https://pubsonline.informs.org/journal/msom

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 26, No. 6, November–December 2024, pp. 1981–2014
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Incentivizing Healthy Food Choices Using Add-On Bundling: A Field Experiment

Nymisha Bandi,[a] Maxime C. Cohen,[a,*] Saibal Ray[a]

[a] Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 1G5, Canada

*Corresponding author

Contact: nymisha.bandi@mcgill.ca, https://orcid.org/0009-0004-1802-0461 (NB); maxime.cohen@mcgill.ca, https://orcid.org/0000-0002-2474-3875 (MCC); saibal.ray@mcgill.ca (SR)

Received: June 20, 2023
Revised: March 31, 2024; June 25, 2024
Accepted: July 7, 2024
Published Online in Articles in Advance: August 12, 2024

https://doi.org/10.1287/msom.2023.0336

Abstract. ***Problem definition*:** How can retailers incentivize customers to make healthier food choices? Price, convenience, and taste are known to be among the main drivers behind such choices. Unfortunately, healthier food options are often expensive and not adequately promoted. However, we are observing recent efforts to nudge customers toward healthier food. ***Methodology/results*:** In this paper, we conducted a field experiment with a global convenience store chain to better understand how different add-on bundle promotions influence healthy food choices. We considered three types of add-on bundles sequentially: (i) an unhealthy bundle (when customers purchased a coffee, they could add a pastry for $1), (ii) a healthy bundle (offering a healthy snack, such as fruit, vegetable, or protein, as a coffee add-on for $1), and (iii) a choice bundle (the option of either a pastry or a healthy snack as an add-on to coffee for $1). In addition to our field experiment, we conducted an online laboratory study to strengthen the validity of our results. ***Managerial implications*:** We found that offering healthy snacks as part of an add-on bundle significantly increased healthy purchases (and decreased unhealthy purchases). Surprisingly, this finding continued to hold for the choice bundle, that is, even when unhealthy snacks were concurrently on promotion. However, we did not observe a long-term stickiness effect, meaning that customers returned to their original (unhealthy) purchase patterns once the healthy or choice bundle was discontinued. Finally, we show that offering an add-on choice bundle is also beneficial for retailers, who can earn higher revenue and profit.

Funding: This research was supported by the James McGill Scholar Award Fund, the Scale AI Chair Program, IIVADO (Institut de valorisation des données) Fundamental Research Project Grant, and two Discovery Grants from the Natural Sciences and Engineering Research Council of Canada.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2023.0336.

Keywords: healthy eating • field experiment • economic incentives • bundling

1. Introduction

Food habits have changed considerably in the last few decades, with a shift toward high-calorie and high-sugar dishes, frequent eating out, and larger food portions along with a reduced intake of fruits, vegetables, and high-fiber items.[1] Diets with higher amounts of salt, sugar, and trans fats and lower amounts of fruits, vegetables, and fibers are typically categorized as unhealthy (Lobstein and Davies 2009). One of the main consequences of this diet change is a higher incidence of obesity and chronic noncommunicable diseases, such as diabetes, heart disease, and stroke (Muhammad et al. 2017). For example, the global prevalence of diabetes nearly doubled from 4.7% in 1980 to 8.5% in 2014 in the adult population (Roglic 2016).

Given that an unhealthy diet has clear adverse health consequences, there have been concerted efforts in recent years to encourage healthy eating through various interventions or nudges (Cadario and Chandon 2020, Hinnosaar 2023). Researchers have explored the effectiveness of interventions, such as descriptive nutrition labeling (Nikolova and Inman 2015), visibility enhancement (Kroese et al. 2016), increased assortment and availability of healthy items (van Kleef et al. 2012), healthy eating calls (Salmon et al. 2015), and price promotions for healthy food (Afshin et al. 2017). Research suggests that price, convenience, and taste are the main drivers behind consumers' choices, whereas listing nutritional benefits and dietary guidelines have little effect (Sogari et al. 2018). It is perhaps unsurprising that price promotions are one of the most popular strategies used by food retailers (see, e.g., Neslin 2002, Cohen et al. 2021). Several studies have delved into assessing the impact of promotions on consumer behavior. Hawkes (2009) surveyed the literature on using promotions as an intervention to affect food choices, especially related

Figure 1. (Color online) Examples of Add-On Bundles

to healthy eating, and observed that some interventions successfully influenced consumers' choices. An (2013) used field experiments to investigate the efficacy of subsidies in encouraging healthy food purchases.

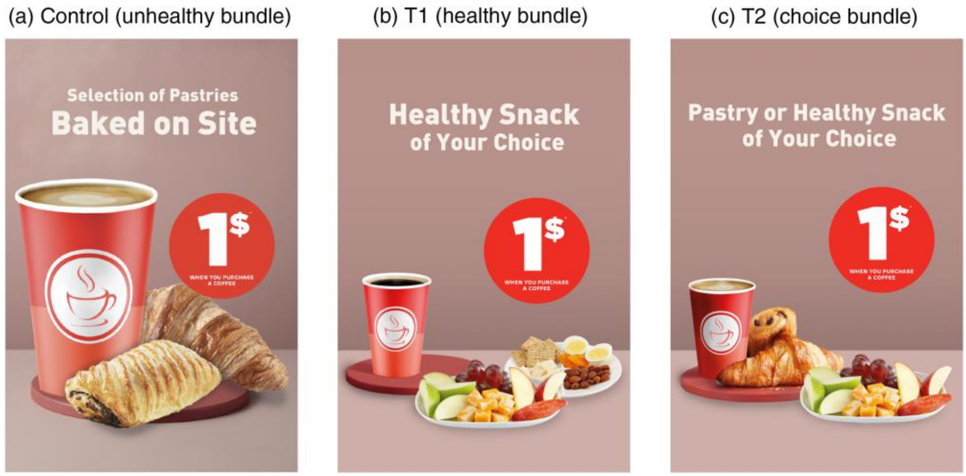
The above studies focus on price discounting on individual products as the focal promotion. However, there is a lack of research examining the potential of bundling as an intervention strategy;[2] although, according to the United Kingdom's Competition Commission (2000), bundle offers typically increase sales more than simple price discounts. The only study in this context is by Carroll et al. (2018), who performed a laboratory study in which participants shopped via a grocery display while inducing a cognitive load (i.e., mental strain). The authors found that discounted bundles could successfully encourage consumers toward a higher fruit and vegetable consumption in the absence of a cognitive load, that is, when consumers were more attentive to what they were buying. Anecdotal evidence also suggests that healthy food bundles can be an effective incentive (Gordon and ICF International 2014). For example, retailers can incentivize customers by bundling a healthy product with a popular item. To further motivate the use of bundling as a mechanism to incentivize healthy food choices, we ran a laboratory experiment to compare bundling with price discounting. We found a significantly higher likelihood of purchasing healthy products and a lower likelihood of purchasing unhealthy products when healthy options were promoted as part of a bundle (see Appendix A, Section A.1, for more details). Bundling also allows retailers to strategically counterbalance any profit loss due to the discounted bundle offer by including a product with a high-profit margin in the bundle. Given the popularity of bundling in retail, there is a need to specifically study the impact of bundling on healthy food purchases, which is the focus of this paper.

In this paper, we focus on add-on bundling, which refers to the retail practice of offering a product (say, B) at a discounted price when consumers purchase another product (say, A) at the regular price, with A often being a popular, high-margin product. Examples of add-on bundles are shown in Figure 1. Many add-on bundles combine an unhealthy item with a popular one. For example, Tim Hortons, a Canadian fast-food chain, offers consumers the option to add a sausage biscuit for 99¢ (originally priced at $3.29) when they buy a coffee. Another such promotion is offered by 7-Eleven, where consumers can get a 32 oz Big Gulp for 99¢ (originally priced at $1.79) when they buy nachos.

Most previous research on incentizing healthy food purchases has focused on supermarkets and grocery stores. Although these are popular places to buy food items, customers also buy shop for groceries in small stores and other limited-service establishments, such as gas stations, dollar stores, and pharmacies (Ver Ploeg et al. 2015). These establishments are often coined as convenience stores (C-stores). Half of all shoppers visit convenience stores at least once a week. For younger generations, particularly those belonging to Generation Z and millennials, frequenting their preferred convenience store is a pivotal aspect of their daily routine, with 43% engaging in daily shopping at independent C-stores (see Warren 2023). More importantly, C-stores are known for their large selection of unhealthy food items (Farley et al. 2009). Studies have shown that the ratio of healthy to unhealthy food available and purchased in C-stores is lower relative to supermarkets (Larson et al. 2009, Stern et al. 2016). Moreover, Bennett et al. (2020) found that the prevalence of promotions for products with high fat, sugar, and salt was also quite high in C-stores (56%).

Our main research goal is to provide strong experimental evidence on the impact of offering add-on bundles for healthy and unhealthy food options in C-stores. Consumer behavior is nuanced and context dependent, often varying based on the product category and the type of product. In this paper, we focus on healthy and unhealthy characteristics of food products in the C-store

Figure 2. (Color online) Promotion Banners Used



context. Specifically, we investigate whether add-on bundles can successfully incentivize customers toward healthier food choices, especially when there may still be concurrent unhealthy add-on bundles available. In addition, we examine the impact of such offers on the retailer's revenue and profit.

To address the above questions, we conducted a field experiment in a branch store of a leading global C-store chain in a major North American city. Like many C-stores, this store carries a disproportionately low volume of healthy products,[3] and healthy products are promoted much less often (around 28%). In this paper, we use three common nutrient profiling techniques (the calories-for-nutrient (CFN) score, ratio of recommended to restricted (RRR) food score, and Food Standard Agency (FSA) rating) to classify food as healthy or unhealthy (for more details, see Section 3.1 and Appendix B). The primary unhealthy foods are bakery items, chocolates, salty snacks, and sugary beverages, accounting for 68% of all sales. One of the most popular unhealthy promotional add-on bundles offered by the store is the option to add a bakery item for $1 (average price of $2.36) when purchasing a hot beverage, such as coffee, tea, or hot chocolate (see Figure 2(a) for an illustration of the promotion). This successful promotion is available in all the stores in the city and has been in place for more than three years. We highlight that 30% of all coffee transactions avail of this promotion and that it accounts for 70% of the sale of bakery items in the store. We consider this setting the status quo (called Control 1 (C1)) for our experiment. We then tested two different add-on bundles. First, in Treatment 1 (T1), we replaced the unhealthy item in the bundle with a healthy alternative. Specifically, customers were offered the option to add a healthy snack for $1 (average price of $3.99) when purchasing a coffee (see Figure 2(b)).[4] We call this a healthy add-on bundle. Subsequently, in Treatment 2 (T2), customers were offered the choice of an add-on bundle that involves either adding a healthy snack or an unhealthy bakery item for $1 when purchasing a coffee (see Figure 2(c)). We call this a choice add-on bundle. Last, we returned to the status quo by offering the original unhealthy add-on bundle in Control 2 (C2). We use the same add-on price of $1 irrespective of the original price of the add-on product to ensure a fair and standardized comparison of the different promotions offered. Each of the above interventions was executed for three consecutive weeks. We used several other control stores in the same city to account for unobserved time heterogeneity. To further support and validate the results of our field experiment, we also ran a laboratory experiment using an online survey that mimicked our field experiment without suffering from the temporal split among the different interventions.

Our main empirical methods relied on several regression specifications, the difference-in-differences (DID) approach, and a relatively new methodology known as synthetic difference-in-differences (SDID). Our control conditions in both methods leveraged the data from all the other stores in the same city by identifying comparable stores. We also used several control variables, such as time fixed effects, to control for city-wide unobserved temporal trends or shocks (e.g., seasonality) and product stockouts. Finally, we conducted a series of robustness tests to showcase the stability of our results. The fact that we found consistent results both in our field experiment (under several model specifications) and in our online survey experiment enhances our confidence in the validity of our results. To address concerns on the one-store study and showcase the generalizability of our results, we conducted a series of A/A tests and DID analyses.

1.1. Summary of Results

As discussed, our field experiment compared three add-on bundles: Controls 1 and 2 (unhealthy bundles), Treatment 1 (healthy bundle), and Treatment 2 (choice bundle). Our results are summarized below.

1.1.1. Analyzing the Impact of Add-On Bundling on Food Choices

Healthy Add-On Bundling. When comparing Control 1 and Treatment 1, we found that replacing the unhealthy bundle with the healthy one resulted in a significant number of customers substituting unhealthy bakery items with healthy snacks when buying a coffee. Sales of healthy snacks increased by 1,107.69%, whereas sales of unhealthy snacks decreased by 36.52%.

Choice Add-On Bundling. More importantly, under Treatment 2 (i.e., the choice bundle), many customers persisted with the healthy option as an add-on instead of opting for the discounted bakery item. Specifically, whereas sales of the healthy add-on bundle (i.e., coffee plus healthy snack) were naturally lower in Treatment 2 than in Treatment 1, overall sales of healthy snacks under Treatment 2 were significantly higher than in Control 1. Sales of the healthier alternative increased by 817.5% relative to Control 1 and decreased by 31.63% relative to Treatment 1. Also, sales of the unhealthy add-on bundle (i.e., coffee plus bakery item) were higher in Treatment 2 than in Treatment 1 but remained almost the same as in Control 1. Last, sales of both bakery items and healthy snacks were similar in Control 1 and Control 2.

The main takeaway from our experiment is that a healthy add-on bundle can incentivize customers toward healthier purchases even in the presence of a concurrent unhealthy bundle. We established the robustness of these results by considering a multitude of models and settings.

1.1.2. Revenue and Profit Analysis. Our experiment also revealed interesting insights into the retailer's profit implications of health interventions. Retailers may not be keen to promote healthy items if doing so might result in a loss. Our results suggest that it is possible to achieve a win–win situation for both customers (who will be more likely to choose a healthy food option) and the retailer (who will earn a higher revenue and profit). Specifically, when comparing Treatment 1 to Control 1, the extra profit earned from the bakery items sold at full price compensated for the loss from the discount offered on healthy snacks, hence maintaining a similar profit level. When comparing Treatment 2 to Control 1 (Control 2), however, we observed a profit increase of 23.93% (28.54%). We also found a 27.41% profit increase from Treatment 1 to Treatment 2. This result is somewhat counterintuitive because the retailer was offering a discount for both healthy and unhealthy items in Treatment 2. Nevertheless, by offering an add-on choice bundle, the total sales of bundles increased significantly relative to when only one type of bundle was offered. Because the common product in the two bundles was a high-margin product (coffee), the loss incurred due to discounts was offset by the additional margin accrued from the coffee purchases. In conclusion, when the add-on bundle is carefully designed, an outcome that is both profitable for the retailer and encourages healthy food choices for consumers can be achieved.

The rest of this paper is organized as follows. We develop our hypotheses in Section 2 followed by the design of our field experiment in Section 3. We present the results of our experiment in Section 4, and we conduct a series of robustness tests in Section 5. Section 6 discusses the managerial insights from our field experiment from the perspective of the retailer and the consumers. Finally, we conclude in Section 7. Several additional analyses and results are relegated to the appendices.

2. Hypotheses Development

In this section, we develop hypotheses to study how consumer choices regarding healthy and unhealthy items are affected when exposed to three different promotional bundles (healthy, unhealthy, and choice). This paper focuses on mixed bundling, where firms offer both bundled and individual products. The attractiveness of a bundle naturally depends on the products included in the bundle and on whether the purchase is driven by hedonic or utilitarian considerations (Khan and Dhar 2010). Hedonic goods (e.g., designer clothes, luxury watches, and unhealthy food) provide more fun, pleasure, and excitement, whereas utilitarian goods (e.g., microwaves, personal computers, and healthy food) are instrumental and functional (see Hirschman and Holbrook 1982). Consumers' purchase decisions depend on their reservation price for the products in the bundle, which is equal to the sum of the conditional reservation prices of the separate products (Stremersch and Tellis 2002).[5] Moreover, Wang (2017) found that the hedonic perception of products has a stronger influence on customers' purchase decisions relative to the utilitarian perception. Our hypotheses are designed to study this purchase behavior under various incentives.

We next present our first hypothesis on the impact of the healthy bundle on the sales of both healthy and unhealthy snacks.

Hypothesis 1. *Offering a healthy bundle with a healthy snack as an add-on to a popular item—instead of an unhealthy bundle—will have the following effects:*

a. *increase sales of healthy snacks and*
b. *decrease sales of unhealthy snacks.*

We focused on two product categories for the above hypothesis—healthy and unhealthy snacks (more details can be found in Section 3.1)—whereas the popular, common item was coffee. Stremersch and Tellis (2002) showed that a mixed bundle increases sales of the constituent products in the bundle relative to the unbundled scenario. After all, this is the main motivation behind using a bundling strategy. Bundling can be viewed as a price discrimination technique. By properly setting the bundle price, the retailer can capture different customer segments with heterogeneous valuations for the individual products in the bundle. In this case, by replacing the unhealthy snacks in the bundle with healthy ones, we can expect an increase in sales of healthy items (Hypothesis 1a). Recall that the second product category (unhealthy snacks) was part of the bundle prior to our intervention. Removing the unhealthy item from the bundle will decrease the sales of unhealthy snacks because consumers' reservation price for unhealthy snacks is lower than the original price (Guiltinan 1987). Consumer choice research states that hedonic items are associated with greater guilt and, thus, require greater justification. Hence, a bundle promotion with a hedonic item (in our case, an unhealthy snack) should be more effective at increasing the purchases of the bundled items than the unbundled items (Khan and Dhar 2010). We thus expect a decrease in sales of unhealthy snacks when the unhealthy bundle is not offered (Hypothesis 1b).

Our second hypothesis is on the impact of the choice bundle on the sales of both healthy and unhealthy snacks.

Hypothesis 2. *Offering a choice bundle—which includes both a healthy and an unhealthy snack option as an add-on to a popular item—will have the following effects*:

a. *decrease the sales of healthy snacks relative to the unhealthy bundle setting and*

b. *increase the sales of unhealthy snacks relative to the unhealthy bundle setting*.

In the literature, we could find two contrasting theories regarding the above hypothesis. We outline both perspectives below, and we ultimately adopt the one with the strongest empirical support.

The choice bundle lets the consumers choose between a healthy and an unhealthy snack as the add-on item based on their inherent preferences, both emotional and cognitive. This choice could be seen as being between consumption for immediate pleasure and consumption for long-term benefits and well-being. It is well known that consumers tend to assign disproportionate weight to short-term benefits and costs (Ainslie 1975). For example, when contemplating a future meal, one may plan to consume healthy options, but when consumption is imminent, one is more likely to prioritize immediate appeal and temptation and opt for unhealthy options. This is driven by temporally inconsistent preferences. This type of choice is also related to the conflict between desire and willpower. People often choose the short-term, easy, gratifying option (Shiv and Fedorikhin 2004). The few who can resist this impulse are the ones who make decisions based on a rigorous assessment of the long-term repercussions behind these choices. Researchers have found that emotions—rather than logic—tend to have a greater impact on choice (Khan et al. 2005).

