Offline Assortment Optimization in the Presence of an Online Channel

Daria Dzyabura  
Stern School of Business, New York University, New York, NY 10012, ddzyabur@stern.nyu.edu

Srikanth Jagabathula  
Stern School of Business, New York University, New York, NY 10012, sjagabat@stern.nyu.edu

With the proliferation of multiple sales channels, a firm’s operational decisions must account for the switching of consumers between different channels during their purchase process. This paper considers the assortment problem faced by a firm selling products that vary on multiple features through an online and an offline channel. The firm carries a ‘large’ collection of products online and must decide on the subset of products to carry in its offline store. The objective of the firm is to maximize sales net returns, which may happen if a product is purchased from the online channel prior to physical evaluation. Most existing work on optimizing the store assortment focuses on a single channel and ignores channel interactions: consumers may visit the offline store and subsequently purchase, possibly different products, from the online channel. We model channel interactions by allowing consumer preferences to change based on the set of product features examined during the store visit. The offline channel then becomes an ‘information’ channel, in addition to a sales channel, for the firm. Our framework allows us to naturally model product returns. We allow a consumer to return a product if it was ordered online, but turned out to be less preferred than the outside alternative.

For the proposed model, we propose techniques to estimate model parameters and solve the assortment decision problem. For the purposes of model estimation and validation, we conducted a study consisting of two within-subject conjoint tasks – one online and one with physical prototypes – with Timbuk2 messenger bags in order to measure the impact of physical evaluation on preferences. We find that physical evaluation leads to a statistically significant difference in preferences for 45% of participants. Next, we consider the assortment decision with and without capacity constraints and with and without returns. We derive theoretical results on the structure of the sales maximizing offline assortment under the multinomial logit (MNL), variants of the nested logit (NL), and Gaussian Mixture of the MNL models. Finally, we demonstrate the robustness of our techniques by applying them to the data from our conjoint study to determine the optimal subset of Timbuk2 messenger bag to carry in the offline stores. Our results show gains of 45% without returns and 5% with returns in net sales from accounting for channel interactions.

Key words: Assortment optimization, multichannel, choice modeling, conjoint, product returns
1. Introduction

Most models of consumer purchase process that are used in operations focus on purchases from a single channel, typically the offline brick-and-mortar store (Kök and Fisher 2007). The proliferation of multiple channels, such as online and mobile, has created a need to model consumer switching between different channels during the purchase process. Even when shopping for medium-priced items such as backpacks, furniture, or electronics, many consumers visit both the store’s web page and the brick and mortar store. Despite the higher variety and convenience that the online channel offers, consumers continue to visit offline brick-and-mortar stores because they may have a hard time judging the value of certain attributes online. For example, consider a consumer shopping for a messenger bag at Timbuk2. He may be certain about his utility for the color Black, but uncertain about his utility for the colors Red and Blue. The consumer can learn about the available attributes from the store web page and form beliefs about his values for the attributes. However, he may need to physically examine a product with a particular attribute in the brick and mortar store to learn that attribute’s true value.

Now consider the decision faced by Timbuk2 of which bags to offer in the offline stores. The objective is to maximize expected sales across both offline and online channels. The firm produces bags that vary on several attributes: suppose it bags come in Black, Red, and Blue exterior colors, can be Large or Small, and can have a Laptop Compartment or a Divider for Files inside. Timbuk2 offers most of the attribute combinations online but their physical stores have a much smaller selection due to the resource and space constraints of the offline channel. Examples of other firms that face a similar decision are Crate and Barrel and West Elm in the furniture category, and auto makers such as Mini that allow customers to customize their vehicles through their website. Even predating the internet, firms like LL Bean offered all their products through a catalogue, but had only a subset in their stores.

