Smith SA, Kalyanam K (2000) A Decision Support System for Vendor Managed Inventory. *Journal of Retailing* 76(4):430–454.

Aflaki A, Feldman P, Swinney R (2019) Becoming Strategic: Endogenous Consumer Time Preferences and Multiperiod Pricing, Working Paper.

Akepanidtaworn K, Di Mascio R, Imas A, Schmidt L (2019) Selling Fast and Buying Slow: Heuristics and Trading Performance of Institutional Investors. *Working paper* .

Anderson ET, Simester DI (2010) Price Stickiness and Customer Antagonism. *The Quarterly Journal of Economics* 125(2):729–765.

Arvan M, Fahimnia B, Reisi M, Siemsen E (2019) Integrating Human Judgement into Quantitative Forecasting Methods: A Review. *Omega* 86:237–252.

Aviv Y, Pazgal A (2008) Optimal Pricing of Seasonal Products in the Presence of Forward-Looking Consumers. *Manufacturing & Service Operations Management* 10(3):339–359.

Bai B, Dai H, Zhang D, Zhang F, Hu H (2020) The Impacts of Algorithmic Work Assignment on Fairness Perceptions and Productivity: Evidence from Field Experiments, Working Paper.

Ball RT, Ghysels E (2017) Automated Earnings Forecasts: Beat Analysts or Combine and Conquer? *Management Science* 64(10):4936–4952.

Barrio PJ, Goldstein DG, Hofman JM (2016) Improving Comprehension of Numbers in the News. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (ACM).

Bastani H, Bastani O, Kim C (2018) Interpreting Predictive Models for Human-in-the-Loop Analytics. *arXiv Preprint arXiv:1705.08504* .

Bean R (2022) Why Becoming a Data-Driven Organization Is So Hard. *Harvard Business Review* .

Bearden JN, Murphy RO, 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):62–71.

Bendoly E, Donohue K, Schultz KL (2006) Behavior in Operations Management: Assessing Recent Findings and Revisiting Old Assumptions. *Journal of Operations Management* 24(6):737–752.

Bertrand M, Schoar A (2003) Managing with Style: The Effect of Managers on Firm Policies. *The Quarterly Journal of Economics* 118(4):1169–1208.

Blattberg RC, Hoch SJ (1990) Database Models and Managerial Intuition: 50% Model + 50% Manager. *Management Science* 36(8):887–899.

Bolton GE, 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.

Bordalo P, Gennaioli N, Shleifer A (2013) Salience and Consumer Choice. *Journal of Political Economy* 121(5):803–843.

Camerer CF (1998) Prospect Theory in the Wild: Evidence from the Field. *Caltech Social Science Working Paper* .

Cameron AC, Trivedi PK (2005) *Microeconometrics: Methods and Applications* (Cambridge University Press), 1st edition.

Caplin A, Dean M (2015) Revealed Preference, Rational Inattention, and Costly Information Acquisition. *American Economic Review* 105(7):2183–2203.

Caro F, Gallien J (2012) Clearance Pricing Optimization for a Fast-Fashion Retailer. *Operations Research* 60(6):1404–1422.

Castelo N, Bos MW, Lehmann DR (2019) Task-Dependent Algorithm Aversion. *Journal of Marketing Research* 56(5):809–825.

Che H, Sudhir K, Seetharaman P (2007) Bounded Rationality in Pricing Under State-Dependent Demand: Do Firms Look Ahead, and If so, How Far? *Journal of Marketing Research* 44(3):434–449.

Cheremukhin A, Popova A, Tutino A (2015) A Theory of Discrete Choice with Information Costs. *Journal of Economic Behavior & Organization* 113:34–50.

Chetty R, Looney A, Kroft K (2009) Salience and Taxation: Theory and Evidence. *American Economic Review* 99(4):1145–1177.

Cui R, Allon G, Bassamboo A, van Mieghem JA (2015) Information Sharing in Supply Chains: An Empirical and Theoretical Valuation. *Management Science* 61(11):2803–2824.

DellaVigna S, Gentzkow M (2019) Uniform Pricing in U.S. Retail Chains. *The Quarterly Journal of Economics* 134(4):2011–2084.

Dietvorst B, Bharti S (2019) People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error, Working Paper.

Dietvorst BJ, Simmons JP, Massey C (2016) Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science* 64(3):1155–1170.

Dijkstra JJ (1999) User agreement with incorrect expert system advice. *Behaviour & Information Technology* 18(6):399–411.

Dijkstra JJ, Liebrand WBG, Timminga E (1998) Persuasiveness of expert systems. *Behaviour & Information Technology* 17(3):155–163.