In another study by Wilcox et al. (2009), the authors explored the impact of having healthy food options in a consideration set on consumers' decisions. Surprisingly, they showed that consumers were more likely to select indulgent foods when a healthy option was available compared with when it was not. This effect was stronger for individuals with a higher level of self-control. This finding supports the idea that the presence of a healthy option might make individuals feel that they have the permission to indulge. Similarly, Martin (2007) suggested that adding healthy items to menus could make unhealthy items seem less threatening and, thus, harder to resist. Using this argument, the hypothesis can be framed as providing a choice to consumers (between healthy and unhealthy food choices) will increase the likelihood of selecting unhealthy snacks and reduce the likelihood of choosing healthy snacks. This forms the basis for the first perspective.

In contrast, from an economics perspective, when given a choice, consumers typically select the option that offers them the highest perceived monetary gain, which, in this case, is the healthy snacks (Janiszewski and Cunha 2004). However, the actual consumer decision-making process is far more complex. It is not solely driven by rationally assessing the economic benefits but it is also heavily influenced by several emotional factors. Furthermore, Fishbach et al. (2003) found that facing a temptation such as unhealthy food activated goals to resist those temptations, helping individuals to act in line with their long-term interests such as eating healthier. In this case, when individuals have a choice between an unhealthy and a healthier option, they activate their health goals, leading to self-regulation and the tendency to avoid indulgent choices. This is based on the assumption that repeated efforts at self-control create links between thoughts of temptations and conflicting goals. This suggests the completely opposite hypothesis that providing a choice to consumers will decrease the likelihood of selecting unhealthy snacks and increase the likelihood of selecting healthy snacks. This forms the second perspective. Because we found more support for the first perspective, we proceeded with formalizing this version of the hypothesis. In this paper, we have the opportunity to leverage our field experiment to distinguish between these two competing arguments.

Hypothesis 3. *Offering a healthy bundle or a choice bundle—instead of an unhealthy one—will not have any effect on sales of unhealthy snacks purchased outside the bundle*.

Each individual develops a reference price for products based on historical prices and other context variables about the product. Consumer purchase behavior is influenced explicitly or implicitly by this reference price (Putler 1992). For each product, there are some individuals who are willing to pay the full price without leveraging any bundled promotions because they have a reference price that is equal to or higher than the price of the product. In the context of our field experiment, the customers who purchased unhealthy snacks at the full price (even when there was an offer to leverage the coffee plus unhealthy snack bundle) fell into that category. Such individuals are not likely to alter their purchase behavior when the bundle is modified to include healthy snacks. Alternatively, we can consider the main motivations behind the bundling strategy, which include market segmentation, new product introduction, and cross-selling (Stremersch and Tellis 2002). At the same time, not every customer will be influenced by this strategy. In the unhealthy bundle in our experiment, the focal product was coffee, and the unhealthy snack was the discounted add-on product. There will naturally be some customers who are interested in only the unhealthy snack, and they will not be influenced by the bundle promotion. These consumers are loyal to the product (in our case, unhealthy snacks) irrespective of the promotions bundled with coffee (because they are most likely not interested in purchasing a coffee). Consequently, we hypothesize that sales of unhealthy snacks purchased outside the bundle will not be affected throughout.

Hypothesis 4. *Our bundling strategies do not have a long-term stickiness effect on sales of healthy and unhealthy items*.

There is a limited understanding on how consumers behave once promotional offers are discontinued. Several researchers have observed in various settings that incentives may only have a short-term impact. For example, in a field experiment to conserve energy, Allcott and Rogers (2014) found that the incentives had no long-term effects on energy consumption once they were discontinued. They observed only short-term effects immediately after the experiment. Similarly, Ni Mhurchu et al. (2010) studied the long-term impact of promoting healthy items and found that there was no significant effect on sales of healthy and unhealthy items once the promotion was discontinued. Motivated by them, we hypothesize that our interventions will not have a long-lasting effect. Hence, we do not expect to observe a stickiness effect in the increased sales of healthy items.

3. Field Experiment

To formally test our hypotheses, we conducted a live field experiment in a physical C-store located in the city center of a North American metropolitan city. We then ran a laboratory study based on an online survey to further showcase the validity and robustness of our findings. In this section, we present the experimental design of our field experiment.

3.1. Design

One of the key steps in our experimental design was to determine the healthy and unhealthy categorization of the offered products. In this paper, we rely on the common convention used by most consumers to define healthy and unhealthy products, while providing support from the nutrition literature. Specifically, products that include fresh fruits or vegetables with low fats, low sugar, and low carbohydrates while having high fiber and other necessary nutrients are considered healthy. To complement this healthy versus unhealthy categorization, we use three common nutrient profiling methods: the RRR food score (Scheidt and Daniel 2004), the CFN score (Lachance and Fisher 1986), and the score developed by the Food Standard Agency.[6] The scores computed using the above three methods are presented in Table 1. More details on this topic can be found in Appendix B. The products with the highest score in all categories were picked as the healthy snacks, which are assortments of healthy items combined and sold as a snack box. More specifically, we considered three types of healthy snacks, namely, fruits, vegetables, and protein (Figure 3, (a)–(c)). The product descriptions can be found in Table 1. Similarly, all pastry items, namely, croissants, cinnamon rolls,

Table 1. Nutrient Profiling Scores and Contents for Healthy and Unhealthy Snacks Used in Our Experiment

Snack type	Contents	CFN	RRR	FSA	Product images
Vegetable box	Celery, broccoli, carrots, pepper, and a dip	452.38	0.23	−8	Figure 3(b)
Fruit box	Apple slices, grapes, and cheese	184.74	0.92	−7	Figure 3(a)
Protein box	Hard boiled egg, almonds, cheese, and crackers	206.24	0.32	−1	Figure 3(c)
Pastry	Croissant, cinnamon roll, apple turnover, fruit Danish chocolate avalanche, chocolate muffin	722.88	0.1	12	Figure 3(d)

Figure 3. (Color online) Healthy and Unhealthy Snacks Used in the Bundle Promotions

apple turnovers, fruit Danishes, chocolate avalanches, and chocolate muffins (Figure 3(d)), are unanimously classified as unhealthy by all nutrient profiling methods (as well as based on common sense).

As discussed, the C-store chain has been running a successful promotion campaign involving an add-on bundle offer. The promotion went as follows: When customers purchased a coffee (a.k.a., hot beverage), they could add a pastry for an additional $1. This was the control condition for our experiment. The banner used in the store to promote this bundle offer can be seen in Figure 2(a). We call this the unhealthy bundle. Our field experiment involved two alternative interventions to test our hypotheses. The first intervention was to study the effect of replacing the unhealthy snack (pastries) in the bundle with a healthy alternative (snack boxes), resulting in Treatment 1. We call this the healthy bundle. The banner used to promote this intervention can be found in Figure 2(b). During this intervention, customers could only add a healthy snack (and not unhealthy pastries) for an additional $1 when they purchased a coffee. The second intervention was to examine the preference between a healthy and an unhealthy snack when they were offered simultaneously as part of separate bundles and the choice was left to the customers. We call this the choice bundle. The promotion offered during Treatment 2 is shown in Figure 2(c). In other words, when customers purchased a coffee, they could add either a healthy or an unhealthy snack for an additional $1. As explained before, we strategically designed these three promotion bundles around coffee purchases because it was by far the most sold item in the store (around 45% of purchases). Per our retail partner's suggestion, we opted to use a uniform promotional value of $1 for all bundles in order to keep the deals simple and equally appealing.

As discussed in Section 4.1, the primary outcome variable in our analyses is the daily number of transactions in each category (healthy and unhealthy products). We then examine the impact of the healthy bundle (T1) and the choice bundle (T2) relative to the unhealthy bundle (C1). We also study the postexperiment effect when the promotion reverts to the unhealthy bundle (C2). Note that the C2 condition checks whether the effects observed in our experiment are not merely attributable to an awareness increase of the healthy products but are causally linked to our intervention. When comparing T1 and C1, we observe the effect of replacing an unhealthy item by a healthy one on the sales of coffee bundles and individual items (healthy and unhealthy). When comparing T2 and C1, we observe the effect of providing consumers with a choice between a healthy and an unhealthy item while keeping the same deal value of adding $1 extra. This allows us to measure the affinity toward healthy items under the same amount spent. Finally, the comparison between C2 and C1 measures the (longer-term) residual effect of the experiment after the interventions are discontinued, if any.

We next discuss the implementation timeline of our field experiment. Because the experiment was conducted at a specific time of the year and each treatment was at a different period, we also conducted a complementary laboratory study (based on an online survey, presented in Section 5.4) to verify that our results were not influenced by unobserved temporal heterogeneity.

3.2. Implementation Timeline

The experiment lasted for a total of 14 weeks, as shown in Table 2, along with the promotions offered. Each intervention was in place for a period of three consecutive weeks. To ensure uniform conditions throughout all phases, we excluded the data collected for two consecutive weeks between February 21, 2022, and March 6, 2022. During this period, the store faced technical issues while switching from the unhealthy bundle to the healthy one, and there were inventory shortages for the healthy snacks (also, the second week had very few transactions because it coincided with a vacation period). During the experiment, the promotions shown in Figure 2 were displayed near the store entrance for visibility and awareness purposes. There were no other promotions on the same products to alleviate interference effects. The store employees were informed of the promotions to ensure they could answer any customer questions about the products or on the nature of the promotions.

The timeline for the different interventions in the experiment can be summarized as follows:

- The first three-week period was considered as the baseline phase in which the (usual) unhealthy bundle was offered (Figure 2(a)). We call this phase C1.
- The next two weeks were excluded (E) because of technical issues beyond our scope.
- During the next three weeks, the healthy bundle was offered (Figure 2(b)). We ensured that no other promotions were offered for the unhealthy products included in the experiment and that everything else remained the same for these two product categories. We call this phase T1.
- During the next three weeks, the choice bundle was offered (Figure 2(c)). We call this phase T2.
- Finally, after the T2 phase, the promotion bundle reverted to the default unhealthy bundle offered in C1. Although this promotion continued throughout the rest of the year, we considered only the first three weeks as the C2 phase.

Our goal was to rigorously analyze the customers' purchase patterns of healthy and unhealthy snacks under each of the four conditions (C1, T1, T2, and C2). We investigated the impact of the different treatments (T1 and T2) as well as C2 relative to the control (C1) by running four types of empirical analyses: (a) treatment effect using an ordinary least squares regression specification (Section 4.2), (b) a negative binomial regression (Section 5.1), (c) several DID approaches (Section 5.2), and (d) the SDID (Section 5.3). The regression and negative binomial regression analyses established the effect of the treatment by considering only the sales from the treated store, whereas the DID and SDID analyses relied on variation in the time series by analyzing the trend changes using 88 other stores from the same chain in the same metropolitan city. We also studied the impact of the time of the day on the consumer behavior using heterogeneous treatment effects (Section 4.3). Finally, we conducted a detailed analysis using market segmentation, a series of A/A tests, and several DID variants to showcase that our results are generalizable (Section 5.5). Because the treated store was located in the city center, we saw a clear drop in sales during weekends. This was due to offices being closed on weekends and general footfall being significantly lower. We thus conducted our analyses by using only the weekday sales (that said, the vast majority of our results continue to hold when we include the weekends, as shown in Appendix D, Section D.2).

4. Data and Results

In this section, we present the data collected and report our results. Our main econometric methods are DID and SDID. Nevertheless, we also use analysis of variance (ANOVA) and regression analyses to showcase the robustness of our estimates. Additionally, we investigate the presence of heterogeneous treatment effects by considering the time of day as a covariate.

4.1. Data and Metrics

In this section, we provide an overview of the data collected in the treated store during our field experiment. There were two types of data: point-of-sale (POS) data and end-of-day (EOD) inventory data. The POS data provide us with detailed information on all the transactions.

Table 2. Live Experiment: Dates and Promotions Offered

Experiment phase	Date range	Number of days	Promotion offered	Promotion banner
Control 1	1/31/22 to 2/20/22	21	Unhealthy add-on bundle	Figure 2(a)
Excluded	2/21/22 to 3/6/22	14	—	—
Treatment 1	3/7/22 to 3/27/22	21	Healthy add-on bundle	Figure 2(b)
Treatment 2	3/28/22 to 4/17/22	21	Choice add-on bundle	Figure 2(c)
Control 2	4/18/22 to 5/8/22	21	Unhealthy add-on bundle	Figure 2(a)

For each transaction, we had access to several features, such as transaction time, the total amount spent, the total discount amount, discount details, payment method, and individual items purchased along with quantities and prices. Similarly, the EOD inventory data recorded the inventory level at the end of the day (i.e., midnight) for each item in the store. Unfortunately, this number was often not accurately recorded because it was computed internally in the system based on approximation rules. A physical inventory count would typically result in more reliable ending inventory numbers. However, for large organizations, this is obviously unfeasible. To mitigate this inaccuracy in inventory records and accurately identify when specific products were out of stock, we used both the sales data on a given day and the value of EOD inventory in the system. Specifically, if there were no sales recorded for a particular item on a given day and the EOD inventory was not positive, we could safely conclude that the item was not available on that day. We stored this information as a binary variable called *Stockouts* and used this as a control variable in our empirical models.

In all our analyses, we used data aggregation at the daily level. The values of the average daily transactions for each phase were as follows: 545.71 (standard deviation (SD) = 206.14) for C1, 576.95 (SD = 237.82) for T1, 689.65 (SD = 179.03) for T2, and 542.15 (SD = 212.41) for C2. The majority of the transactions were coffee purchases, with a daily average of 267.0 (SD = 101.61) for C1, 258.81 (SD = 107.56) for T1, 326.0 (SD = 93.93) for T2, and 220.45 (SD = 113.97) for C2. Coffee transactions accounted for 46% of the overall store transactions, and 34% of those transactions included one of the bundle promotions used in our experiment.

4.1.1. Data Filtering. To ensure that our results are representative, we carefully applied basic filtering rules. We eliminated the top 1% of observations based on the distribution of each key metric. For example, to analyze the total sales during the different phases of the experiment, we first looked at all the transactions and eliminated the top 1% that had unusually large basket sizes. Similarly, we eliminated the transactions with the top 1% highest sales amounts. Finally, we considered the total transactions recorded on each day during the experiment and eliminated the one day with the highest value (i.e., the top 1%, assuming that this value was exceedingly high). To showcase the robustness of our results, we varied this filtering threshold between 1% and 3%. We also considered outlier removal using three standard deviations away from the mean. We observed consistent results under each of these outlier removal approaches. We highlight that the results remain consistent without the application of these data filtering methods (see Appendix D, Section D.1). However, the estimated effects appear to be more extreme when outliers are not omitted. We thus proceed with presenting the results after removing the outliers to provide a more representative estimation of the effects.

4.1.2. Key Metrics. We used the following two metrics to capture customer preferences toward healthy and unhealthy food choices:

1. *Number of add-on bundles sold.* This is the total number of add-on bundles purchased as well as the number of healthy and unhealthy add-on bundles purchased during the experiment period.
2. *Number of transactions that included specific types of items.* This is the total number of transactions in which either a healthy or an unhealthy snack was purchased.

We then aggregated these metrics at the day level to guide our empirical analyses and estimate the various treatment effects.

4.2. Regression Results

We examined the impact of the different treatments—T1 (healthy bundle), T2 (choice bundle), and C2 (unhealthy reverted bundle)—on the average daily number of bundles (all, healthy, and unhealthy) purchased and the average daily transactions containing items related to our experiment (healthy and unhealthy snacks). As mentioned, each treatment lasted for three consecutive weeks (i.e., 21 days). As discussed, we focused on the data from weekdays to reflect a more representative picture (nevertheless, our results remained consistent when including weekend observations) and removed the day with the largest number of transactions (outlier). We thus had a sample of 59 days. We let Y_{ip} denote the values of the two metrics (bundles sold and quantity sold) on day i for different product groups p. The number of bundles sold is analyzed for bundle group $p = \{$all, healthy, unhealthy$\}$, whereas the quantity sold is analyzed for product group $p = \{$healthy snacks, unhealthy snacks$\}$. We used the following regression model to estimate the treatment effects:

$$Y_{ip} = \alpha + \sum_{I \in \{T_1, T_2, C_2\}} \beta_I \cdot I + \beta_{so} \cdot SO_{ip} + \mu_i + \epsilon_{ip}, \qquad (1)$$

where I represents the different treatment T_1, T_2, C_2 for day i (see Table 2), so that β_I is the coefficient for the treatments $I \in \{T_1, T_2, C_2\}$. Term T_1 indicates being exposed to Treatment 1, T_2 indicates being exposed to Treatment 2, C_2 indicates being exposed to Control 2, and SO_{ip} is a binary variable that indicates whether a stockout has occurred on day i for product p. Note that we performed a mediation analysis to ensure that the stockout variable can be used in the above regression (see Appendix C). The term μ_i represents time fixed effects to capture any unobserved time-specific demand shocks. We considered various types of time fixed effects, such as day-of-week effects and the week number during the treatment period. The key parameters in

Table 3. Impact of Different Interventions Using Weekday Sales on Bundles Sold

Panel A. Healthy and unhealthy bundles

	Healthy add-on bundle				Unhealthy add-on bundle			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	54.07***	54.07***	54.32***	54.90***	−54.87***	−54.87***	−54.74***	−55.65***
	(3.73)	(3.73)	(4.01)	(4.03)	(9.51)	(8.74)	(9.74)	(8.97)
T2	38.65***	38.80***	38.84***	39.41***	−10.54	−9.95	−10.48	−10.24
	(3.79)	(3.80)	(3.96)	(3.97)	(9.68)	(8.91)	(9.80)	(9.00)
C2	2.53	2.53	2.79	3.37	−10.53	−10.53	−10.53	−10.53
	(3.73)	(3.73)	(4.01)	(4.03)	(9.51)	(8.74)	(9.60)	(8.81)
No. observations	59	59	59	59	59	59	59	59
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.85	0.86	0.85	0.86	0.43	0.57	0.43	0.57

Panel B. All coffee add-on bundles

Treatment	Model (1)	Model (2)	Model (3)	Model (4)
T1	−3.67	−3.67	−3.67	−3.67
	(10.44)	(9.33)	(10.44)	(9.33)
T2	25.67*	26.40***	25.67*	26.40***
	(10.63)	(9.51)	(10.63)	(9.51)
C2	−8.27	−8.27	−8.27	−8.27
	(10.44)	(9.33)	(10.44)	(9.33)
No. observations	59	59	59	59
Time FEs	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes
R^2	0.18	0.39	0.18	0.39

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^*p < 0.05$; $^{***}p < 0.001$.