This leads to the assortment optimization decision, which is a canonical problem in operations studied in a single channel setting. The main challenge in solving this two-channel assortment optimization problem is to account for the interaction between the two channels. Due to the presence of the online channel, consumers are not constrained by the in-store assortment, but rather use it as an information source to help them make a choice. The offline assortment impacts the purchase decision because it is common for customers to visit an offline store, examine the offered products, and then purchase from the online channel a product that may be different from any of the products offered offline. In the above example, if the only bag on display at the physical store is a Large Blue bag, a customer may examine it, determine that she likes the Blue color, and

\[\text{http://www.timbuk2.com/customizer}\]
purchase a Small Blue bag from the online store. Because the firm’s products share attributes, the physical evaluation of one product, the Large Blue bag, results in a change in the utility of other products, like the Small Blue bag. Thus, the two channels impact one another. In order to capture such interactions between channels, we need to understand how the offline assortment influences the consumers’ purchase decision.

With this motivation, we consider the following specific problem. A firm sells products through two channels, online and offline, with the objective of maximizing sales across both channels. The online channel is assumed to have a wider assortment so that the offline assortment is a subset of the online assortment. Products are multi-attribute so that each product can be represented by a vector in the attribute space. Customers’ valuation of certain product attributes may change upon physical evaluation of a product with the attribute. We focus on the segment of customers who are indifferent between purchasing online or offline. In general, customers may exhibit strong channel preferences by visiting either only the online or only the offline channel. The segment of customers who do not visit the offline channel do not impact the offline assortment decision and can be safely ignored for the purposes of the decision. Conversely, the segment of customers who visit only the offline channel do affect our assortment decision, and for them the decision problem reduces to the traditional single channel assortment optimization problem. Instead of dealing with the mixture of the two segments, we focus only on the customers who visit both channels. Our solution would be a good approximation to cases in which most of the customers visit both channels. More generally, our solution would provide the starting point to a determining the assortment that appeals to both segments (refer to Section 7 for further discussion of this point).

**Overview of the Model.** We consider a firm that is selling products through an online and an offline channel. There is a universe of \( n \) products \( N = \{1, 2, 3, \ldots, n\} \), which the firm carries online. Let 0 denote the no-purchase alternative. The firm must determine the subset \( M \subseteq N \) of products to carry in the offline store. The objective of the firm is to maximize the expected sales net returns, if any, across both channels. We assume consumers are utility maximizing: each customer assigns a utility value to each product and chooses to purchase the product (including the no purchase option) with the maximum utility.

In order to determine the optimal offline assortment, we need to model the impact of the offline assortment on the purchase behavior of the customer. To that end, we allow the utility to depend on the subset of products that were physically evaluated. Let \( M \) be the offline assortment and assume that the customer physically evaluates all the products in \( M \) upon visiting the offline store. Then, \( u_i(M) \) denotes the utility assigned to product \( i \) given offline assortment \( M \). The dependence of utilities on the offline assortment may be thought of as resulting from information obtained...
through physical evaluation of the products. Note that the utilities of products not in $M$ are allowed to change as a result of offline store visit. Because the customer may still purchase any product from the entire universe (available online) and is not constrained by the products $M$, he will purchase $\arg \max_{i \in N \cup \{0\}} u_i(M)$ after the offline store visit. This makes the problem different from the standard assortment optimization problems where the offered assortment is a constraint for the customer.

We use a linear-in-parameter utility specification, so that utility of product $i$ is $\beta^T x_i + \epsilon_i$, where $\beta \in \mathbb{R}^J$ is the attribute part-worth vector; $\epsilon_i$ captures the random component of the utility; and $\beta^T x_i$ – defined as $\sum_{j \in J} \beta_j x_{ij}$ – captures the deterministic component of the utility$^2$. We let the parameters of the utility function depend on the offline assortment $M$. Specifically, if $S_M$ is the set of all features of the products in $M$, then the partworth vector depends on $S_M$ as follows:

$$\beta^M_j = \begin{cases} w_{1j}, & \text{if } j \notin S_M, \\ w_{2j}, & \text{if } j \in S_M, \end{cases}$$

where we use the notation $\beta^M$ to explicitly capture the dependence of the partworth vector on $M$. It follows from our notation that if there is no offline store, then $S_M = \emptyset$ and $\beta^\emptyset = w_1$. Note that by allowing the parameters of the utility function to depend on the offline assortment, we capture changes to the utilities of products that were not physically evaluated. Our model framework is itself more general: if the utility function was coming from a parametric family $U(\Theta) = \{u(\cdot; \theta) : \theta \in \Theta\}$, then we would allow the parameter vector $\theta$ to be a function of the offline assortment.