Donohue K, Özer Ö, Zheng Y (2019) Behavioral Operations: Past, Present, and Future. *Manufacturing & Service Operations Management* Articles in Advance.

Donohue K, Schultz K (2018) The Future Is Bright. *The Handbook of Behavioral Operations*, chapter 18, 619–651 (John Wiley & Sons, Ltd).

Eisen MB (2011) Amazon's $23,698,655.93 Book About Flies. *Michael Eisen's Blog* URL `http://www.michaeleisen.org/blog/?p=358`.

Elmaghraby W, Jank W, Karaesmen IZ, Zhang S (2012) An exploratory analysis of B2B price changes. *Journal of Revenue and Pricing Management* 11(6):607–624.

Elmaghraby W, Jank W, Zhang S, Karaesmen IZ (2015) Sales Force Behavior, Pricing Information, and Pricing Decisions. *Manufacturing & Service Operations Management* 17(4):495–510.

Fahimnia B, Pournader M, Siemsen E, Bendoly E, Wang C (2019) Behavioral Operations and Supply Chain Management–A Review and Literature Mapping. *Decision Sciences* 50(6):1127–1183.

Feng X, Gao J (2020) Is Optimal Recommendation the Best? A Laboratory Investigation Under the Newsvendor Problem. *Decision Support Systems* 131:113251.

Fildes R, Goodwin P, Lawrence M, Nikolopoulos K (2009) Effective Forecasting and Judgmental Adjustments: An Empirical Evaluation and Strategies for Improvement in Supply-Chain Planning. *International Journal of Forecasting* 25(1):3–23.

Fisher M, Raman A (2010) *The New Science of Retailing: How Analytics Are Transforming the Supply Chain and Improving Performance* (Harvard Business Press).

Flicker B (2018) Managerial Insight and "Optimal" Algorithms. *Working paper* .

Galasso A, Simcoe TS (2011) CEO Overconfidence and Innovation. *Management Science* 57(8):1469–1484.

Gallino S, Moreno A, Rooderkerk RP (2019) Omnichannel Fulfillment Dilemmas: Customer Preferences and Manager Perceptions, Working Paper.

Gilbert B, Graff Zivin J (2014) Dynamic Salience with Intermittent Billing: Evidence from Smart Electricity Meters. *Journal of Economic Behavior & Organization* 107:176–190.

Gino F, Pisano G (2008) Toward a Theory of Behavioral Operations. *Manufacturing & Service Operations Management* 10(4):676–691.

Goldfarb A, Ho TH, Amaldoss W, Brown AL, Chen Y, Cui TH, Galasso A, Hossain T, Hsu M, Lim N, Xiao M, Yang B (2012) Behavioral Models of Managerial Decision-Making. *Marketing Letters* 23(2):405–421.

Greene WH (2012) *Econometric Analysis* (Prentice Hall), 7th edition.

Hammond JH, Raman A (1995) Sport Obermeyer, Ltd. TN. Technical report, Harvard Business School.

Hardcopf R, Gonçalves P, Linderman K, Bendoly E (2017) Short-Term Bias and Strategic Misalignment in Operational Solutions: Perceptions, Tendencies, and Traps. *European Journal of Operational Research* 258(3):1004–1021.

Heckman JJ (1979) Sample Selection Bias as a Specification Error. *Econometrica* 47(1):153–161.

Ho TH, Lim N, Cui TH (2010) Reference Dependence in Multilocation Newsvendor Models: A Structural Analysis. *Management Science* 56(11):1891–1910.

Hoch SJ, Schkade DA (1996) A Psychological Approach to Decision Support Systems. *Management Science* 42(1):51–64.

Hogarth RM (1981) Beyond Discrete Biases: Functional and Dysfunctional Aspects of Judgmental Heuristics. *Psychological Bulletin* 90(2):197–217.

Ibanez MR, Clark JR, Huckman RS, Staats BR (2017) Discretionary Task Ordering: Queue Management in Radiological Services. *Management Science* 64(9):4389–4407.

Ibrahim R, Kim SH (2019) Is Expert Input Valuable? The Case of Predicting Surgery Duration, Working Paper.

Ibrahim R, Kim SH, Tong J (2021) Eliciting Human Judgment for Prediction Algorithms. *Management Science* 67(4):2314–2325.

Kahneman D (1973) *Attention and Effort* (Prentice-Hall).

Kahneman D, Knetsch JL, Thaler RH (1991) Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives* 5(1):193–206.

Käki A, Kemppainen K, Liesiö J (2019) What to do when decision-makers deviate from model recommendations? Empirical evidence from hydropower industry. *European Journal of Operational Research* 278(3):869–882.