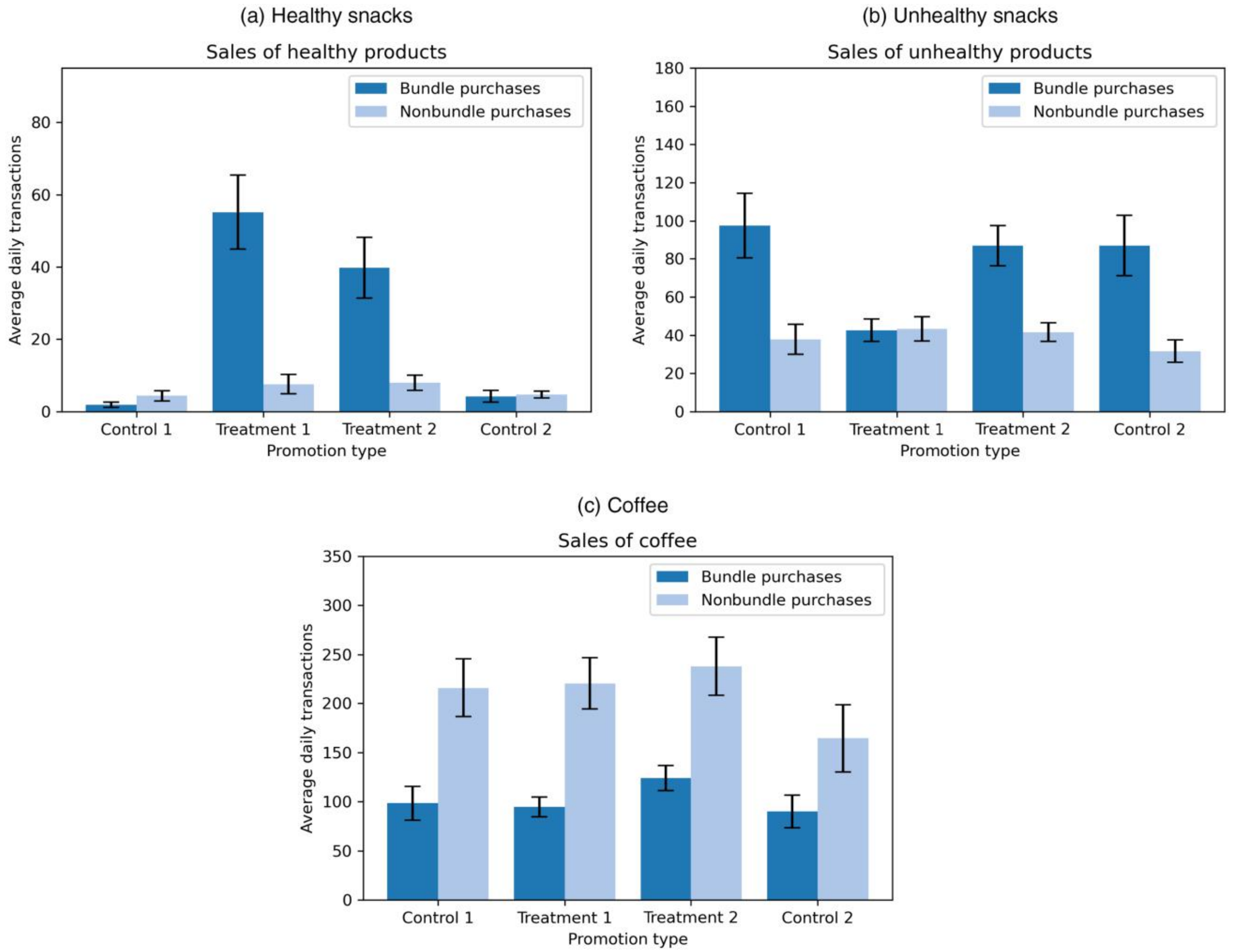
Equation (1) are $\beta_{T_1}, \beta_{T_2}, \beta_{C_2}$, which capture the impact of each type of promotion bundle on the customers' snack preferences.

4.2.1. Bundles Sold. We estimated Equation (1) for healthy, unhealthy, and total add-on bundles sold. Specifically, we considered four different models for each case. Model (1) reports the treatment effect estimates ($\beta_{T_1}, \beta_{T_2}, \beta_{C_2}$) without controlling for stockouts and without including time fixed effects. Models (2)–(4) explicitly account for stockouts and time fixed effects (all possible combinations). As mentioned, we considered three types of time fixed effects: (i) day-of-week effects, (ii) week number during the treatment period, and (iii) a combination of both. We report only the results for day-of-week time fixed effects, but we found consistent results in all three cases. Table 3 shows that the results are consistent across all model specifications (i.e., with and without time fixed effects and with and without controlling for stockouts) for healthy, unhealthy, and all add-on bundles.

For illustration purposes, Figure 4 plots the sales of coffee add-on bundles and the healthy and unhealthy add-on bundles for each treatment along with a 95% confidence interval.[7] Figure 4(c) suggests that T2 increased the sales of coffee add-on bundles by 26.08%. This indicates a strong, statistically significant positive effect on the sales of healthy snacks during T2 (Table 3, panel B). Recall that these add-on bundles consist of either the healthy or the unhealthy bundle. Figure 4(a) suggests that T1 (respectively, T2) increased the sales of the healthy bundles by 4,784.96% (respectively, 3,420.35%). The high percentages are because of low purchases observed in the control period. In other words, we observed a strong statistically significant positive effect on the sales of healthy snacks during both T1 and T2 (Table 3, panel A). Finally, Figure 4(b) suggests that T1 decreased the sales of unhealthy bundles by 56.29%, whereas T2 did not have a significant effect on the bundle sales. We next proceed to analyze the impact of the different treatments on the number of transactions of healthy and unhealthy snacks.

4.2.2. Number of Transactions. We estimated Equation (1) for the average daily number of transactions containing healthy snacks, unhealthy snacks, and unhealthy snacks without coffee. We report the treatment effect estimates ($\beta_{T_1}, \beta_{T_2}, \beta_{C_2}$) in Table 4 and find that the

Figure 4. (Color online) Average Daily Transactions for Different Types of Products Under Each Phase of the Experiment



results are consistent across all model specifications (i.e., with and without time fixed effects and with and without controlling for stockouts) for all three cases. In addition, when we changed the control condition to C2 (instead of C1), the results for both healthy and unhealthy snacks remained consistent.

For illustration purposes, Figure 4 plots the average daily sales of healthy and unhealthy snacks for each treatment along with the 95% confidence interval. Figure 4(a) suggests that T1 (respectively, T2) increased the sales of healthy snacks by an impressive 1,107.69% (respectively, 817.5%). In other words, we observed a strong, statistically significant positive effect on the sales of healthy snacks during both T1 and T2 (Table 4, panel A), which supports Hypothesis 1a but rejects Hypothesis 2a. This implies that although customers were offered a choice between a healthy and an unhealthy snack in T2, there was still a significant demand for healthy snacks relative to the case when there was no promotion on healthy snacks (C1). We also observed that the treatment effect was not significant for C2, indicating that there was no stickiness in the effect once the promotion was discontinued. This supports Hypothesis 4. Figure 4(b) shows that T1 led to a 36.52% drop in the sales of unhealthy snacks, whereas T2 and C2 did not have significant effects. Similarly, panel A of Table 4 reports a statistically significant decrease in the average daily sales of unhealthy snacks during T1 but no effect during T2, hence supporting Hypotheses 1b and 2b. This suggests that by replacing unhealthy snacks with healthy snacks in the bundle, consumers' purchases can be significantly shifted toward healthy choices. It is important to note that during T2, the unhealthy snack sales did not decrease, but the healthy snack sales increased significantly. This is because there were more individuals who were interested in the bundle. Thus, we tapped into consumer groups of both healthy and unhealthy snacks by providing a discount on both. Table 4, panel B, also reveals that the sales of unhealthy snacks purchased without coffee were not affected by the experiment, hence supporting Hypothesis 3. This implies that

Table 4. Impact of Different Interventions Using Weekday Sales on Quantity Sold

Panel A. Healthy and unhealthy snacks

	Healthy snacks				Unhealthy snacks			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	57.60***	57.60***	57.29***	58.16***	−49.40***	−49.40***	−48.47***	−49.35***
	(4.25)	(4.17)	(4.57)	(4.51)	(12.83)	(11.79)	(13.13)	(12.12)
T2	42.51***	42.61***	42.29***	43.02***	−6.77	−6.20	−6.33	−6.18
	(4.32)	(4.24)	(4.52)	(4.45)	(13.06)	(12.02)	(13.20)	(12.17)
C2	3.20	3.20	2.89	3.76	−16.67	−16.67	−16.67	−16.67
	(4.25)	(4.17)	(4.57)	(4.51)	(12.83)	(11.79)	(12.93)	(11.91)
No. observations	59	59	59	59	59	59	59	59
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.83	0.85	0.83	0.85	0.24	0.40	0.24	0.40

Panel B. Unhealthy snacks without coffee

Treatment	Model (1)	Model (2)	Model (3)	Model (4)
T1	5.47	5.47	6.27	6.30
	(4.60)	(4.45)	(4.67)	(4.53)
T2	3.77	3.75	4.14	4.06
	(4.68)	(4.53)	(4.69)	(4.54)
C2	−6.13	−6.13	−6.13	−6.13
	(4.60)	(4.45)	(4.60)	(4.45)
No. observations	59	59	59	59
Time FEs	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes
R^2	0.12	0.24	0.14	0.25

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^{***}p < 0.001$.

customers who usually purchase unhealthy snacks irrespective of the offered promotion were not affected by the type of bundle.

As discussed, all our results remained valid when including the weekend observations (for more details, see Appendix D, Section D.2). We also conducted an additional robustness test by excluding the days where stockouts occurred, and we observed consistent results (see Appendix D, Section D.3).

4.3. Heterogeneous Treatment Effects

In this section, we estimate the model from Equation (1) for the quantity sold while adding interaction terms between the treatment conditions (T_1, T_2, C_2) and the time of the day. To this end, we introduce a new covariate that segments the day into distinct time intervals. Specifically, we define early morning as the period between 12:00 a.m. and 5:00 a.m. (D_1), morning for the interval between 5:00 a.m. and 11:00 a.m. (D_2), noon for the period between 11:00 a.m. and 3:00 p.m. (D_3), evening for the period between 3:00 p.m. and 7:00 p.m. (D_4), and night for the hours following 7:00 p.m. up to 12:00 a.m. (D_5). We then use the following regression specification to estimate the treatment effects:

$$Q_{ip} = \alpha + \sum_{I \in \{T_1, T_2, C_2\}} \beta_I \cdot I + \sum_{j=2}^{5} \beta_{D_j} \cdot D_j + \sum_{I \in \{T_1, T_2, C_2\}} \sum_{j=2}^{5} \beta_{I, D_j} \cdot (I \times D_j) + \beta_s \cdot SO_{ip} + \mu_i + \epsilon_{ip},$$

where I represents the different treatments T_1, T_2, C_2, β_I is the coefficient for treatment I, D_j represents the different time windows as defined above, and β_{D_j} is the coefficient for time window D_j. We are interested in examining the interaction terms β_{I, D_j} to understand the effect of different treatments during different time periods. We present the findings on the interaction effects in Appendix E. We observe a substantial and statistically significant increase in the sales of healthy products during T1 across morning, noon, and evening periods, with percentage changes of 4,367.5%, 831.5%, and 971.9%, respectively, compared with C1. Likewise, during T2, we note a significant positive impact on healthy product sales across the same periods, with changes of 3,257.5%, 636%, and 628.1%, respectively,

Table 5. Impact of Different Interventions Using Weekday Sales on Quantity Sold Under a Negative Binomial Regression

Panel A. Healthy and unhealthy snacks

	Healthy snacks				Unhealthy snacks			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	2.49***	2.51***	2.39***	2.41***	−0.45***	−0.46***	−0.45***	−0.47***
	(0.38)	(0.38)	(0.41)	(0.41)	(0.12)	(0.12)	(0.12)	(0.12)
T2	2.22***	2.25***	2.20***	2.21***	−0.05	−0.05	−0.05	−0.05
	(0.39)	(0.39)	(0.40)	(0.41)	(0.12)	(0.12)	(0.12)	(0.12)
C2	0.48	0.50	0.38	0.41	−0.13	−0.13	−0.13	−0.13
	(0.39)	(0.39)	(0.42)	(0.42)	(0.12)	(0.12)	(0.12)	(0.12)
No. observations	59	59	59	59	59	59	59	59
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.61	0.62	0.62	0.62	0.24	0.38	0.24	0.38

Panel B. Unhealthy snacks without coffee

Treatment	Model (1)	Model (2)	Model (3)	Model (4)
T1	0.14	0.12	0.16	0.15
	(0.37)	(0.37)	(0.38)	(0.38)
T2	0.10	0.09	0.10	0.10
	(0.38)	(0.38)	(0.38)	(0.38)
C2	−0.18	−0.17	−0.18	−0.17
	(0.37)	(0.37)	(0.37)	(0.37)
No. observations	59	59	59	59
Time FEs	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes
R^2	0.11	0.19	0.12	0.2

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
***$p < 0.001$.

compared with C1. Notably, there is no discernible effect for C2 on healthy product sales at any time of the day. We also observe that the impact is most pronounced during the noon time period for both T1 and T2.

Conversely, a substantial and statistically significant negative effect is observed on the sales of unhealthy products during T1 across morning, noon, and evening periods, with percentage changes of 29.35%, 48.22%, and 34.04%, respectively, compared with C1. Similar to the healthy products, the effect is more prominent during the noon period. Unexpectedly, a negative effect was observed during C2 around noon, a phenomenon not identified in our preliminary analysis. However, consistent with prior observations, there is no significant effect on the sales of unhealthy products during T2. Moreover, Appendix E unveils that the sales of unhealthy products, when purchased without coffee, remain unaffected by the experiment regardless of the time of the day. A marginal negative effect is found for C2 during the morning hours.

5. Additional Analyses

In this section, we show the consistency of our findings by conducting several robustness tests, and we provide additional details on the generalizability of our findings. We do so by estimating a negative binomial regression, the DID approach, and the SDID methodology, as well as running an online survey to confirm our results and alleviate some potential concerns. Last, we address the potential one-store study limitation of our field experiment.

5.1. Negative Binomial Regression

Keeping in mind the count nature of our transaction data, we consider a negative binomial regression as a robustness test. The treatment effect estimates are presented in Table 5, demonstrating consistent results across various model specifications. These specifications include models with and without time fixed effects as well as with and without controlling for stockouts.

The findings reveal that exposure to Treatment 1 and Treatment 2 leads to a significant increase in the sales of healthy snacks by 1,106.12% and 820.73%, respectively. Control 2, however, exhibits no discernible effect on the sales of healthy snacks. Treatment 1 results in a notable reduction of unhealthy purchases by 36.23%, whereas Treatment 2 and Control 2 do not exhibit a significant effect. Consistent with our prior findings, the number of consumers who are purchasing unhealthy snacks without coffee are unaffected by the interventions.

5.2. Difference in Differences

Thus far, we have estimated various regression models without accounting for a possible treatment selection bias. To address this concern, we compared the sales in the treated store to other untreated stores in the same city during the same period. We relied on a DID specification to quantify the impact of the treatment conditions by contrasting the treated group's performance relative to an untreated group. We specify our DID model as follows:

$$Q_{ips} = \alpha + \sum_{I\in\{T_1,T_2,C_2\}} \beta_I \cdot I + \beta_T Treat_s + \sum_{I\in\{T_1,T_2,C_2\}} \gamma_I \cdot I \times Treat_s + \beta_{so} \cdot SO_{ips} + \mu_i + Store_s + \epsilon_{ips}, \qquad (2)$$

where Q_{ips} is the quantity sold on day i for product group p = {healthy snacks, unhealthy snacks} in store s, $Store_s$ represents store fixed effects, I represents the different treatments T_1, T_2, C_2 for day i, and

$$Treat_s = \begin{cases} 1 & \text{if observation occurs in the treated store s,} \\ 0 & \text{otherwise.} \end{cases}$$

Because we are interested in quantifying the impact of the treatments on the treated group relative to the control group, we focus on the interaction coefficients $\gamma_{T_1}, \gamma_{T_2}, \gamma_{C_2}$. The validity of the DID technique is based on the parallel trends assumption, namely, that no time-varying differences exist between the treatment and control groups. We tested the parallel trends assumption graphically by plotting the sales of both the treated and untreated stores and comparing the trends before the experiment. An alternative statistical technique to test the parallel trends assumption is by using the following equation (O'Neill et al. 2016, Han et al. 2019, Cui et al. 2020):

$$Q_{ips} = \alpha + \beta_1 \cdot d_i + \beta_2 Treat_s + \beta_3 \cdot d_i \times Treat_s + \epsilon_{ips}, \qquad (3)$$

where d_i represents the day counter, counting up to the start of the experiment during the pretreatment period. The above equation measures the effect of time on sales in a difference-in-differences fashion and was run using 12 weeks of data prior to the experiment. If the estimated coefficient $\beta_3 = 0$, then both groups would have the same slope before the experiment started and, hence, the parallel trends assumption would be satisfied.

In our analyses, we utilize various combinations of stores as the control group and show consistency in our results. More specifically, we have a pool of $n = 88$ stores from the focal retail chain in the same metropolitan city to choose from. We then consider the following two approaches to select our control group: (i) stores based on a close geographical distance from the treated store and (ii) stores based on similar coffee purchasing patterns (because coffee is our focal product).

The data used to estimate our various DID specifications were the historical point-of-sales data and end-of-day inventory data from the 88 stores starting from September 27, 2021. For better representation, we removed the days between December 20, 2021, and January 30, 2022, because of end-of-year holidays and city-wide COVID-19 restrictions.

5.2.1. Geographical Distance. The first control group for the DID analysis was to rely on all the (untreated) stores within a radius of 1 km from the treated store. There were two such stores that sold all the products used in our field experiment. We estimated the model in Equation (2) and report the interaction estimated coefficient $\boldsymbol{\beta_d}$ in Table 6. Model (1) reports the treatment effect without including any control variables, whereas Models (2)–(4) explicitly account for stockouts and time fixed effects. We can see that the sales of healthy snacks were significantly higher in the treated store during both T1 and T2, but did not show any significant change during C2. Similarly, the sales of unhealthy snacks decreased significantly for the treated store during T1 with no effect during T2 and C2. Finally, the sales of unhealthy snacks purchased without coffee remained unaffected by the different interventions. All these results are perfectly aligned with the results obtained in the previous section.

To test the parallel trends assumptions, we use the model in Equation (3) and inspect the estimated value of β_3. Our goal is to check whether $\beta_3 = 0$ with a p-value exceeding 0.05. The estimated β_3 values reported in Table G.3 in Appendix G satisfy the assumption. In Figure G.1, we also graphically convey that the parallel trends assumption is satisfied. Specifically, we plot the average weekly sales for the different product categories and observe a clear parallel trend.