Finally, a key feature of our modeling framework is that it allows us to endogenously model product returns. A customer may return a product if he realizes that the post-evaluation utility of the product is less than that of the outside option, which may happen when the product is ordered online without prior physical evaluation. Therefore, conditional on being purchased, an item $i$ is returned if $u_i(w_2) < u_0$, where $u_0$ is the utility of the no purchase option. If we allow returns, then the firm should maximize sales net returns.

We provide a detailed discussion of the model assumptions in Section 2.1.

**Overview of the results.** In the context of the above model, we answer the questions of model validation, parameter estimation, and assortment optimization. A broad overview of our results is as follows.

1. **Model validation and parameter estimation.** We use preference elicitation tools to provide evidence for the phenomenon that attribute partworths change upon physical evaluation and demonstrate how the parameters of our model can be estimated. We conducted a conjoint study in

$^2$ We suppose that any customer-level heterogeneity is captured by the attribute part-worth vector $\beta$. 

which participants were asked to rate Timbuk2 messenger bags. Conjoint analysis is a method for quantitative preference measurement that is widely used by firms when conducting marketing research. Conjoint methods are used when the goal is to estimate individual level preferences for multi-attribute products. Because our model requires fine-grained information on how consumer preferences change upon physically evaluating a set of products, we conducted two within-subject conjoints: one evaluating product descriptions on a computer followed by the other evaluating physical products. The need for fine-grained preference information also makes it challenging to use secondary data such as purchase transactions, which are popular for choice model estimation in operations (Farias et al. 2013).

We find that at the population level, the partworths of three features (a type of exterior design, size of the bag, and water bottle pocket in the bag) are statistically significantly different (at $p < 0.05$) pre and post physical evaluation of the bags (see Table 2). Three more features (another type of exterior design, price, and strap pad) exhibit statistically significant differences at $p < 0.1$. In addition, at the individual level, the participant models pre and post physical evaluation were statistically significantly different (based on an F-test) for 45% (28 out of 62) and 48% (30 out of 62) of the participants at $p < 0.05$ and $p < 0.1$ respectively. The results provide evidence that customers update their decision criteria upon physical evaluation.

The key question now is which partworths the customer will use for making purchases online after the store visit. Would the customer revert back to the online partworths or would the offline partworths persist? In order to answer this question, we asked another group of participants to complete the tasks in reverse order: first the physical evaluation task, followed by the online evaluation task. For this group, our individual-level analysis shows that the models are statistically significantly different ($p < 0.05$) for none of the participants (0 out of 20). The result from this second group of participants provides evidence that the partworth of a feature used for the purchase decision depends only on whether the customer has been exposed to the feature in a physical product. It does not depend on the channel in which the purchase decision is made. Once the customer has been exposed to the feature, he will apply the new partworths to both his online and offline purchasing decisions. This simplifies our optimization problem, because we do not have to take into account in which order the customer visits the two channels.

2. Assortment Optimization. The assortment optimization problem that arises in the context of our model is computationally challenging to solve in general. Hence, taking a cue from the existing body of literature on assortment optimization, we analyze several important special cases of increasing complexity. We consider the cases with and without returns separately.

3 At $p < 0.1$ significance, the models were statistically significantly different for 10% (2 out of 20) of the participants.
(a) Single-class MNL, uncapacitated, without returns. We first consider the case when customers make choices according to a single-class MNL model and there is no capacity constraint. In this case, we show that it is optimal to offer all the under-valued features and “hide” or not offer the over-valued features in the offline store (see Theorem 4.1). We say feature $j$ is under-valued if the pre physical evaluation partworth $w_{1j}$ is strictly less than the post physical evaluation partworth $w_{2j}$. Conversely, we say feature $j$ is over-valued if the pre physical evaluation partworth $w_{1j}$ is larger or equal to the post physical evaluation partworth $w_{2j}$. In other words, the firm should be offering more information to customers on features they tend to under-value. Multiple assortments can achieve the desired combination (exposure to under-valued and hiding of over-valued features) of features in the store and the smallest (in terms of cardinality) such assortment can be constructed in an efficient manner.