Karlinsky-Shichor Y, Netzer O (2019) Automating the B2B Salesperson Pricing Decisions: Can Machines Replace Humans and When?, Working Paper.

Kayande U, De Bruyn A, Lilien GL, Rangaswamy A, van Bruggen GH (2008) How Incorporating Feedback Mechanisms in a DSS Affects DSS Evaluations. *Information Systems Research* 20(4):527–546.

Kazerouni A, van Roy B (2017) Learning to Price with Reference Effects, Working Paper.

Kesavan S, Kushwaha T (2019) Field Experiment on the Profit Implications of Merchants' Discretionary Power to Override Data-Driven Decision-Making Tools, Working Paper.

Kleinmuntz DN (1985) Cognitive Heuristics and Feedback in a Dynamic Decision Environment. *Management Science* 31(6):680–702.

Kocabıyıkoğlu A, Göğüş CI, Gönül MS (2015) Revenue Management vs. Newsvendor Decisions: Does Behavioral Response Mirror Normative Equivalence? *Production and Operations Management* 24(5):750–761.

Kocabıyıkoğlu A, Göğüş CI, Hekimoğlu MH (2018) The Impact of Decision Types on Revenue Management Decisions: An Experimental Study. *Decision Sciences* 49(2):225–249.

Kremer M, Mantin B, Ovchinnikov A (2017) Dynamic Pricing in the Presence of Myopic and Strategic Consumers: Theory and Experiment. *Production and Operations Management* 26(1):116–133.

Lazear EP (1986) Retail Pricing and Clearance Sales. *American Economic Review* 76(1):14–32.

Leung E, Paolacci G, Puntoni S (2018) Man Versus Machine: Resisting Automation in Identity-Based Consumer Behavior. *Journal of Marketing Research* 55(6):818–831.

Li J, Moreno A, Zhang DJ (2019) Agent Pricing in the Sharing Economy: Evidence from Airbnb. Hu M, ed., *Sharing Economy: Making Supply Meet Demand*, 485–503, Springer Series in Supply Chain Management (Springer International Publishing).

Li KJ, Jain S (2015) Behavior-Based Pricing: An Analysis of the Impact of Peer-Induced Fairness. *Management Science* 62(9):2705–2721.

Lin W, Kim SH, Tong J (2021) Does Algorithm Aversion Exist in the Field? An Empirical Analysis of Algorithm Use Determinants in Diabetes Self-Management, Working Paper.

Liu M, Bi W, Chen X, Li G (2014) Dynamic Pricing of Fashion-Like Multiproducts with Customers' Reference Effect and Limited Memory. *Mathematical Problems in Engineering* 2014.

Locke EA, Cartledge N, Koeppel J (1968) Motivational Effects of Knowledge of Results: A Goal-Setting Phenomenon? *Psychological Bulletin* 70(6, Pt.1):474–485.

Locke EA, Latham GP (2006) New Directions in Goal-Setting Theory. *Current Directions in Psychological Science* 15(5):265–268.

Logg JM, Minson JA, Moore DA (2019) Algorithm Appreciation: People Prefer Algorithmic to Human Judgment. *Organizational Behavior and Human Decision Processes* 151:90–103.

Longoni C, Bonezzi A, Morewedge CK (2019) Resistance to Medical Artificial Intelligence. *Journal of Consumer Research* 46(4):629–650.

Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases. *Marketing Science* 38(6):937–947.

Luong A, Kumar N, Lang KR (2020a) Algorithmic Decision-Making: Examining the Interplay of People, Technology, and Organizational Practices through an Economic Experiment, Working Paper.

Luong A, Kumar N, Lang KR (2020b) Human–machine Collaboration and Algorithmic Decision-making in Organizations: Do Humans Still Matter?, Working Paper.

Mak V, Rapoport A, Gisches EJ (2018) Dynamic Pricing Decisions and Seller-Buyer Interactions under Capacity Constraints. *Games* 9(1):10.

Mak V, Rapoport A, Gisches EJ, Han J (2014) Purchasing Scarce Products Under Dynamic Pricing: An Experimental Investigation. *Manufacturing & Service Operations Management* 16(3):425–438.

Mcintosh A, Reys BJ, Reys RE (1992) A Proposed Framework for Examining Basic Number Sense. *For the Learning of Mathematics* 12(3):2–44.

McLaughlin B, Spiess J (2022) Algorithmic Assistance with Recommendation-Dependent Preferences.

Mišić VV, Perakis G (2019) Data Analytics in Operations Management: A Review. *Manufacturing & Service Operations Management* 22(1):158–169.

Nasiry J, Popescu I (2011) Dynamic Pricing with Loss-Averse Consumers and Peak-End Anchoring. *Operations Research* 59(6):1361–1368.