We repeat this analysis using all the untreated stores within 2 and 3 km radii of the treated store as control groups. We report the DID estimates for the 2 km radius as the control group in Appendix G, Section G.1. The results are consistent with the estimates from our main specification.

5.2.2. Clustering Using Coffee Sales. The second control group we considered is by clustering stores based on historical coffee sales. The details of this analysis can be found in Appendix G, Section G.2. The results

Table 6. DID Estimates Using the Stores Within a 1 km Radius

Panel A. Healthy and unhealthy snacks

	Healthy snacks				Unhealthy snacks			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	57.27***	57.27***	57.54***	57.48***	−55.03***	−52.40***	−55.03***	−52.40***
	(3.00)	(2.97)	(3.00)	(2.98)	(9.40)	(9.17)	(9.40)	(9.17)
T2	42.27***	42.27***	42.67***	42.59***	−4.70	−5.86	−4.70	−5.86
	(3.00)	(2.97)	(3.01)	(2.99)	(9.40)	(9.17)	(9.40)	(9.17)
C2	2.17	2.17	2.64	2.55	−5.70	−4.78	−5.70	−4.78
	(3.00)	(2.97)	(3.01)	(2.99)	(9.40)	(9.17)	(9.40)	(9.17)
No. observations	177	177	177	177	177	177	177	177
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.90	0.91	0.90	0.91	0.86	0.87	0.86	0.87

Panel B. Unhealthy snacks without coffee

Treatment	Model (1)	Model (2)	Model (3)	Model (4)
T1	2.58	2.74	2.52	2.64
	(6.83)	(6.63)	(6.83)	(6.63)
T2	1.59	1.59	1.54	1.50
	(6.83)	(6.63)	(6.83)	(6.63)
C2	−6.93	−6.87	−7.00	−6.98
	(6.83)	(6.63)	(6.83)	(6.63)
No. observations	177	177	177	177
Time FEs	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes
R^2	0.86	0.87	0.86	0.87

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^{***}p < 0.001$.

for healthy snacks, unhealthy snacks, and unhealthy snacks without coffee were all consistent with those observed in our previous analyses.

5.3. Synthetic DID

The conventional DID approach is often challenged by limitations such as the assumption of parallel trends and the potential selection bias of control stores. To address these challenges, synthetic control methods have been proposed, which involve computing weights for control stores to create a weighted average that mimics the treated store's characteristics (Bekkerman et al. 2021, Yilmaz et al. 2024). However, this approach is contingent upon the treated store lying within the convex hull of the control stores, which may not always be the case. In our scenario, the treated store happens to fall outside this convex hull, hence indicating that the combined sales of the control stores are lower relative to the treated store. Consequently, using the synthetic control method would require violating certain weight constraints. In light of this, we explore alternative techniques and rely on the SDID methodology, a novel estimation approach that combines the benefits of DID and synthetic control methods while enhancing the precision of treatment effect estimation (Arkhangelsky et al. 2021).

The details on how to compute the treatment effects under the SDID can be found in Appendix H. The estimation results are presented in Table 7. We can see that the sales of healthy snacks, unhealthy snacks, and unhealthy snacks without coffee are all consistent with our previous analyses. This strengthens the validity of our results. We highlight that the parallel trends assumption is not a strong requirement for SDID. Nonetheless, it is still satisfied as shown in Table G.3.

5.4. Online Survey

Although the findings in the previous section remained consistent across multiple model specifications, there is still a possibility that the results are driven by the specific time period when the experiment was conducted and are affected by the time differences across treatments. To test our interventions' effects without time-related biases, we conducted a custom online survey with 2,000 individuals using the Prolific platform, an online subject recruitment platform that explicitly caters to researchers (see, e.g., Palan and Schitter 2018).

Table 7. SDID Estimates Using 15 Weeks of Pretreatment

	Healthy			Unhealthy			Unhealthy w/o coffee		
Treatment	T1	T2	C2	T1	T2	C2	T1	T2	C2
	49.48***	28.05***	−5.53	−49.45***	0.93	−13.01	3.66	3.44	−7.64
	(0.46)	(0.49)	(2.49)	(3.32)	(3.48)	(3.86)	(1.47)	(1.53)	(2.06)
No. observations	9,345	9,345	9,345	9,345	9,345	9,345	9,345	9,345	9,345
R^2	0.77	0.65	0.29	0.19	0.24	0.19	0.16	0.14	0.05

***$p < 0.001$.

5.4.1. Study Design. The survey aimed to replicate the decision-making process in a physical store, focusing on coffee, pastries, and healthy snack boxes (i.e., the same products as in our field experiment). Participants were given a budget and shown a specific promotion, as depicted in Figure 2. The budget amounts were determined by looking at the coffee-based transactions in the treated physical store with an average transaction amount of \$3.45. We decided to round up the budget (i.e., to \$4) to be used as a low budget for half of the participants in the laboratory study. This allowed participants to purchase a coffee while also being able to take advantage of the promotion (i.e., add a pastry or a healthy snack for an additional \$1). Similarly, a high budget of \$8 was used for half of the participants to allow them to purchase all three items offered in the survey if they desired to do so. After randomly assigning a budget value to the participants, we further randomly assigned them to one of three promotions (Figure 2): the control group viewed the promotion in Figure 2(a), the T1 group viewed Figure 2(b), and the T2 group viewed Figure 2(c). Participants who completed the survey successfully received a monetary compensation of \$1.

The main benefit of this online survey was that it fully eliminated any time-dependent effects that could have been present in our field experiment. Because all three promotions were run simultaneously in the form of a survey, it provides the perfect data to strengthen our results and overcome the shortcoming of having different intervention timings.

5.4.2. Results. The survey was set up to be completed by 2,000 respondents on the Prolific platform. We filtered out the survey responses that failed to pass the attention check questions, leaving us with 1,979 responses. We applied a filter to remove responses that took longer than three standard deviations from the mean to complete the survey, resulting in a final sample of 1,945 records. We investigate the balancedness of the experimental groups to make sure that the differences between the groups originated solely from the treatment and were unaffected by other factors. A chi-squared test indicates that the percentage of participants assigned to different promotions did not vary significantly based on budget, $\chi^2(2, N = 1{,}945) = 2.21$, $p = 0.33$, hence confirming that the sample is properly balanced. We use the following model specification to estimate the treatment effects:

$$P(R_{ip} = 1) = \frac{1}{1 + e^{-(\alpha + \sum_{I \in \{T_1, T_2\}} \beta_I \cdot I + \gamma B_i^H)}}, \quad (4)$$

where R_{ip} is a binary variable that indicates participant i's preference for product group $p = \{$healthy snacks, unhealthy snacks, coffee$\}$, T_1 and T_2 are binary variables indicating the treatment assigned to participant i, and

$$B_i^H = \begin{cases} 1 & \text{if participant } i\text{'s budget is \$8,} \\ 0 & \text{otherwise.} \end{cases}$$

The estimated coefficients $(\beta_{T_1}, \beta_{T_2})$ from Equation (4) are reported in Table 8. Model (1) presents the treatment effects for various snack types without considering participant budget, whereas Model (2) controls for the budget allocated to each participant. The results with and without controlling for the budget amount were consistent.

Under T1, the preference for healthy snacks was 1,536% higher than in the control, whereas under T2, it was 744% higher. In contrast, the preference for unhealthy snacks under T1 was 90% lower than in the control, and under T2, it was 43% lower. The lower preference for unhealthy snacks under T1 aligns with our field experiment results. However, T2's lower preference for unhealthy snacks seems to contradict our earlier result (Figure 4(b)). Fortunately, this discrepancy is totally intuitive and expected. The fixed number of respondents per intervention in the survey, unlike in our field experiment, prevented us from gauging the increase in affinity for the choice bundle observed in the field. Additionally, our analysis revealed that individuals with a higher budget are more inclined to purchase healthy snacks, but a higher budget does not alter the preference for unhealthy snacks.

In summary, the results from our online survey strongly support the findings from our field experiment, hence reinforcing the conclusion that both a healthy add-on bundle and a choice add-on bundle

Table 8. Effect of Different Interventions While Controlling for the Budget Value

	Healthy		Unhealthy		Unhealthy without coffee	
	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
Treatment 1	4.94***	4.98***	−4.82***	−4.83***	0.63	−0.61
	(0.21)	(0.22)	(0.20)	(0.20)	(0.44)	(0.44)
Treatment 2	2.69***	2.71***	−2.39***	−2.39***	−0.14	−0.15
	(0.19)	(0.19)	(0.17)	(0.17)	(0.52)	(0.52)
B_i^H		0.31**		−0.14		−0.45
		(0.13)		(0.13)		(0.38)
No. of observations	1,945	1,945	1,945	1,945	1,945	1,945
Budget	No	Yes	No	Yes	No	Yes
R^2	0.47	0.47	0.47	0.47	0.002	0.003

$^{**}p < 0.01$; $^{***}p < 0.001$.

significantly increase the likelihood of selecting healthy food choices. It also validates the fact that our results are not driven by the time differences across the different conditions.

5.5. Addressing the One-Store Study Limitation

A limitation of our field experiment is its implementation in a single store, potentially raising concerns on the generalizability of our results. To address this limitation and bolster the generalizability of our results, we undertake a exhaustive analysis considering factors such as the store's geographical location, customer demographics, and prevailing market conditions. We first perform an extensive market segmentation using store characteristics (store type, products sold, foot traffic, average basket size, average basket value, average daily coffee transactions, average daily pastry transactions, and average daily healthy snack transactions) and customer demographics based on store location (age, gender, type of dwelling, families, households, marital status, language, income, immigration and ethnocultural diversity, housing, education, journey to work, mobility, and migration). We find stores that are similar to the treated store based on the above set of features. This was followed by a series of A/A tests to ensure that the observed results remain robust and are not unduly influenced by store-specific factors of the treated store. We consider these similar stores as the control stores and perform a DID analysis to estimate the treatment effects for the treated store. We find consistent results indicating that the store selection for the field experiment did not substantially influence the results reported in this paper (see Appendix F for details).

6. Managerial Insights

In this section, we delve into the impact of our field experiment on the two main stakeholders: the retailer and the consumers. We begin by analyzing the revenue and profit, highlighting the effects of the interventions on the retailer. Subsequently, we shift our focus to exploring changes in consumer behavior across various interventions and summarize our takeaways.

6.1. Retailer's Perspective: Revenue and Profit Analysis

We investigate the impact of our interventions on the retailer's revenue and profit. Specifically, we examine the impact of the different bundles on the revenue and profit from the three product categories (healthy snacks, unhealthy snacks, and coffee). From the retail chain's perspective, a negative effect on revenue or on the profit would reduce the incentive to deploy this type of intervention at scale. The profit is computed using the difference between the selling price and the purchasing cost of each product. The profit value remained constant for all the products throughout our experiment. We then aggregate the profit values from the three product categories for the four phases of the experiment for further analysis.

The pairwise t-tests on revenue and profit between the four phases of our experiment are reported in Table 9. We found that the revenue and profit in T1 were not significantly different when compared with either C1 or C2. This is driven by the fact that the retailer earned a higher revenue and profit by charging

Table 9. Pairwise Comparisons of Revenue and Profit Between the Different Interventions

	Revenue		Profit	
	Mean difference	p-value	Mean difference	p-value
C1–T1	−76.88	0.33	22.12	0.65
C1–T2	−246.69	**0.01**	−130.02	**0.04**
C1–C2	63.81	0.51	19.47	0.75
T1–T2	−169.81	**0.05**	−152.14	**0.02**
T1–C2	140.69	0.15	−2.65	0.97
T2–C2	310.50	**0.01**	149.49	**0.04**

Note. The bold font indicates p-values ≤ 0.05 and emphasizes the significant change in revenue and profit between the different interventions.

the full price on unhealthy snacks during T1. Our analysis shows that these additional revenue and profit approximately offset the loss from the promotion on healthy snacks. This ultimately led to approximately similar average revenue and profit levels in the three stages. In addition, as we can see from Figure 4(c), there was no increase or decrease in coffee sales between T1 and C1 (or C2), hence ensuring that there was no overall negative impact on revenue or profit. However, the revenue and profit in T2 had a statistically significant positive effect relative to either C1 or C2 (and even T1). Specifically, we observe a 23.93% (respectively, 28.54%) profit increase and a 28.31% (respectively, 38.45%) revenue increase during T2 relative to C1 (respectively, C2). Because the sales of unhealthy snacks during T2 were not significantly different relative to C1 and C2 (from our results in Section 4), we attribute the increase in revenue and profit to the increase in sales of coffee bundles during T2 compared with C1 (which is found to be 25.21%). Indeed, the profit margin for coffee is much higher than the profit margins for healthy and unhealthy snacks, so that the profit increase from coffee largely compensates the profit loss incurred from healthy snacks. It is interesting to highlight that by offering both healthy and unhealthy snacks via a choice bundle in T2, the retailer can generate higher revenue and profit relative to offering only the healthy bundle in T1. The revenue (respectively, profit) in T2 was 17.91% (respectively, 29.1%) higher than in T1.

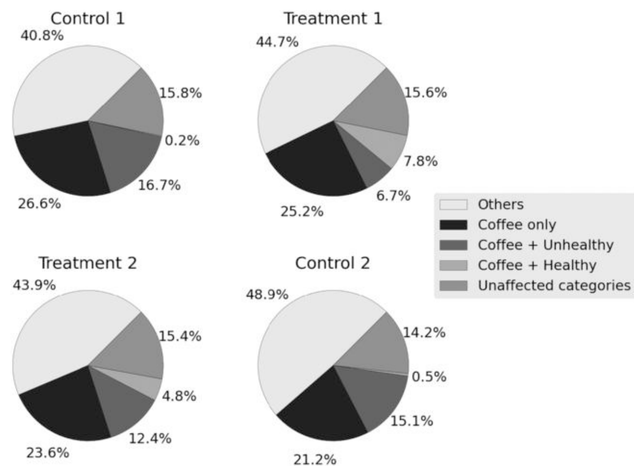
In summary, the above findings bear the following practical implications:

1. Offering an add-on bundle with only healthy items does not have a significant positive impact on revenue and profit. Thus, the retailer's bottom line is unaffected.
2. Offering an add-on choice bundle with either a healthy or an unhealthy item leads to a significant increase in both revenue and profit relative to a separate bundle (healthy or unhealthy).

6.2. Consumers' Perspective

It is interesting to examine how consumers' food preferences vary under the different interventions of our field experiment. In Figure 5, we plot the proportions of sales for the different types of purchases. We readily observe that the number of customers who purchased healthy snacks (C1 = 0.54%, T1 = 0.92%, T2 = 0.92%, C2 = 0.62%) and unhealthy snacks (C1 = 7.63%, T1 = 7.50%, T2 = 6.51%, C2 = 6.27%) outside the bundle, and the number of customers who purchased coffee together with items outside the experiment (C1 = 7.61%, T1 = 7.22%, T2 = 7.93%, C2 = 7.33%) remained roughly the same throughout all four phases of our experiment. These transactions were aggregated and represented as "unaffected categories." Under T1, we observe an increase of 7.6% (= 7.8 − 0.2) in healthy bundle purchases compared with C1, which is likely coming from customers who were purchasing only a coffee before (i.e., adapters) and from some of the customers who were purchasing the unhealthy bundle (i.e., switchers). The proportion of sales from other product categories actually increases by a slight 3.9% (= 44.7 − 40.8), hence indicating that there is no cannibalization effect from other product categories. We also observe that the decrease in unhealthy bundles was not entirely compensated by the increase in healthy bundles; that is, a portion of the customers who stopped purchasing unhealthy bundles switched to healthy bundles, whereas the remaining switched to "others." Similarly, under T2, the increase in healthy bundles amounts to 4.6% (= 4.8 − 0.2) compared with C1. The proportion of sales from other product categories increases by 3.1% (= 43.9 − 40.8). Thus, the primary reason behind the increase in healthy bundles was customers adopting the healthy bundle instead of purchasing only a coffee and from customers switching from the unhealthy to the healthy bundle. As discussed, we found that more than half of the customers continued to purchase the healthy bundle even when they were offered a choice between healthy and unhealthy snacks. When the promotion reverted back to the original unhealthy bundle (C2), the preferences for healthy and unhealthy snacks reached similar levels as in C1, so there is no long-term stickiness effect.

Figure 5. Customers' Preferences for Various Product Categories Under the Different Interventions



Notes. "Others" refers to all purchases outside the categories in our experiment (coffee, healthy snacks, and unhealthy snacks). "Coffee only" corresponds to transactions where only coffee was purchased. "Coffee + Unhealthy" corresponds to coffee purchases with pastry items. "Coffee + Healthy" corresponds to coffee purchases with healthy snacks."Unaffected categories" corresponds to purchases of unbundled healthy or unhealthy snacks or coffee plus other products.

7. Conclusion

According to Thaler and Sunstein (2021), nudging is becoming a key method to positively influence people's behavior. Nudging for social good often involves assisting individuals in adopting healthier and more sustainable lifestyles by leveraging their mental shortcuts, emotions, and surroundings (Chaurasia et al. 2022). In this context, private firms can also have a social impact when interacting with their customers. The study by Kroese et al. (2016) is a good example of nudging for social good, where visibility enhancement was used to nudge customers toward a healthier food alternative. A second example is from Cohen et al. (2023), who used nudging to encourage a more environmentally sustainable carpooling behavior for daily commuting.

In this paper, we focused on incentivizing retail customers to make healthier food choices by offering add-on bundles with healthy snacks. We investigated the impact of these bundling strategies on customers' purchases. We conducted two studies—a field experiment in a physical store and an online survey—to study this question. We considered three bundle combinations: (i) an unhealthy bundle (status quo), (ii) a healthy bundle, and (iii) a choice bundle. We found strong evidence that healthy snacks are purchased much more frequently when offered as part of a bundle. At the same time, the sales of unhealthy snacks are significantly reduced when they are not part of the add-on bundle. Unfortunately, however, there was no long-term stickiness; the preferences reverted back to the original levels when we stopped offering promotions on healthy snacks. Ultimately, we found that strategic add-on bundling incentivizes healthy food choices even when unhealthy items are included in a choice bundle. We conducted a series of robustness tests to showcase that our results are not driven by the time differences between treatments or by the store selected for the experiment. We also conveyed that well-designed bundles can increase the retailer's revenue and profit. Specifically, offering a choice bundle boosted the revenue and profit by 28.31% and 23.93%, respectively. Thus, offering such an add-on choice bundle is beneficial for customers (who can enjoy healthy food at a discount) and for retailers (who can earn higher revenue and profit). One of the limitations of our field experiment is that it was conducted in a single store, which may raise the concern of the generalizability of our findings. As discussed previously, we conducted a series of tests to address this concern, and we found that the experimental store is representative of several other stores, hence providing reassurance that the results are likely to be generalizable.