(b) Single-class MNL, capacitated, without returns. We next consider the case in which the offline store of the firm is constrained by a fixed capacity, measured in terms of the size of the offer set, and customers make choices according to an MNL model. The capacity constraint affects how many under-valued levels of each attribute the customer can be exposed to. Specifically, it is optimal to offer at most $C$ of the most significant levels for each attribute $k$, where we measure the significance of an attribute level by the magnitude of its discrepancy $\exp(w_{2k}) - \exp(w_{1k})$ (see Theorem 4.2).

(c) Nested logit, capacitated, without returns. Our results for the MNL model extend to the case when customers choose according to the following variant of the nested logit model: products are partitioned into two nests with one nest containing the ‘no-purchase’ option and the other nest containing all the other products. We show that the result for the capacitated, single-class MNL model extends (see Theorem 4.4): for each attribute $k$, the optimal assortment offers the top at most $C$ levels according to $\exp(w_{2k}) - \exp(w_{1k})$.

(d) Gaussian mixture of MNL models, uncapacitated, without returns. The results for the uncapacitated, single-class MNL model extend to the case when consumers are heterogeneous in their pre and post physical evaluation part-worths and sample $w_1$ and $w_2$ according to a multivariate Gaussian distribution with mean vectors $\nu_1$ and $\nu_2$ and a diagonal covariance matrix. We show that when there is no capacity constraint, it is optimal to offer under-valued features, for which $\nu_{2j} \geq \nu_{1j}$, and “hide” over-valued features, for which $\nu_{2j} < \nu_{1j}$ (see Theorem 4.5).

(e) Uncapacitated, with returns: We then consider the setting in which product returns are allowed. We assume that a customer returns the product if the post-evaluation utility of a product is less than that of the outside option, which may happen if the product is ordered online without prior physical evaluation. When there is no capacity constraint, we show the following strong result: irrespective of the underlying choice structure, it is optimal for the firm to offer all of the features
in the offline store. In other words, it is in the best interest of the firm to provide full information to the customer (see Theorem 5.1).

The assortment optimization problem becomes significantly more complicated and we are not able to theoretically derive a complete characterization of the optimal in-store assortment in the more general settings. As a result, we propose a Greedy heuristic and demonstrate it accuracy through a simulation study.

3. Numerical study. We demonstrate the effectiveness of our approach through two simulation studies and a case study with the data collected on Timbuk2 messenger bags. The goal of the simulation studies is three-fold: (a) demonstrate the gain in net sales a firm can expect from accounting for channel interactions; (b) demonstrate the accuracy of the Greedy algorithm; and (c) gain insight into the firm’s operational decision on how much offline store capacity to invest in. We find that the Greedy heuristic provides good approximations to the optimal solutions (an average optimality gap of 2%) and the loss from ignoring channel interactions can be significant (up to 16% loss in sales). In addition, the marginal value of offline store capacity (measured in terms of the increase in net sales from an additional unit of capacity) crucially depends on the ‘discrepancy’ between the partworths before and after physical evaluation. As expected, capacity becomes more valuable to the firm as the ‘discrepancy’ increases (see Figure 5 and Figure 6 in Section 6). The results from our case study provide further evidence to the finding that sales gains from accounting for channel interactions are substantial: 45% without returns and 5% with returns on an average. We also find that ignoring channel interactions can decrease net sales (with and without returns) as we increase the offline store capacity, underscoring the significance of accounting for channel interactions.

Related Work. Our paper falls at the intersection of marketing and operations and builds on work in both areas. More specifically, the work is related to four streams of literature: assortment optimization, product returns, multichannel retailing, and preference elicitation.