Neate R (2014) Amazon Sellers Hit by Nightmare Before Christmas as Glitch Cuts Prices to 1p. *The Guardian* URL `https://www.theguardian.com/money/2014/dec/14/amazon-glitch-prices-penny-repricerexpress`.

Oh HK, Oliva R (2020) Better Together? How Managers Can Complement Algorithms, Working Paper.

Osadchiy N, Bendoly E (2015) Are Consumers Really Strategic? Implications from an Experimental Study, Working Paper.

Özer Ö, Zheng Y (2012) Behavioral Issues in Pricing Management. *The Oxford Handbook of Pricing Management* (OUP Oxford).

Özer Ö, Zheng Y (2015) Markdown or Everyday Low Price? The Role of Behavioral Motives. *Management Science* 62(2):326–346.

Peters E (2012) Beyond Comprehension: The Role of Numeracy in Judgments and Decisions. *Current Directions in Psychological Science* 21(1):31–35.

Peters E, Västfjäll D, Slovic P, Mertz C, Mazzocco K, Dickert S (2006) Numeracy and Decision Making. *Psychological Science* 17(5):407–413.

Phillips R, Şimşek AS, van Ryzin G (2015) The Effectiveness of Field Price Discretion: Empirical Evidence from Auto Lending. *Management Science* 61(8):1741–1759.

Qiu L, Whinston AB (2017) Pricing Strategies under Behavioral Observational Learning in Social Networks. *Production and Operations Management* 26(7):1249–1267.

Ramachandran K, Tereyağoğlu N, Xia Y (2018) Multidimensional Decision Making in Operations: An Experimental Investigation of Joint Pricing and Quantity Decisions. *Management Science* 64(12):5544–5558.

Rooderkerk RP, DeHoratius N, Musalem A (2021) Retail Analytics: The Quest for Actionable Insights from Big Data on Consumer Behavior and Operational Execution, Working Paper.

Samuelson W, Zeckhauser R (1988) Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty* 1(1):7–59.

Schweitzer ME, Cachon GP (2000) Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence. *Management Science* 46(3):404–420.

Sims CA (2003) Implications of Rational Inattention. *Journal of Monetary Economics* 50(3):665–690.

Snyder C, Keppler S, Leider S (2022) Algorithm Reliance Under Pressure: The Effect of Customer Load on Service Workers.

Staats BR, Dai H, Hofmann D, Milkman KL (2016) Motivating Process Compliance Through Individual Electronic Monitoring: An Empirical Examination of Hand Hygiene in Healthcare. *Management Science* 63(5):1563–1585.

Sun J, Zhang D, Hu H, van Mieghem JA (2019) Predicting Human Discretion to Adjust Algorithmic Prescription: A Large-Scale Field Experiment in Warehouse Operations. *Working paper* .

Talluri KT, van Ryzin GJ (2005) *The Theory and Practice of Revenue Management* (Springer Science & Business Media).

Tan T, Staats BR (2016) Behavioral Drivers of Routing Decisions: Evidence from Restaurant Table Assignment. *Working paper* .

Tiefenbeck V, Goette L, Degen K, Tasic V, Fleisch E, Lalive R, Staake T (2016) Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation. *Management Science* 64(3):1458–1476.

Tversky A, Kahneman D (1974) Judgment Under Uncertainty: Heuristics and Biases. *Science* 185(4157):1124–1131.

van Beuningen B (2018) *Automated Store Ordering Versus Manual Store Ordering at Jumbo*. Master's thesis, Eindhoven University of Technology.

van den Boer A (2015) Dynamic Pricing and Learning: Historical Origins, Current Research, and New Directions. *Surveys in Operations Research and Management Science* 20(1):1–18.

van der Staak BB, Basten RJ, van de Calseyde PP, Demerouti E, de Kok A (2020) Some-Touch Forecasting: A Novel Method to Combine Human Judgment with Statistical Algorithms, Working Paper.

van Donselaar KH, Gaur V, van Woensel T, Broekmeulen RACM, Fransoo JC (2010) Ordering Behavior in Retail Stores and Implications for Automated Replenishment. *Management Science* 56(5):766–784.

Weerasinghe K, Perera HN, Hurley J (2021) Behaviorally Informed Task Sequencing for Efficient Warehouse Operations, Working Paper.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data* (MIT Press), 2nd edition.

Zhang DJ, Dai H, Dong L, Qi F, Zhang N, Liu X, Liu Z, Yang J (2019) The Long-term and Spillover Effects of Price Promotions on Retailing Platforms: Evidence from a Large Randomized Experiment on Alibaba. *Management Science* .