Acknowledgments

The authors thank department editor Wedad Elmaghraby, the associate editor, and the anonymous reviewers for their insightful comments, which have helped improve this paper. The authors also thank their C-store partner for the collaboration opportunity and for sharing the data.

Appendix A. Laboratory Experiments

A.1. Comparing Add-On Bundling and Price Discounting

A.1.1. Objective. We focus on the add-on bundling strategy because of the limited literature on its impact on consumer behavior compared with discounting. Although bundle offers often boost sales more than discounts (as noted by Competition Commission 2000), studies detailing bundling effects on food choices and retailer revenue are scarce. We believe that add-on bundling offers a more financially sustainable approach for retailers to promote healthy food choices. To validate this intuition, we conducted a supplementary Prolific survey to formally compare the add-on bundling strategy with price discounting. We discuss the survey design and promotions in detail in the next subsection.

A.1.2. Experiment Design. The survey includes three conditions: control, bundle, and discount (see Figure A.1). The control condition is the unhealthy bundle; namely, when

Figure A.1. (Color online) Promotion Banners Used in Our Laboratory Experiment

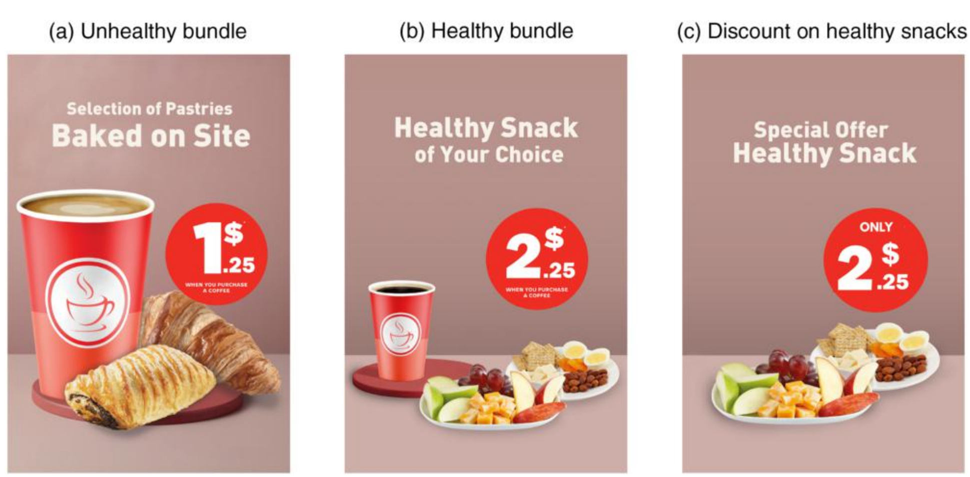


Table A.1. Promotions Offered in Our Laboratory Experiment

Experiment phase	Promotions offered	Discount offered on unhealthy snacks (%)	Discount offered on healthy snacks (%)	Promotion banner
Control 1 (C1)	Unhealthy add-on bundle	47.35	0	Figure A.1(a)
Treatment 1 ($T1_D$)	Healthy add-on bundle	0	47.55	Figure A.1(b)
Treatment 2 ($T2_D$)	Discount on healthy snacks	0	47.55	Figure A.1(c)

customers purchase a coffee beverage, they could add a pastry for an additional \$1.25 (resulting in a 47.35% discount). This is the same as the control condition (C1) in our field experiment. The only difference is that the add-on price for the unhealthy snacks is now set to \$1.25 instead of \$1 (Figure A.1(a)).[8] In the second condition (bundle), we offer the healthy bundle; specifically, when customers purchase a coffee, they could add a healthy snack for an additional \$2.25 (Figure A.1(b)). The original price of healthy snacks is \$4.29, and we are providing a discount of 47.55% to the consumers in the second condition. In the third condition (discount), we offer a discount on healthy snacks, which are sold at \$2.25 instead of the \$4.29 regular price (Figure A.1(c)). We are setting the price of the healthy snacks at the same discount of 47.55% as in the bundling condition. The only difference is that the consumer does not need to purchase the coffee to activate this lower price. The reason we selected a discounted price of \$2.25 is to ensure that the percentage discount on the healthy snacks (approximately 47%) matches the discount on unhealthy snacks. The survey aims to closely replicate the decision-making process from the physical store. The three promotions are simultaneously and randomly assigned to the participants as seen in Table A.1. Likewise, we randomly assigned each participant a budget of either \$6 (low) or \$10 (high) to be spent in the store. We ran the survey on the Prolific platform by recruiting a total of 500 participants to complete the survey.

A.1.3. Results. From the 500 responses collected, we first removed all the observations from individuals who failed to answer the attention-check questions. That left us with a total of 491 observations. We then performed a balancedness test to ensure that the differences across groups are only due to the difference in treatment and not to other factors. Our analysis indicated that participants were indeed assigned in a properly balanced randomized fashion, $\chi^2(2, N = 491) = 1.55, p = 0.46$. Figure A.2 displays the selection percentages for healthy snacks and unhealthy snacks under the different interventions (control, $T1_D$, and $T2_D$).

To assess the effectiveness of add-on bundling relative to discounting, and to evaluate the promotion of healthy snacks through either strategy versus the unhealthy bundle, we conducted pairwise *t*-tests. Table A.2 summarizes the results for the pairwise *t*-tests for both healthy and unhealthy snacks. Our analysis revealed a statistically significant increase of 12.23% in the likelihood of selecting healthy snacks when they are offered as part of a bundle, relative to being offered under a price discount. In addition, the likelihood of selecting unhealthy snacks is 15.57% lower when using a healthy add-on bundle relative to a healthy price discount. These findings support our intuition that using a bundling strategy to incentivize consumers toward healthy food choices is more effective than offering a price discount.

In addition, a compelling rationale for the superiority of the add-on bundling strategy over discounting lies in the advantage it offers to retailers. This aspect holds significant importance for retailers as, when they offer price discounts, they incur a clear profit loss. In contrast, under an add-on bundle, retailers can sell more units of the core item (in our

Figure A.2. (Color online) Summary Statistics of Our Laboratory Experiment

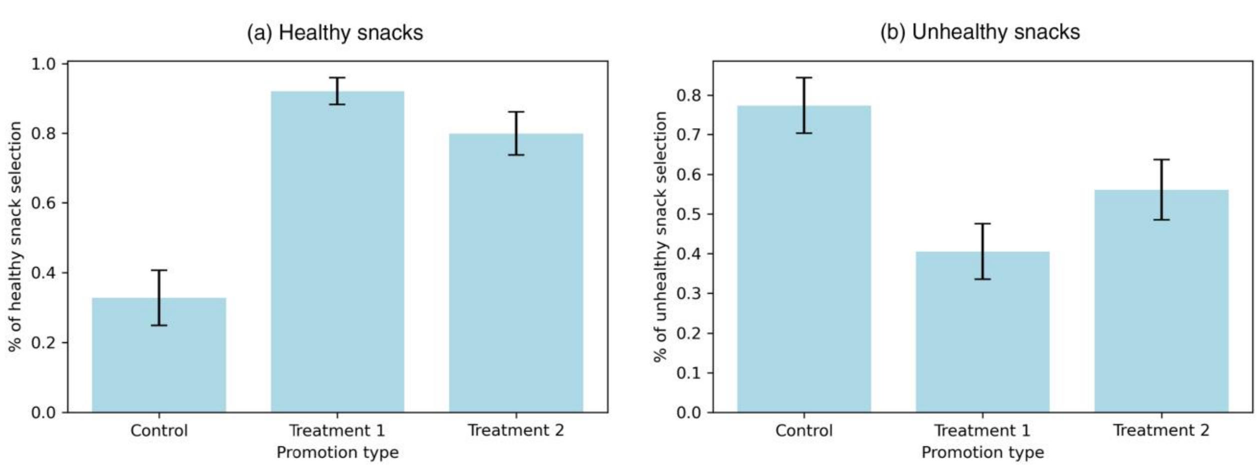


Figure A.3. (Color online) Flow of Our Online Survey

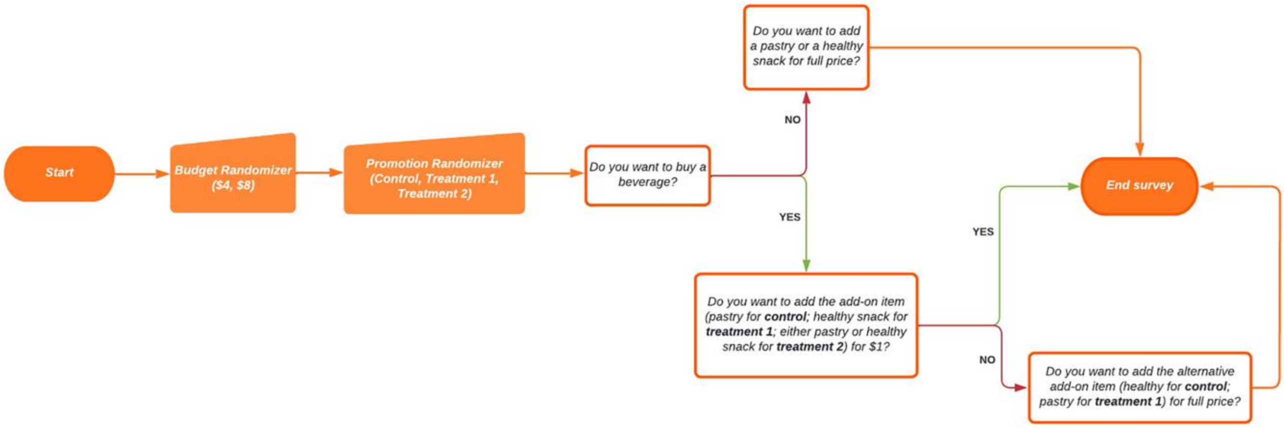
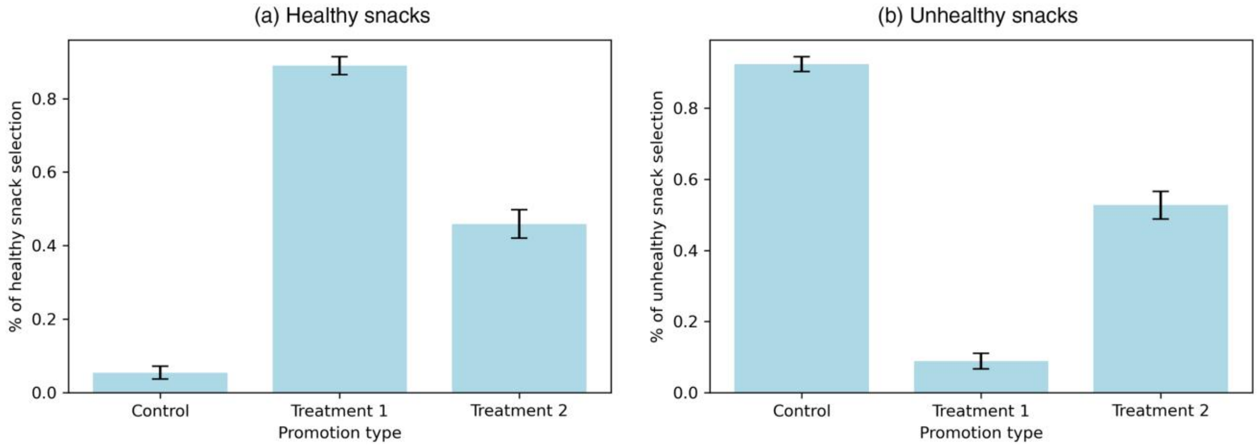


Figure A.4. (Color online) Summary Statistics of Our Laboratory Experiment



case, coffee) at full price. Thus, a well-crafted add-on bundling promotion can serve as a profit-generating strategy for retailers. We demonstrate this effect in our analysis in Section 6.1.

A.2. Add-On Bundling

The online survey was designed in Qualtrics and hosted on Prolific. The survey was anonymous and voluntary, and the first page of the survey served as a consent form (see Figure I.10 in the Online Appendix). We performed a cross-sectional study of American adults through the Prolific platform. The survey took three to four minutes to complete, and participants received a $1 compensation. Informed consent was not required because the data were anonymized. At each survey stage, participants could see the remaining budget and the prices of the items available for selection. This ensured that the participants could plan their purchases. The survey included two attention-check questions (Figure I.12, (a) and (b), in the Online Appendix) placed at the beginning and the end of the survey to augment data quality. We used the responses to the attention check questions to filter the respondents with high quality. The flow of the survey is as shown in Figure A.3. The complete transcript of the survey is provided in the Online Appendix. The summary statistics of the survey are shown in Figure A.4. For illustration purposes, Figure A.4 plots the preferences for healthy snacks, unhealthy snacks, and unhealthy snacks without coffee for each treatment along with a 95% confidence interval. These trends were consistent with the ones observed in our field experiment (see Figure 4).

Table A.2. Pairwise Comparisons of Consumer Preferences Between the Different Interventions

	Healthy snacks		Unhealthy snacks	
	Mean difference	*p*-value	Mean difference	*p*-value
C–$T1_D$	−0.59	0.00	0.37	0.00
C–$T2_D$	−0.47	0.00	0.21	0.00
$T1_D$–$T2_D$	0.12	0.001	−0.16	0.003

Appendix B. Definition of Healthy Food Items

The classification of items as healthy or unhealthy is an important aspect of our experimental design. Most individuals use the nutrient information on the packaging to classify products as healthy or unhealthy. One of the most common techniques used to categorize a food item as healthy or unhealthy relies on the amounts of nutrients and fats, carbohydrates, and sugar per kilocalorie of food. In addition, there exist several profiling techniques that are used to categorize food items. At a high level, healthy food items are considered to be nutrient dense; namely, they provide substantial levels of vitamins and minerals while containing relatively few calories (Drewnowski and Fulgoni 2008). The nutrient composition of the products under consideration in our field experiment is listed in Table B.1. We can then use this information to compute the score from nutrient profiling methods. Specifically, we use the CFN score, the RRR score, and the FSA rating. The three nutrient profiling models used to classify products as healthy or unhealthy can be summarized as follows:

1. *The CFN score* (Lachance and Fisher 1986): The lower the CFN value, the lower the calories needed to obtain the nutrients associated with a given food (and hence the healthier the food is). This is equivalent to computing how densely packed with nutrients a particular food is. However, this metric does not consider the nutrients that can be harmful when excessively consumed, such as sugar and carbohydrates. The CFN score can be computed as follows:

$$CFN = \frac{ED}{\sum_{i=1}^{13} \%DV_i/13},$$

where ED is the energy density of the food item measured in kilocalories and the denominator corresponds to the average daily value percentages of 13 nutrients, namely, protein, thiamin, riboflavin, niacin, folate, calcium, iron, zinc, magnesium, and vitamins A, C, B6, and B12. To compute the CFN score, one needs to scale the nutrients available in 100 g of the food item.

2. *The RRR score* (Scheidt and Daniel 2004): A higher value of RRR translates into a healthier food item. This metric computes the ratio of recommended nutrient values (e.g., vitamins) with the restricted ones (e.g., sugars, fats, carbohydrates). This metric is more comprehensive because it relies on both the recommended and nonrecommended nutrients. The RRR score can be computed by using the following formula:

$$RRR = \frac{\sum_{i=1}^{6} Nutrient\_recommended_i/6}{\sum_{i=1}^{5} Nutrient\_restricted_i/5}.$$

The recommended nutrients are protein, fiber, vitamins A and C, calcium, and iron. The restricted nutrients are energy, saturated fats, sugar, cholesterol, and sodium. The score is computed per serving of the item.

3. *The FSA rating*:[9] This rating provides an integer value, and any food item with a value below four is considered healthy. It also accounts for both the recommended and nonrecommended nutrients. In addition, it explicitly considers whether a particular food item contains fruits, vegetables, and nuts. To compute the FSA score, one needs to scale the nutrients available in 100 g of the food item. The FSA scoring algorithm can be divided into the following three steps:

   a. Compute the total "A" points = (points for energy) + (points for saturated fats) + (points for sugar) + (points for sodium).
   b. Compute the total "C" points = (points for percentages of fruits, vegetables, and nut content) + (points for fiber) + (points for protein)
   c. Final score = Total A points − total C points, if total A points are lower than 11. Otherwise, we do not count points toward protein, unless total C points are higher than 5.

The computed values for the healthy snacks and the average values for the pastry items are reported in Table 1. For the pastry items, we use an average value for simplicity. Overall, it is clear that the pastry items have a much higher score relative to the three healthy snacks. One exception is the fruit snack box that contains natural sugar, as opposed to the pastry items that have artificial sweeteners. This key difference is accounted for in the FSA score but not in the other profiling methods.