The problem of assortment optimization has been a fundamental problem of interest to the operations community for the past several decades (see Jagabathula (2014), Rusmevichientong et al. (2010), Gallego and Topaloglu (2012), Davis et al. (2014), Rusmevichientong et al. (2014), Bront et al. (2009)). The canonical problem is to find the subset of products that yields the maximum expected revenues when customers exhibit substitution behavior. The focus of the literature has been on optimization: how to solve the decision problem efficiently in the context of well-studied choice models such as the multinomial logit (MNL), nested logit (NL), and the mixed logit models. Gaur and Honhon (2006) study the assortment and inventory problem under the locational choice model. Honhon et al. (2010) consider the impact of stock-out based substitution on assortment and
inventory planning. The rich work in this area has resulted in efficient algorithms for assortment optimization along with theoretical results on computational hardness and theoretical guarantees. The focus of most of the work has been on a retailer selling frequently purchased products through a single-channel; the canonical example is a super market selling groceries. We deviate from this stream of existing literature by focusing on firms that both design and sell products through multiple retail channels (examples include Timbuk2, Crate & Barrel, West Elm, Bonobos, etc.). The products may be infrequently purchased so that the consumers may be novices and need to get familiar with the product. The products also vary on multiple attributes, so that in-store assortment changes from being a constraint to being a channel for information on the attributes. We capture the ‘information value’ of the in-store assortment through a novel model that operates in the feature space, and then study the assortment optimization question in the context of the new model. We show that the new modeling approach results in optimization problems that different in the solution and problem structure from the ones traditionally studied. We develop new methodology for solving these optimization problems. More importantly, our work lays the foundation for a rich set of optimization problems that are becoming increasingly important in practice.

A key feature of our modeling framework is that it naturally allows us to model product returns. Product returns have become increasingly common, especially in the context of a firm that also sells through an online channel. The value of products that US consumers return to retailers exceeds $100 billion each year (see Shear et al. (2002), Su (2009)). Existing work on consumer behavior in operations management has studied consumer returns policies when a monopolistic firm is selling a single product to a population of customers who may be strategic in their purchase decisions. Specifically, consumers are modeled as facing valuation uncertainty and the valuations are realized only after purchase. The valuation uncertainty captures the ‘fit risk’ that consumers face due to potential product misfit if the product valuation turns out to be low (see Su (2009)). Several operational policies have been considered for a firm to compensate a customer for bearing the ‘fit risk’ and thus encourage purchases: advance purchase discounts (Dana 1998, Xie and Shughan 2001), call options on capacity (Gallego and Sahin 2006), reservations (Alexandrov and Lariviere 2012). Along similar lines, returns policies can serve as an insurance mechanism for the consumer. Existing work has focused on several aspects of consumer returns policies such as preventing inappropriate returns from customers who have no intention of keeping their purchases (such as buying a dress before a party) (Hess et al. 1996), distinction between “no questions asked” and “verifiable problems only” (Chu et al. 1998), optimal level of “hassle” on returns policies (Davis et al. 1998), optimal level of logistics investment (Yalabik et al. 2005), etc. Finally, there is rich literature on studying the operational impact of a firms returns policies (such as full vs partial refunds); see Su (2009) and the references therein. The key distinction of the above work from our work is that we focus
on a firm is selling multiple products instead of a single product. Further, our focus is on models that may be estimated from consumer preference data with the goal of making tactical assortment decisions.

The multichannel retailing literature in marketing and information systems has focused on the interaction between online and offline channels. Most of the focus of this literature has focused on the competition between online and brick and mortar retailers and the implications for pricing decisions. For example, Forman et al. (2009) study customers’ propensity to check prices for an identical item they are looking to purchase at multiple retailers, some of which are brick and mortar, and others are online. Brynjolfsson et al. (2009) demonstrate the online retailers compete with brick and mortar retailers significantly more on niche products than on mainstream products. Mehra et al. (2013) model the showrooiming phenomenon when the two channels compete with each other. In their framework, the customer receives only a noisy signal about a product’s quality online, and this uncertainty is resolved if the consumer visits a brick and mortar store to examine the product, at some cost. They demonstrate the implications for equilibrium pricing. Note that these papers explore the option of the customer to buy the exact same product online versus offline. We differ from this literature in that in our model (1) the firm tries to optimize sales across the two channels, (2) the customer may examine a product in the physical store and then choose to purchase online a product not offered in the store, and (3) the focus is on product assortment rather than pricing decisions with the assumption that the product offerings are differentiated and the prices online and offline are matched so that a differential pricing decision becomes less important.