Appendix C. Mediation Analysis

In this section, we perform a mediation analysis to confirm that the stockout variable is not a significant mediator in the relationship between promotional interventions and consumer purchase behavior. It could be argued that the stockout variable is on the causal path between the intervention and the purchase outcome. Hence, it could theoretically absorb the true treatment effect and should not be included in the

Table B.1. Nutrient Information for Healthy and Unhealthy Snacks

Snack type	Total energy (kilocalorie)	Fats (g)	Sugar (g)	Carbohydrates (g)	Fiber (g)	Protein (g)	Potassium (mg)	Calcium (mg)	Iron (mg)
Vegetable box	190	0.5	6	10	3	2	500	40	0.75
Fruit box	230	10	22	27	3	8	250	225	0.4
Protein box	430	27	16	32	4	18	500	125	3
Croissant	272	14	7.5	31	1.7	5.5	79	0	0
Cinnamon roll	184	16	17	32	0.8	3.1	0	18	0
Apple turnover	285	13	25	41	1.6	2.4	0	18	0.75
Fruit Danish	263	20	20	34	1.3	3.8	59	33	1.26
Chocolate avalanche	320	16	11	37	3	5	0	40	1.8
Chocolate muffin	318	14	27	45	0.8	3.8	0	38	0

Table C.1. Mediation Effects

Treatment	Healthy			Unhealthy			Unhealthy without coffee		
	ACME	ADE	Total	ACME	ADE	Total	ACME	ADE	Total
T1	−0.56	58.11***	57.56***	−0.06	−49.27***	−49.33***	−0.83	6.37	5.54
	[−4.94,2.67]	[49.00, 67.07]	[48.82, 66.47]	[−6.70,6.52]	[−74.24, −25.62]	[−74.62, −26.40]	[−3.99,1.01]	[−2.62, 15.03]	[−3.50, 14.52]
T2	−0.41	43.05***	42.64***	−0.04	−6.50	−6.46	−0.36	4.04	3.68
	[−3.29,2.28]	[35.02, 51.51]	[34.05, 51.43]	[−5.06,5.26]	[−29.32, 17.87]	[−29.62, 18.21]	[−3.06,1.74]	[−4.90, 12.30]	[−5.59, 12.43]
C2	−0.62	3.94	3.32	−0.09	−16.14	−16.23	−0.01	−6.29	−6.28
	[−4.52,2.92]	[−5.01, 13.20]	[−4.53, 12.02]	[−4.53,3.78]	[−38.66, 7.98]	[−38.30, 7.25]	[−2.04,2.31]	[−15.03,2.55]	[−15.37,3.12]

Note. The 95%-confidence interval errors are reported in parentheses.
$^{***}p < 0.001$.

regression analysis. Mediation analysis helps us understand both the direct and indirect effects, hence providing insights into the specific role of stockouts in this relationship. Our model considers the quantity sold (Q_{ip}) as the outcome variable and the stockout occurrences (SO_{ip}) as the mediator. More precisely, the model is defined as follows:

$$Q_{ip} = \alpha_1 + \sum_{I \in \{T_1, T_2, C_2\}} \beta_I \cdot I + \beta_s \cdot SO_{ip} + \mu_i + \epsilon^1_{ip},$$

$$SO_{ip} = \alpha_2 + \sum_{I \in \{T_1, T_2, C_2\}} \beta^s_I \cdot I + \epsilon^2_{ip},$$

where $\beta_{T_1}, \beta_{T_2}, \beta_{C_2}$ are the coefficients of treatments T_1, T_2, C_2, respectively; μ_i are time fixed effects; and $\epsilon^1_{ip}, \epsilon^2_{ip}$ represent the error terms. We compute the average causal mediation effects (ACMEs), the average direct effects (ADEs), and the combined total effect for each product category (healthy snacks, unhealthy snacks, and unhealthy snacks without coffee) based on the same methodology as in Section 4.2. ACME quantifies the average outcome change resulting from the mediation process, while ADE represents the average outcome change directly linked to the treatment. The combined total effect encapsulates the overall treatment impact, considering both direct and indirect effects through one or more mediators. For simplicity, we present the results exclusively for the model with time fixed effects and stockouts (Model (4)). The mediating variable, *Stockouts*, shows a statistically insignificant ACME across all product categories (see Table C.1), indicating that it does not mediate our analysis. Moreover, the ADE's proximity to the combined total effect reinforces our confidence in the estimates from Section 4.2.

Appendix D. Robustness Tests

D.1. Data Filtering

Excluding the extreme observations mitigates the potential impact of outlier values on the observed effects. By removing the instances with exceptionally large purchases, we aimed to obtain a more conservative (and representative) estimate of the treatment effects, hence ensuring that our results were not disproportionately influenced by a small number of high-value transactions. We acknowledge that the frequency of such occurrences is low but felt that their exclusion would contribute to the robustness of our analysis. We replicated the analysis outlined in Section 4.2 without applying any data filtering procedure. We focused on two dependent variables: bundles sold and the number of transactions. For conciseness and consistency, our reported results are based on the data set excluding weekends, following the methodology outlined in the main paper. However, we highlight that the results exhibit consistency even when including weekend observations.

Figure D.1 presents the average daily sales of healthy and unhealthy snacks for each treatment, along with a 95% confidence interval. Additionally, Table D.1 reports the treatment effect estimates ($\boldsymbol{\beta}$) for healthy, unhealthy, and all add-on bundles. These estimates align with the insights reported in Table 3. Similarly, Table D.2 reports the treatment effects for the number of transactions for healthy and unhealthy snacks, demonstrating once again an alignment with the results from

Figure D.1. (Color online) Average Daily Transactions for Different Products Under Each Experiment Phase Without Removing Outliers

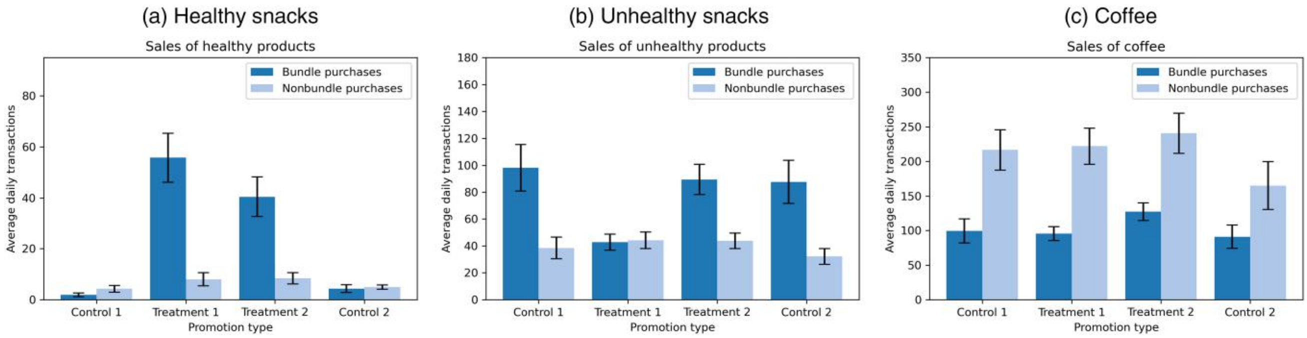


Table D.1. Impact of Different Interventions Using Weekday Sales on Bundles Sold Without Data Filtering

	Healthy add-on bundle				Unhealthy add-on bundle				All coffee add-on bundles			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	54.47***	54.47***	54.67***	54.94***	−55.40***	−55.40***	−52.80***	−54.59***	−3.67	−3.67	−3.67	−3.67
	(3.68)	(3.69)	(3.80)	(3.82)	(9.63)	(8.79)	(10.10)	(9.33)	(10.44)	(9.33)	(10.44)	(9.33)
T2	39.13***	39.13***	39.24***	39.37***	−8.67	−8.67	−6.71	−8.06	25.67*	26.40***	25.67*	26.40***
	(3.68)	(3.69)	(3.73)	(3.74)	(9.63)	(8.79)	(9.90)	(9.13)	(10.63)	(9.51)	(10.63)	(9.51)
C2	2.47	2.47	2.67	2.94	−10.47	−10.47	−9.71	−10.06	−8.27	−8.27	−8.27	−8.27
	(3.68)	(3.69)	(4.80)	(3.82)	(9.63)	(8.79)	(76)	(8.99)	(10.44)	(9.33)	(10.44)	(9.33)
No. observations	60	60	60	60	60	60	60	60	60	60	60	60
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.85	0.86	0.85	0.86	0.42	0.55	0.43	0.55	0.18	0.39	0.18	0.39

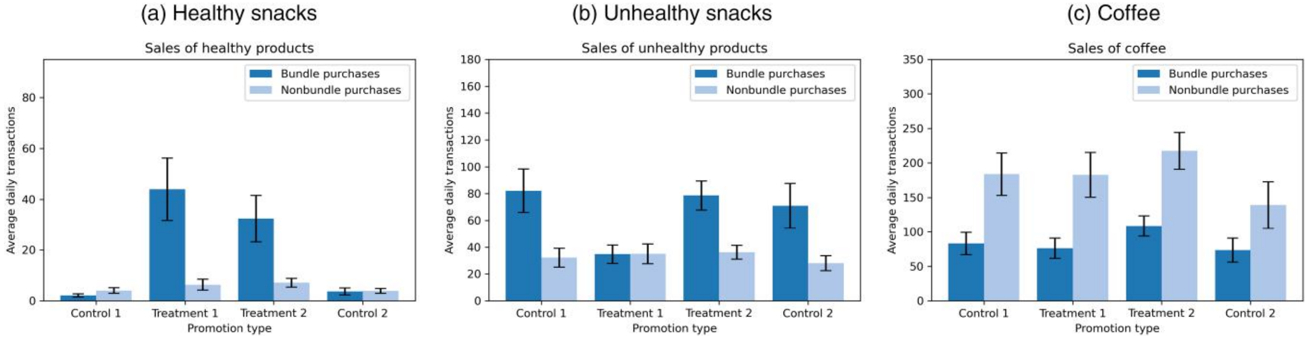
Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^{*}p < 0.05$; $^{***}p < 0.001$.

Table D.2. Impact of Different Interventions Using Weekday Sales on Quantity Sold Without Data Filtering

	Healthy snacks				Unhealthy snacks				Unhealthy without coffee			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	58.20***	58.20***	58.36***	58.74***	−49.73***	−49.73***	−44.78***	−46.73***	5.67	5.67	8.02	7.86
	(4.17)	(4.09)	(4.31)	(4.23)	(13.06)	(11.98)	(13.60)	(12.66)	(4.71)	(4.60)	(4.85)	(4.78)
T2	43.27***	43.27***	43.35***	43.54***	−3.33	−3.33	−0.38	−1.08	5.33	5.33	7.10	6.98
	(4.17)	(4.09)	(4.23)	(4.15)	(13.06)	(11.98)	(13.34)	(12.39)	(4.71)	(4.60)	(4.76)	(4.68)
C2	3.13	3.13	3.29	3.68	−16.73	−16.73	−14.26	−15.23	−6.27	−6.27	−5.09	−5.17
	(4.17)	(4.09)	(4.31)	(4.23)	(13.06)	(11.98)	(13.15)	(12.19)	(4.71)	(4.60)	(4.69)	(4.61)
No. observations	60	60	60	60	60	60	60	60	60	60	60	60
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.84	0.86	0.84	0.86	0.24	0.41	0.27	0.42	0.13	0.23	0.17	0.26

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^{***}p < 0.001$.

Figure D.2. (Color online) Average Daily Transactions for Different Types of Products Under Each Phase of the Experiment (Including Weekend Observations)



Note. The error bars indicate the 95% confidence intervals.

Table 4. The only difference in the results reported here compared with Section 4.2 is the magnitude of the treatment effects.

D.2. Regression with Weekend Sales

In this section, we show that the results presented in Section 4.2 are robust to the inclusion of weekend observations. Accordingly, we include the data from all the weekends and reestimate the treatment effects for both the bundles sold and the quantity sold as we did in the main paper. Figure D.2 shows the average daily transactions of both bundled and unbundled purchases of healthy snacks, unhealthy snacks, and coffee under the different treatments. As we can see, the trends are very similar to the ones observed without including the weekend observations. Table D.3 supports our previous results on coffee bundle preferences being higher during T2. The sales of healthy and unhealthy coffee add-on bundles align as well. Similarly, Table D.4 reports the treatment effects for the overall products purchased with and without controlling for stockouts and time fixed effects. All the results are consistent with the estimates presented in Table 4 from Section 4.2. When you consider the consumer preference for coffee add-on bundles, T2 has a positive effect, whereas T1 and C2 have no significant impact on it (Table D.3). T1 and T2 have a strong positive effect on the number of healthy add-on bundles sold, whereas C2 does not affect it. Finally, T1 has a negative effect on the number of unhealthy add-on bundles purchased, and T2 and C2 had no significant effect on it (Table D.3). T1 and T2 have a positive effect, whereas C2 does not affect the number of transactions with healthy snacks (Table D.4). When we include weekend observations, we find that T1 (respectively, T2) increased the number of transactions with healthy snacks by 926.12% (respectively, 705.12%). Table D.4 indicates that T1 reduced the overall purchases of unhealthy snacks, whereas T2 and C2 had no statistically significant effect. The sales of unhealthy snacks decreased by 38.97% during T1. None of the interventions had a significant effect on the unhealthy snacks purchased without coffee. Overall, we are not introducing data selection bias by excluding weekends from the data in the main analysis. The analysis in this section confirms our findings from the main paper.

Table D.3. Impact of Different Interventions Using Weekday and Weekend Sales on Bundles Sold

	Healthy add-on bundle				Unhealthy add-on bundle				All coffee add-on bundles			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	42.67***	42.67***	41.21***	42.76***	−47.43***	−47.43***	−45.86***	−47.71***	−7.00	−7.00	−7.00	−7.00
	(4.35)	(3.72)	(4.59)	(3.96)	(9.65)	(7.15)	(9.74)	(7.29)	(11.42)	(7.26)	(11.42)	(7.26)
T2	31.06***	31.40***	30.37***	31.44***	−3.54	−2.53	−2.80	−2.63	25.46*	26.80**	25.46*	26.80**
	(4.41)	(3.77)	(4.46)	(3.85)	(9.77)	(7.24)	(9.78)	(7.30)	(11.56)	(7.36)	(11.56)	(7.36)
C2	1.52	1.52	0.31	1.60	−11.29	−11.29	−11.29	−11.29	−9.90	−9.90	−9.90	−9.90
	(4.35)	(3.72)	(4.52)	(3.90)	(9.65)	(7.15)	(9.64)	(7.19)	(11.42)	(7.26)	(11.42)	(7.26)
No. observations	83	83	83	83	83	83	83	83	83	83	83	83
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.65	0.76	0.65	0.76	0.28	0.64	0.29	0.64	0.13	0.67	0.13	0.67

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table D.4. Impact of Different Interventions Using Weekday and Weekend Sales on Quantity Sold

	Healthy snacks				Unhealthy snacks				Unhealthy snacks without coffee			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	45.38***	45.86***	43.13***	45.32***	−44.52***	−44.52***	−41.90***	−44.33***	2.90	2.90	2.90	2.90
	(4.90)	(4.11)	(5.13)	(4.33)	(13.30)	(9.37)	(13.36)	(9.55)	(4.60)	(3.54)	(4.60)	(3.54)
T2	34.55***	35.37***	33.47***	35.15***	0.51	1.80	1.76	1.87	4.05	4.33	4.05	4.33
	(4.96)	(4.17)	(4.99)	(4.23)	(13.47)	(9.49)	(13.42)	(9.57)	(4.66)	(3.58)	(4.66)	(3.58)
C2	1.76	2.24	−0.12	1.80	−15.38	−15.38	−15.38	−15.38	−4.10	−4.10	−4.10	−4.10
	(4.90)	(4.11)	(5.05)	(4.27)	(13.30)	(9.37)	(13.23)	(9.43)	(4.60)	(3.54)	(4.60)	(3.54)
No. observations	83	83	83	83	83	83	83	83	83	83	83	83
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.62	0.73	0.62	0.73	0.16	0.61	0.18	0.61	0.04	0.48	0.04	0.48

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^{***}p < 0.001$.

D.3. Stockout Occurrences

In this section, we show that the results presented in Section 4.2 are robust to the exclusion of stockout observations. Accordingly, we exclude all the days on which a stockout occurred and reestimate the treatment effects for the quantity sold. The number of stockout days and frequency of stockouts for different products during different treatments are reported in Table D.5. First, we highlight that coffee was never stocked out, and there was no issue with the coffee machine during our field experiment. The stockout variable used in the analysis is binary. Because we are looking at a product category, we set the stockout variable to one for healthy products when there is one or fewer varieties available in the store. Similarly, for the unhealthy snacks, we set the stockout variable to one when there is less than 25% variety available in the store. This variable is calculated using the inventory information at the day level.

We further conducted an analysis similar to the analysis in Section 4.2 by removing the stockout days from the data. For consistency, we show the results only for the weekday data. The treatment effects are reported in Table D.6. For healthy snacks, removing the stockout data resulted in a reduction of six data points from the weekdays. Though the coefficients of the treatment effects slightly changed relative to the main paper, the direction and significance of the treatment effects remained the same. Similarly, for unhealthy snacks and unhealthy snacks without coffee, the number of days remaining after removing the stockout days is 57 weekdays. We observe that the results are consistent for both Treatment 1 and Treatment 2. For Control 2, we observe a significant decrease in transactions in Model (2) when we include time fixed effects. We repeated this analysis by including the weekend data and observed similar results. In summary, the vast majority of the results remain consistent with our main results even when we remove the data for the stockout days.

Table D.5. Frequency and Number of Stockout Days for Healthy and Unhealthy Snacks During Our Experiment

Product category	Treatment	Number of days	Frequency (%)
Healthy	Control 1	6	28.57
	Treatment 1	0	0
	Treatment 2	3	14.28
	Control 2	1	4.76
Unhealthy	Control 1	3	14.28
	Treatment 1	4	19.04
	Treatment 2	1	4.76
	Control 2	1	4.76

Table D.6. Impact of Different Interventions Using Weekday Sales on Quantity Sold After Removing Stockout Days

	Healthy snacks		Unhealthy snacks		Unhealthy without coffee	
Treatment	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
T1	56.35***	57.43***	−55.59***	−57.73***	2.50	1.92
	(4.79)	(4.74)	(13.14)	(11.65)	(4.60)	(4.31)
T2	40.55***	41.68***	−12.96	−14.42	0.80	0.26
	(4.95)	(4.90)	(13.36)	(11.84)	(4.68)	(4.37)
C2	1.95	3.03	−22.86	−25.00*	−9.10	−9.68*
	(4.79)	(4.74)	(13.14)	(11.65)	(4.60)	(4.31)
No. observations	54	54	57	57	57	57
Time FEs	No	Yes	No	Yes	No	Yes
R^2	0.82	0.84	0.28	0.48	0.13	0.30

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
*$p < 0.05$; ***$p < 0.001$.

Appendix E. Heterogeneous Treatment Effects: Impact of Different Interventions Using Weekday Sales and Hour of the Day on Quantity Sold

	Healthy snacks		Unhealthy snacks		Unhealthy without coffee	
Treatment	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
T1: Morning	17.47***	17.47***	−13.07*	−13.07*	1.87	1.87
	(2.15)	(2.13)	(5.44)	(5.22)	(2.11)	(2.11)
T1: Noon	22.20***	22.20***	−23.47***	−23.47***	0.33	0.33
	(2.15)	(2.13)	(5.44)	(5.22)	(2.11)	(2.11)
T1: Evening	14.87***	14.87***	−11.80*	−11.80*	1.93	1.93
	(2.15)	(2.13)	(5.44)	(5.22)	(2.11)	(2.11)
T1: Night	3.07	3.07	−1.07	−1.07	1.33	1.33
	(2.15)	(2.13)	(5.44)	(5.22)	(2.11)	(2.11)
T2: Morning	13.03***	13.03***	−5.18	−5.18	−0.90	−0.90
	(2.19)	(2.17)	(5.53)	(5.31)	(2.15)	(2.11)
T2: Noon	16.98***	16.98***	−9.10	−9.10	−0.18	−0.18
	(2.19)	(2.17)	(5.53)	(5.31)	(2.15)	(2.11)
T2: Evening	9.61***	9.61***	4.48	4.48	3.75	3.75
	(2.19)	(2.17)	(5.53)	(5.31)	(2.15)	(2.11)
T2: Night	2.90	2.90	3.03	3.03	1.1	1.1
	(2.19)	(2.17)	(5.53)	(5.31)	(2.15)	(2.11)
C2: Morning	1.67	1.67	−4.60	−4.60	−4.47*	−4.47*
	(2.15)	(2.13)	(5.44)	(5.22)	(2.11)	(2.07)
C2: Noon	0.27	0.27	−13.73*	−13.73**	−4.00	−4.00
	(2.15)	(2.13)	(5.44)	(5.22)	(2.11)	(2.07)
C2: Evening	0.80	0.80	−2.20	−2.20	0.27	0.27
	(2.15)	(2.13)	(5.44)	(5.22)	(2.11)	(2.07)
C2: Night	0.47	0.47	3.87	3.87	2.07	2.07
	(2.15)	(2.13)	(5.44)	(5.22)	(2.11)	(2.07)
No. observations	295	295	295	295	295	295
Time FEs	No	Yes	No	Yes	No	Yes
Stockouts	No	Yes	No	Yes	No	Yes
R^2	0.79	0.79	0.73	0.76	0.62	0.64

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Appendix F. Store Selection Bias

Because the treatment store was not selected at random and was selected in collaboration with the retailer, we devote this section to conducting a series of A/A tests to demonstrate that the outcomes are not impacted by this specific selection. To achieve this, we initially performed a store clustering procedure to effectively segment the market.

In the city where the experiment was conducted, the convenience store chain we partnered with operates a total of 89 stores (including the treated store). We narrowed it down to 77 stores selling the three product categories central to our experiment—namely, coffee, pastry items, and healthy snack boxes. The next step in the store selection process was to segment the customers from different stores. The first round of customer segmentation was based on whether a store sells fuel or not. The retailer operates both traditional stand-alone C-stores and C-stores located within gas stations. Recognizing potential differences in customer shopping behavior between these two types of stores, our study focuses on the traditional stand-alone C-stores. This left us with 35 stores. We then proceeded with customer segmentation based on the demographic data of the store location. For this task, we used demographic data (age, gender, type of dwelling, families, households, marital status, language, income, immigration and ethnocultural diversity, housing, education, labor, journey to work, mobility, and migration) along with the store characteristics, such as foot traffic, average basket size, average basket value, average daily coffee transactions, average daily pastry transactions, and average daily healthy snack transactions to perform a store clustering procedure to identify similar stores. The number of features turned out to be 44. To cluster the stores using the above high-dimensional data, we used several techniques: (i) *K*-means with dimensionality reduction, (ii) agglomerative clustering, and (iii) spectral clustering. The number of clusters obtained using each of these clustering techniques is reported in Table F.1. It is evident that the agglomerative clustering approach with three clusters performs best (i.e., it has the highest silhouette score). We proceeded to use this technique to find the stores that are most similar to the treated store. There were nine other stores in the same city that had similar characteristics as the treated store in terms of demographics and store characteristics. We would like to emphasize that both agglomerative clustering and spectral clustering yield clusters with 90% similarity in the stores in the cluster with the treated store. The next step was to perform a series of A/A tests to showcase that the treated store is representative, and ultimately showcase the generalizability of our results.

For A/A testing, we used the preintervention data and the data from the control periods (Control 1, Control 2) of the experiment across the 10 stores that were clustered together. The following equation was used for the estimation:

$$\begin{aligned} Y_{ips} &= \alpha + \beta_1 \cdot day\_count_i + \beta_T Treat_s \\ &\quad + \gamma \cdot day\_count_i \times Treat_s + \mu_i + Store_s + \epsilon_{ips}, \end{aligned}$$

where Y_{ips} is $Foot\_traffic_{ips}, Basket\_value_{ips}, Revenue_{ips}$, and $Coffee\_transactions_{ips}$. The variable $Treat_s$ takes the value one for the store in which we ran the experiment and zero for all other stores. The term day_count_i counts the number of days starting from the earliest date in our data. In our case, it was 18 weeks before the experiment began. We used other time fixed effects denoted by μ_i. We used these four metrics to compare the treated store to the other stores in the same cluster during the preexperiment period and to the control period of the field experiment. It is important to note that the promotion running during this time in all the stores is set to the unhealthy add-on bundle. In the above specification, if γ is statistically significant, it would imply that the trends in our metrics are different between the treated store and the similar untreated stores from the same cluster. We report the γ estimates in Table F.2. We find that the estimates are all statistically insignificant, thus indicating that the A/A tests passed. This demonstrates that the treated store and the control stores have similar characteristics, such as foot traffic, basket value, revenue, and number of coffee transactions at the day and product levels. We then proceeded to perform a DID analysis using these 10 stores (treated store and nine control stores from the same cluster) by using the following specification:

$$\begin{aligned} Q_{ips} &= \alpha + \sum_{I \in \{T_1, T_2, C_2\}} \beta_I \cdot I + \beta_T Treat_s \\ &\quad + \sum_{I \in \{T_1, T_2, C_2\}} \gamma_I \cdot I \times Treat_s + \beta_s \cdot SO_{ips} + \mu_i + Store_s + \epsilon_{ips}, \end{aligned}$$

where Q_{ips} is the quantity sold on day i for product group $p = \{\text{healthy, unhealthy}\}$ in store s, $Store_s$ represents store fixed effects, and

$$Treat_s = \begin{cases} 1 & \text{if observation occurs in the treated store,} \\ 0 & \text{otherwise.} \end{cases}$$

Because we were interested in quantifying the impact of the treatments on the treated store relative to the control group, we focused on the interaction coefficients $\gamma_{T_1}, \gamma_{T_2}, \gamma_{C_2}$. Table F.3 reports the DID estimates for healthy snacks, unhealthy snacks, and unhealthy snacks without coffee. We observed that the overall results align with those from our main analysis. The direction of the effects remained the same, whereas the estimate of the treatment effects varied slightly.

Table F.1. Optimal Number of Clusters and Silhouette Score for Different Clustering Techniques

Clustering technique	Number of clusters	Silhouette score
K-means with principal component analysis	2	0.08
Agglomerative clustering	3	0.62
Spectral clustering	3	0.60

Table F.2. Estimates for A/A Tests

	Foot traffic	Basket value	Revenue	Coffee transactions
Day × treat	0.0098	0.0022	1.38	0.0009
	(0.12)	(0.003)	(0.99)	(0.05)
N	1,680	1,680	1,680	1,680
R^2	0.70	0.28	0.512	0.74

Table F.3. DID Estimate Using the Stores from Demographics Clustering

	Healthy snacks				Unhealthy snacks				Unhealthy without coffee			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	57.57***	57.57***	57.83***	57.83***	−51.98***	−51.98***	−52.26***	−52.26***	4.27	4.27	4.06	4.06
	(1.45)	(1.45)	(1.42)	(1.42)	(5.85)	(5.79)	(5.87)	(5.81)	(2.73)	(2.72)	(2.74)	(2.73)
T2	42.87***	42.87***	43.12***	43.12***	−5.71	−5.71	−5.99	−5.99	3.07	3.07	2.86	2.86
	(1.45)	(1.45)	(1.42)	(1.42)	(5.85)	(5.79)	(5.87)	(5.81)	(2.73)	(2.72)	(2.74)	(2.73)
C2	2.88*	2.88*	3.04	3.04	−10.59	−10.59	−10.78	−10.78	−8.87*	−8.87*	−9.01*	−9.01*
	(1.45)	(1.45)	(1.42)	(1.42)	(5.85)	(5.79)	(5.86)	(5.80)	(2.73)	(2.72)	(2.73)	(2.72)
No. observations	600	600	600	600	600	600	600	600	600	600	600	600
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.91	0.91	0.92	0.92	0.84	0.85	0.84	0.85	0.76	0.76	0.76	0.76

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^*p < 0.05$; $^{***}p < 0.001$.

Appendix G. DID Results

G.1. Robustness Tests with a Larger Radius

We conducted DID analysis using untreated stores within a 2 km radius of the treated store as a control group, totaling seven stores. The results, presented in Table G.1, align with the main specification estimates and strengthen our confidence in our results. We verified the parallel trends assumption for each model, ensuring consistency. Extending the analysis to a 3 km radius with 18 stores yielded consistent results.

G.2. Clustering Using Coffee Sales

The second control group we considered was determined by clustering stores based on historical coffee sales. Because coffee was the primary product in all the add-on bundles used in our experiment, it seemed natural to select control stores with a similar level of coffee sales. To perform the clustering, we used the weekly aggregated sales of coffee during the 12 weeks before the experiment for each store. We constructed a data set that represents the weekly coffee sale patterns in all stores. This data were scaled before clustering to account for the data variability across the different stores. We implemented the K-means method to cluster the scaled data. Using the elbow method, we determined the optimal number of clusters to be three. We then identified the cluster that contained the treated store and used all the other stores in the same cluster (four of them) as our control stores for the DID analysis. We report the interaction estimated coefficient $\boldsymbol{\beta}_d$ in Table G.2 for healthy snacks, unhealthy snacks, and unhealthy snacks without coffee. The results for healthy snacks, unhealthy snacks, and unhealthy snacks without coffee were consistent with those observed in all previous analyses. The results of testing the parallel trends assumption are reported in Table G.3.

G.3. Parallel Trends Assumption

This section discusses the parallel trends assumption required for the DID analysis from Sections 5.2.1 and G.2. Table G.3 reports the estimated values of β_3 when estimating Equation (3). The goal is to check whether $\beta_3 = 0$ with a statistically significant p value. Figures G.1 and G.2 plot the number of healthy, unhealthy, and unhealthy without coffee transactions as a

Table G.1. DID Estimates Using the Stores Within a 2 km Radius

	Healthy snacks				Unhealthy snacks				Unhealthy without coffee			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	57.48***	57.40***	57.39***	57.48***	−55.98***	−56.05***	−56.72***	−56.79***	2.36	2.31	2.09	2.04
	(28.61)	(28.76)	(27.52)	(27.74)	(−4.65)	(−4.65)	(−4.72)	(−4.73)	(0.42)	(0.42)	(0.38)	(0.38)
T2	43.05***	43.01***	42.99***	43.05***	−15.37	−15.80	−15.73	−16.15	−0.28	−0.49	−0.43	−0.65
	(20.80)	(20.87)	(20.47)	(20.58)	(−1.23)	(−1.27)	(−1.27)	(−1.30)	(−0.05)	(−0.09)	(−0.08)	(−0.11)
C2	2.34	2.22	2.25	2.30	−12.03	−12.13	−12.54	−12.66	−9.06	−9.10	−9.24	−9.29
	(1.16)	(1.11)	(1.08)	(1.11)	(−0.92)	(−0.93)	(−0.96)	(−0.97)	(−1.50)	(−1.50)	(−1.53)	(−1.54)
No. observations	649	649	649	649	649	649	649	649	649	649	649	649
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Stock out	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.90	0.90	0.90	0.90	0.43	0.44	0.44	0.44	0.23	0.23	0.23	0.23

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
$^{***}p < 0.001$.

Table G.2. DID Estimates Using Clustering Based on Coffee Sales

	Healthy snacks				Unhealthy snacks				Unhealthy without coffee			
Treatment	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)	Model (1)	Model (2)	Model (3)	Model (4)
T1	57.62***	57.62***	57.73***	57.73***	−64.67***	−67.72***	−64.67***	−64.67***	−5.73	−6.14	−5.73	−6.14
	(2.49)	(2.48)	(2.46)	(2.45)	(10.36)	(10.26)	(9.78)	(9.55)	(6.84)	(6.65)	(6.84)	(6.65)
T2	43.33***	43.33***	43.13***	43.12***	−16.62	−19.44	−18.35	−18.25	−12.16	−9.82	−12.16	−9.82
	(2.49)	(2.48)	(2.46)	(2.45)	(10.36)	(10.25)	(9.78)	(9.56)	(6.84)	(6.65)	(6.84)	(6.65)
C2	3.31	3.31	3.10	3.10	−17.63	−16.46	−17.37	−17.26	−1.71	−3.22	−1.71	−3.22
	(2.49)	(2.48)	(2.46)	(2.45)	(10.36)	(10.25)	(9.78)	(9.56)	(6.84)	(6.65)	(6.84)	(6.65)
No. observations	236	236	236	236	236	236	236	236	236	236	236	236
Time FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Stockouts	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.90	0.91	0.90	0.91	0.58	0.6	0.58	0.6	0.66	0.69	0.66	0.69

Notes. The standard errors are reported in parentheses. FEs, Fixed effects.
***$p < 0.001$.

Table G.3. Testing the Parallel Trends Assumption for the DID Analysis

	Healthy snacks		Unhealthy snacks		Unhealthy w/o coffee	
	β_3	p-value	β_3	p-value	β_3	p-value
Stores within a 1 km radius	0.0098	0.312	0.1465	0.393	−0.0197	0.761
Clustering based on coffee sales	0.023	0.051	0.1970	0.216	−0.0086	0.914
SDID	0.0054	0.732	0.35	0.138	0.024	0.787

Figure G.1. (Color online) Parallel Trends Assumption for Stores Within a 1 km Radius

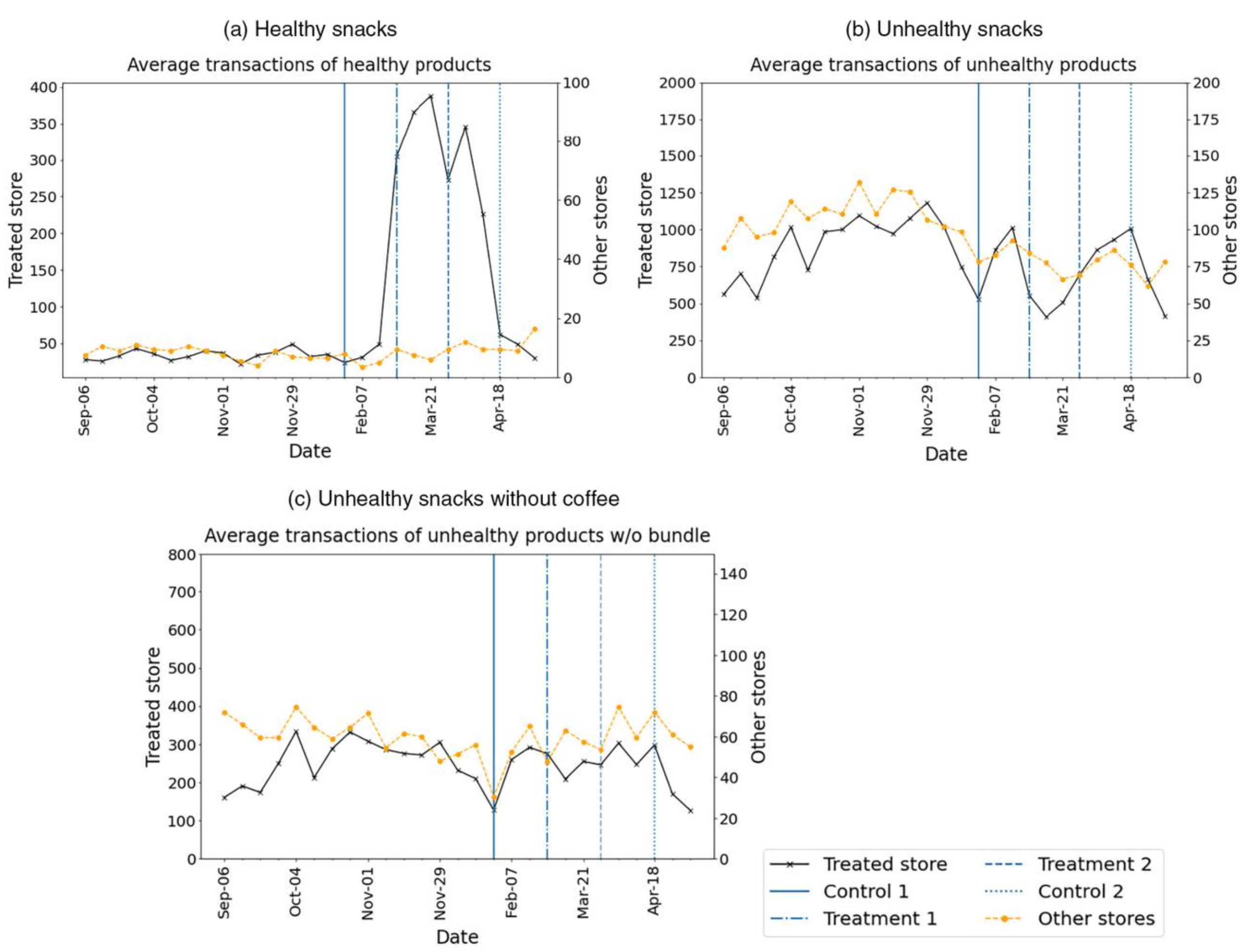
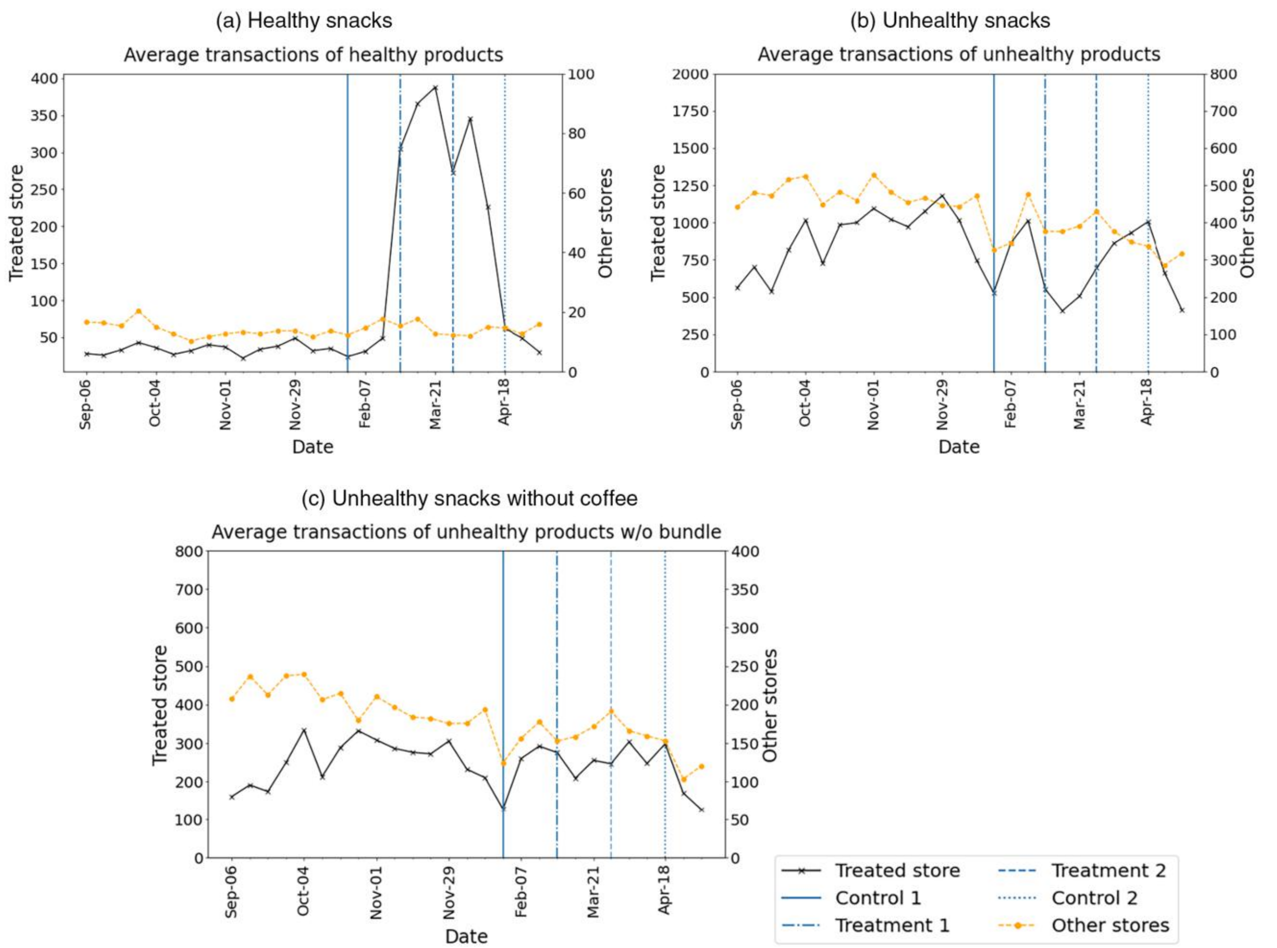


Figure G.2. (Color online) Parallel Trends Assumption Using Clustering Based on Coffee Sales



function of the time for the different treatments in the treated store along with the average over the control stores. As we can see, the parallel trends assumption is satisfied based on the results from both Figures G.1 and G.2 and Table G.3.