Finally, we build on existing methods of preference elicitation in marketing to measure online and offline preferences. We used conjoint analysis (e.g. Green and Srinivasan (1978)) to elicit participants’ individual level preferences for each product attribute. We use an incentive aligned mechanism similar to the one proposed by Ding (2007) to improve the external validity of the measurement. The offline task of our study is similar to the work of Luo et al. (2008) which used physical prototypes to incorporate subjective characteristics into conjoint. We contribute to this literature by measuring within-subject differences between online and offline partworths. Additionally, we build on the model proposed by Dzyabura (2014) of attribute weights changing in response to product evaluation, but there, preferences are measured using an unstructured direct elicitation task, rather than conjoint analysis.

2. Model and Problem Formulation

In this section, we present a formal description of our model and state the precise problem we aim to solve. Suppose $N$ is the product universe consisting of $n$ products indexed $1, 2, \ldots, n$. The ‘outside option’ or the ‘no purchase option’ is denoted by 0. The products differ on multiple
attributes: each product is described by \( K \) attributes with attribute \( k \) having \( L_k \) levels. As an example, products may differ on Color and Size, which we term attributes. The attribute Color may take one of the three values: Black, Blue, and Red; hence, it has three levels. Similarly, the attribute Size may take one of the two values: Small and Large; so, it has two levels. We refer to each attribute-level combination a ‘feature’. Thus, in the example above, the tuple (Color, Black) is a feature. It follows from our notation that there are a total of \( J = \sum_{k=1}^{K} L_k \) possible features. Assuming an arbitrary indexing, we let \( J \) denote the set of features \{1, 2, \ldots, J\}.

We encode product attributes using a boolean vector, so that product \( i \) is represented by the feature vector \( x_i \in \{0, 1\}^J \) with \( x_{ij} = 1 \) if and only if feature \( j \) is present in product \( i \). We sometimes write \( x_{ik\ell} \) to refer to \( x_{ij} \) when feature \( j \) refers to the attribute-level combination \((k, \ell)\). We use dummy encoding, so we require every feasible product to possess exactly one of the \( L_k \) levels for each attribute \( k \). In other words, we assume that the \( L_k \) levels include all possible values that attribute \( k \) can take in the collection of products of our interest. More precisely, the set of feasible feature vectors is defined by \( \{ x \in \{0, 1\}^J : \sum_{\ell=1}^{L_k} x_{k\ell} = 1 \text{ for } k = 1, 2, \ldots, K \} \).

Under the above setting, we consider a firm selling products through an online and an offline channel. We assume that the selection of products carried by the firm in the offline store is a subset of the selection carried online. We focus on the universe of products carried online by the firm and thus let \( N \) denote the selection of products carried online. Given this, our goal is to solve the following decision problem: determine the assortment (or subset) of products to offer in an offline store with the objective of maximizing expected sales both from the offline and online channels. The main challenge in solving the above decision problem is to account for any channel interaction effects: customers who visit the offline store may eventually purchase from the online channel. Our goal is to study the impact of channel interaction effects on the offline assortment decision, for which we focus on the segment of customers who are indifferent between purchasing online or offline and always visit the offline store before making the purchase decision. In general, customers may exhibit strong channel preferences by visiting either only the online or only the offline channel. The segment of customers who do not visit the offline store do not impact the offline assortment decision and can be safely ignored for the purposes of the decision. Conversely, the segment of customers who visit only the offline channel do affect our assortment decision, and for them the decision problem reduces to the traditional single channel assortment optimization problem. Instead of dealing with the mixture of the two segments, we focus only on the customers who visit both the channels. Our solution would be a good approximation to cases in which most of the customers visit both channels. More generally, our solution would provide the starting point to a determining the assortment that appeals to both segments (see Section 8 for further discussion of this point).
We model the impact of the offline store assortment on consumer purchase decisions using the Random Utility Maximization (RUM) framework: consumers are utility maximizers and in each choice instance, the consumer assigns utility $u_i$ to product $i$ and purchases the product with the maximum utility i.e., purchases the product $\arg \max_{i \in N \cup \{0\}} u_i$. We use a linear-in-parameter utility specification, so that utility $u_i$ is related to product attributes through $u_i = \beta^T x_i + \varepsilon_i$, where $\beta \in \mathbb{R}^J$ is the attribute part-worth vector; $\varepsilon_i$ captures the random component of the utility; and $\beta^T x_i$ – defined as $\sum_{j \in J} \beta_j x_{ij}$ – captures the deterministic component of the utility. We suppose that any customer-level heterogeneity is captured by the attribute part-worth vector $\beta$.