Appendix H. Synthetic DID

Algorithm H.1 (SDID)

Data: Data: Q_{ips}, I

Result: Output: $\hat{\gamma}_I$

1 Compute the regularization parameter ζ using the below:

$$\zeta = (N_{tr} * T_{post})^{1/4}\hat{\sigma} \text{ with}$$

$$\hat{\sigma}^2 = \frac{1}{N_{co}(T_{pre}-1)\sum_{i=1}^{N_{co}}\sum_{t=1}^{T_{pre}-1}(\Delta_{it}-\bar{\Delta})}$$

where

$$\Delta_{it} = Y_{i(t+1)} - Y_{it} \text{ and } \bar{\Delta} = \sum_{i=1}^{N_{co}}\sum_{t=1}^{T_{pre}-1}\Delta_{it}$$

2 Compute the unit weights $\hat{w}_s$ by solving the below optimization:

$$(\hat{w}_0, \hat{w}_s) = \underset{w_0\in\mathbb{R}, w\in\Omega}{\arg\min}\sum_{t=1}^{T_{pre}}\left(w_0 + \sum_{i=1}^{N_{co}} w_i Y_{it} - \frac{1}{N_{tr}}\sum_{N_{co}+1}^{N} Y_{it}\right)^2 + \zeta^2 T_{pre}||w||_2^2,$$

$$\Omega = \left\{ w \in \mathbb{R}_+^N : \sum_{i=1}^{N_{co}} w_i = 1, w_i = N_{tr}^{-1} \text{ for all } i = N_{co}+1 \right\};$$

3 **for** *each* $(I, I_s, I_e) \in \{(T_1, T_{pre}+1, T_{T_1}), (T_2, T_{T_1}+1, T_{T_2}), (C_2, T_{T_2}+1, T_{C_2})\}$ **do**

Compute the time weights $\hat{\boldsymbol{\lambda}}^I$:

$$(\hat{\lambda}_0, \hat{\boldsymbol{\lambda}}^I) = \underset{\lambda_0 \in \mathbb{R}, \boldsymbol{\lambda}^I \in \Lambda}{\arg\min} \sum_{i=1}^{N_{co}} \left(\lambda_0 + \sum_{t=1}^{T_{pre}} \lambda_t Y_{it} - \frac{1}{T_{post}} \sum_{I_s}^{I_e} Y_{it} \right)^2,$$

$$\Lambda = \left\{ \lambda \in \mathbb{R}_+^T : \sum_{t=1}^{T_{pre}} \lambda_t = 1, \lambda_t = T_{post}^{-1} \text{ for all } t = I_s, \ldots, I_e \right\}$$

Compute the SDID estimator via the weighted DID regression:

$$(\hat{\gamma_I}, \hat{\alpha}, \hat{\beta_I}, \hat{\beta}_{so}, \hat{Store}_s, \hat{\mu}_i) = \underset{\alpha, \beta_I, \gamma_I, \beta_{so}, Store_s, \mu_i}{\arg\min} \left\{ \sum_{s=1}^{N=88} \left[\sum_{i=1}^{T_{pre}} \theta_I + \sum_{i=I_s}^{I_e} \theta_I \right] \right\}$$

where,

$$\theta_I = (Q_{ips} - \alpha - \beta_I \cdot I - \gamma_I \cdot T \times Treated_s - \beta_{so} \cdot SO_{ips} - \mu_i - Store_s)^2 \hat{w}_s \hat{\lambda}_i^I$$

In this section, we provide more detailed explanations for implementing the SDID estimation process:

1. *Unit weights calculation*: The initial step involves determining the unit weights assigned to all control stores. Unit weights are designed so that the average outcome for the treated units is approximately parallel to the weighted average of control units. Importantly, this process remains consistent across all interventions and does not necessitate repetition for each specific treatment.

2. *Time weights*: Once the unit weights for the control stores are determined, the analysis proceeds to calculate the time weights. Time weights are crucial for addressing temporal dynamics in the experiment. Time weights are designed so that the average posttreatment outcome for each of the control units differs by a constant from the weighted average of the pretreatment outcomes for the same control units.

Figure H.1. (Color online) Time Weights for SDID

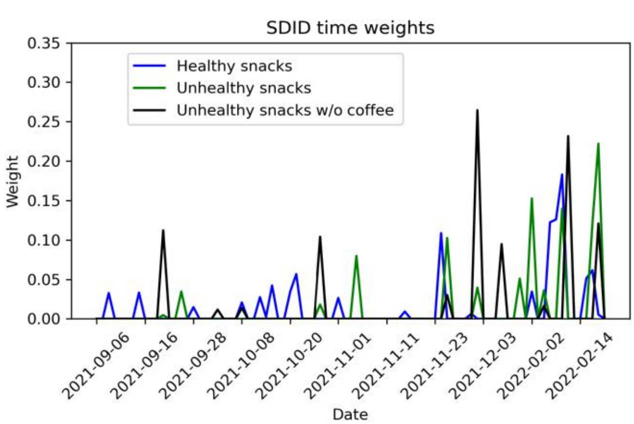


Table H.1. Unit Weights for SDID

Healthy		Unhealthy		Unhealthy w/o coffee	
Store ID	Unit weights	Store ID	Unit weights	Store ID	Unit weights
1	0.0787	3	0.3006	21	0.142211
8	0.0694	2	0.2014	3	0.101928
83	0.0690	39	0.1976	19	0.100787
87	0.0678	21	0.1628	79	0.082726
42	0.0668	74	0.1296	74	0.078028
21	0.0649	87	0.0053	70	0.066583
5	0.0401	88	0.0028	42	0.056149
31	0.0392	83	0.0000	26	0.051796
2	0.0366	22	0.0000	36	0.047387
17	0.0349	38	0.0000	50	0.04181
11	0.0345	77	0.0000	88	0.039547
7	0.0285	26	0.0000	30	0.036402
32	0.0222	20	0.0000	2	0.035255
24	0.0216	11	0.0000	34	0.026853
43	0.0202	5	0.0000	18	0.023384

3. *SDID estimates*: Building on the unit and time weights, the SDID estimates are then computed. This involves combining the unit weights with the time weights to derive a comprehensive measure of the causal effect for each specific intervention using the equation shown in Algorithm H.1. The SDID estimate is a key output that reflects the net impact of the intervention, accounting for both unit-specific characteristics and temporal variations.

As indicated in Algorithm H.1, the unit weights ($\hat{w}_s$) are computed only once for all the treatments. We report the highest 15 estimated unit weights (out of 88) in Table H.1. The time weights ($\hat{\lambda}_i^I$), on the other hand, are computed for each treatment separately. We report the time weights ($\hat{\lambda}_i$) for 15 weeks in Figure H.1. Both the unit weights and the time weights are calculated so that they sum up to one. The SDID estimates ($\gamma_{T_1}, \gamma_{T_2}, \gamma_{C_2}$) are reported in Table 7. We note that this algorithm is an adapted version based on the original framework presented in Arkhangelsky et al. (2021).

Endnotes

[1] See https://ncdalliance.org/why-ncds/risk-factors-prevention/unhealthy-diets-and-malnutrition.

[2] Bundling involves buying multiple units of the same product or a combo offer that includes several products. It is usually presented as "Pay $X when you buy both products A and B."

[3] In our data, less than 15% of the overall transactions involved healthy items. The healthy and unhealthy categorization was based on the FSA technique. See Appendix B for more details.

[4] Using a slight abuse of notation, we used coffee to represent all hot beverages in the rest of this paper because coffee purchases accounted for 93% of the total hot beverage sales.

[5] The reservation price of a product is the maximum price a consumer is willing to pay for the product. The conditional reservation price is the reservation price of a product conditional on the consumer buying another product.

[6] See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216094/dh_123492.pdf.

[7] In this paper, all confidence intervals are reported at the 95% level.

[8] We note that the add-on price was increased by 25¢ because of inflation across all food categories in 2023 relative to 2021 when our original field experiment was conducted.

[9] See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216094/dh_123492.pdf.

References

Afshin A, Peñalvo JL, Del Gobbo L, Silva J, Michaelson M, O'Flaherty M, Capewell S, Spiegelman D, Danaei G, Mozaffarian D (2017) The prospective impact of food pricing on improving dietary consumption: A systematic review and meta-analysis. *PLoS One* 12(3):e0172277.

Ainslie G (1975) Specious reward: A behavioral theory of impulsiveness and impulse control. *Psych. Bull.* 82(4):463–496.

Allcott H, Rogers T (2014) The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *Amer. Econom. Rev.* 104(10):3003–3037.

An R (2013) Effectiveness of subsidies in promoting healthy food purchases and consumption: A review of field experiments. *Public Health Nutrition* 16(7):1215–1228.

Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S (2021) Synthetic difference-in-differences. *Amer. Econom. Rev.* 111(12):4088–4118.

Bekkerman R, Cohen MC, Liu X, Maiden J, Mitrofanov D (2021) The impact of the opportunity zone program on residential real estate. Preprint, submitted March 9, http://dx.doi.org/10.2139/ssrn.3780241.

Bennett R, Zorbas C, Huse O, Peeters A, Cameron AJ, Sacks G, Backholer K (2020) Prevalence of healthy and unhealthy food and beverage price promotions and their potential influence on shopper purchasing behaviour: A systematic review of the literature. *Obesity Rev.* 21(1):e12948.

Cadario R, Chandon P (2020) Which healthy eating nudges work best? A meta-analysis of field experiments. *Marketing Sci.* 39(3): 465–486.

Carroll KA, Samek A, Zepeda L (2018) Food bundling as a health nudge: Investigating consumer fruit and vegetable selection using behavioral economics. *Appetite* 121(February):237–248.

Chaurasia S, Pati RK, Padhi SS, Jensen JM, Gavirneni N (2022) Achieving the United Nations Sustainable Development Goals 2030 through the nutraceutical industry: A review of managerial research and the role of operations management. *Decision Sci.* 53(4):630–645.

Cohen MC, Kalas JJ, Perakis G (2021) Promotion optimization for multiple items in supermarkets. *Management Sci.* 67(4):2340–2364.

Cohen MC, Fiszer MD, Ratzon A, Sasson R (2023) Incentivizing commuters to carpool: A large field experiment with Waze. *Manufacturing Service Oper. Management* 25(4):1263–1284.

Competition Commission (2000) Supermarkets: A report on the supply of groceries from multiple stores in the United Kingdom (Stationery Office, London).

Cui R, Li M, Li Q (2020) Value of high-quality logistics: Evidence from a clash between SF Express and Alibaba. *Management Sci.* 66(9):3879–3902.

Drewnowski A, Fulgoni V III (2008) Nutrient profiling of foods: Creating a nutrient-rich food index. *Nutrition Rev.* 66(1):23–39.

Farley TA, Rice J, Bodor JN, Cohen DA, Bluthenthal RN, Rose D (2009) Measuring the food environment: Shelf space of fruits, vegetables, and snack foods in stores. *J. Urban Health* 86(5):672–682.

Fishbach A, Friedman RS, Kruglanski AW (2003) Leading us not into temptation: Momentary allurements elicit overriding goal activation. *J. Personality Soc. Psych.* 84(2):296.

Gordon E, ICF International (2014) Approaches for promoting healthy food purchases by SNAP participants: ICF International. Report, U.S. Food and Nutrition Service, U.S. Department of Agriculture, Washington, DC.

Guiltinan JP (1987) The price bundling of services: A normative framework. *J. Marketing* 51(2):74–85.

Han BR, Sun T, Chu LY, Wu L (2019) Connecting customers and merchants offline: Experimental evidence from the commercialization of last-mile stations at Alibaba. *40th Internat. Conf. Inform. Systems* (Association for Information Systems, Atlanta).

Hawkes C (2009) Sales promotions and food consumption. *Nutrition Rev.* 67(6):333–342.

Hinnosaar M (2023) The persistence of healthy behaviors in food purchasing. *Marketing Sci.* 42(3):521–537.

Hirschman EC, Holbrook MB (1982) Hedonic consumption: Emerging concepts, methods and propositions. *J. Marketing* 46(3):92–101.

Janiszewski C, Cunha M Jr (2004) The influence of price discount framing on the evaluation of a product bundle. *J. Consumer Res.* 30(4):534–546.

Khan U, Dhar R (2010) Price-framing effects on the purchase of hedonic and utilitarian bundles. *J. Marketing Res.* 47(6):1090–1099.

Khan U, Dhar R, Wertenbroch K (2005) A behavioral decision theory perspective on hedonic and utilitarian choice. Pham MT, Higgins ET, Ratneshwar S, Mick DG, eds. *Inside Consumption: Frontiers of Research on Consumer Motives, Goals, and Desires* (Routledge, London), 166–187.

Kroese FM, Marchiori DR, De Ridder DT (2016) Nudging healthy food choices: A field experiment at the train station. *J. Public Health* 38(2):e133–e137.

Lachance PA, Fisher MC (1986) Educational and technological innovations required to enhance the selection of desirable nutrients. *Clinical Nutrition* 5(6):257–264.

Larson NI, Story MT, Nelson MC (2009) Neighborhood environments: Disparities in access to healthy foods in the US. *Amer. J. Preventive Medicine* 36(1):74–81.

Lobstein T, Davies S (2009) Defining and labelling "healthy" and "unhealthy" food. *Public Health Nutrition* 12(3):331–340.

Martin A (2007) Did McDonald's give in to temptation? *New York Times* (July 2), https://www.nytimes.com/2007/07/22/business/yourmoney/22feed.html.

Muhammad A, D'Souza A, Meade B, Micha R, Mozaffarian D (2017) The influence of income and prices on global dietary patterns by country, age, and gender. Report, U.S. Department of Agriculture, Washington, DC.

Neslin SA (2002) Sales promotion. Weitz BA, Wensley R, eds. *The Handbook of Marketing* (SAGE, Thousand Oaks, CA), 311–338.

Ni Mhurchu C, Blakely T, Jiang Y, Eyles HC, Rodgers A (2010) Effects of price discounts and tailored nutrition education on supermarket purchases: A randomized controlled trial. *Amer. J. Clinical Nutrition* 91(3):736–747.

Nikolova HD, Inman JJ (2015) Healthy choice: The effect of simplified point-of-sale nutritional information on consumer food choice behavior. *J. Marketing Res.* 52(6):817–835.

O'Neill S, Kreif N, Grieve R, Sutton M, Sekhon JS (2016) Estimating causal effects: Considering three alternatives to difference-in-differences estimation. *Health Services Outcomes Res. Methodology* 16(1):1–21.

Palan S, Schitter C (2018) Prolific.ac—A subject pool for online experiments. *J. Behav. Experiment. Finance* 17(March):22–27.

Putler DS (1992) Incorporating reference price effects into a theory of consumer choice. *Marketing Sci.* 11(3):287–309.

Roglic G (2016) WHO global report on diabetes: A summary. *Internat. J. Noncommunicable Diseases* 1(1):3–8.

Salmon SJ, De Vet E, Adriaanse MA, Fennis BM, Veltkamp M, De Ridder DTD (2015) Social proof in the supermarket: Promoting healthy choices under low self-control conditions. *Food Quality Preference* 45:113–120.

Scheidt DM, Daniel E (2004) Composite index for aggregating nutrient density using food labels: Ratio of recommended to restricted food components. *J. Nutrition Ed. Behav.* 36(1):35–39.

Shiv B, Fedorikhin A (2004) Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. *J. Consumer Res.* 26(3):278–292.

Sogari G, Velez-Argumedo C, Gómez MI, Mora C (2018) College students and eating habits: A study using an ecological model for healthy behavior. *Nutrients* 10(12):1823–1839.

Stern D, Ng SW, Popkin BM (2016) The nutrient content of US household food purchases by store type. *Am. J. Preventive Medicine* 50(2):180–190.

Stremersch S, Tellis GJ (2002) Strategic bundling of products and prices: A new synthesis for marketing. *J. Marketing* 66(1): 55–72.

Thaler RH, Sunstein CR (2021) *Nudge* (Yale University Press, New Haven, CT).

van Kleef E, Otten K, van Trijp HCM (2012) Healthy snacks at the checkout counter: A laboratory and field study on the impact of shelf arrangement and assortment structure on consumer choices. *BMC Public Health* 12(1):1–10.

Ver Ploeg M, Mancino L, Todd JE, Clay DM, Scharadin B (2015) Where do Americans usually shop for food and how do they travel to get there? Initial findings from the National Household Food Acquisition and Purchase Survey. Report, U.S. Department of Agriculture, Washington, DC.

Wang EST (2017) Different effects of utilitarian and hedonic benefits of retail food packaging on perceived product quality and purchase intention. *J. Food Products Marketing* 23(3):239–250.

Warren M (2023) C-store IQ: 2023 purchasing report. *Canada Convenience Store News* (July 5), https://ccentral.ca/c-store-iq-2023-purchasing-report.

Wilcox K, Vallen B, Block L, Fitzsimons GJ (2009) Vicarious goal fulfillment: When the mere presence of a healthy option leads to an ironically indulgent decision. *J. Consumer Res.* 36(3):380–393.

Yilmaz O, Son Y, Shang G, Arslan HA (2024) Causal inference under selection on observables in operations management research: Matching methods and synthetic controls. *J. Oper. Management*, ePub ahead of print June 19, https://doi.org/10.1002/joom.1318.