We model the impact of the offline assortment through changing consumer part-worths. In particular, if $M \subset N$ is the subset of products offered in the offline store, then the consumer updates his part-worths upon visiting the store to $\beta^M$ defined as

$$\beta^M_j = \begin{cases} w_{2j}, & \text{if } j \in S^M, \\ w_{1j}, & \text{otherwise,} \end{cases}$$

where $S^M$ denotes the subset $\{ j \in J : x_{ij} = 1 \text{ for some } i \in M \}$ of features that are offered in the store as part of assortment $M$. As a result, upon visiting the offline store, the consumer chooses to purchase the product with the highest utility as computed with the updated part-worths. Note that since the consumer may also purchase from the online channel after the store visit, he is not restricted to purchasing from the subset $M$ of products in the offline store. Since $N$ is the selection of products available in the online channel, the consumer purchases the product $\arg \max_{N \cup \{0\}} u_i(\beta^M)$, using the notation $u_i(\beta^M)$ to make the dependence of utility on the part-worth vector explicit.

Given the above model, the probability that a consumer makes a purchase after evaluating the subset $M$ of products in the offline store is given by

$$\phi(M) = \sum_{i \in N} \Pr_{\varepsilon} \left( u_i(\beta^M) \geq u_{i'}(\beta^M) \forall i' \in N \cup \{0\} \right) = 1 - \Pr_{\varepsilon} \left( u_0 \geq u_i(\beta^M) \forall i \in N \right).$$

The decision problem faced by the firm can then be formulated as

$$\arg \max_{M \subset M} \phi(M) = \arg \max_{M \in M} \left( 1 - \Pr_{\varepsilon} \left( u_0 \geq u_i(\beta^M) \forall i \in N \right) \right),$$

(1)

where we allow the firm to restrict the offered subset to the collection $M$ of feasible in-store assortments.

The formulation in [1] ignores any products a customer may return after purchase. Returns are common, especially when a firm sells through an online channel. Our modeling framework allows us to endogenously model product returns. Within the context of our modeling framework, a customer may return a product if he realizes that the post-evaluation utility of the product is less than that of the outside option, which may happen when the product is ordered online without prior
physical examination. If we allow returns, then the firm should maximize sales while minimizing returns. This issue of minimizing returns is an important problem for firms for which a significant fraction of sales come from the online channel. Existing work on returns (e.g., Anderson et al. (2009)) models returns as exogenous events. To the best of our knowledge, our work is one of the first to account for returns using the utility framework. As expected, allowing returns complicates the decision problem. We study the issue of returns in the next section, and focus on the problem of maximizing sales in this section.

Returns become relevant only for customers who purchase. Conditioned on the fact that the customer purchases product \( i \), he may return it if the post-evaluation utility of the product after physical evaluation upon receipt is less than the utility of the no purchase option. Since after physical evaluation of the product, the utility becomes \( u_i(w_2) \), the customer will not return the product if and only if \( u_i(w_2) \geq u_0 \). Putting everything together, the probability that a customer will purchase, but not return the product, is given by

\[
\phi_R(M) = \sum_i \Pr(\left. u_i(\beta M) \geq u_i'(\beta M) \forall i' \in N \cup \{0\} \right. \text{ and } u_i(w_2) \geq u_0),
\]

where \( \phi_R(M) \) captures the expected sales net returns. The decision problem of determining the subset of products that maximizes the probability that a customer will purchase, but not return, may be expressed as

\[
\arg \max_{M \in M} \phi_R(M) = \arg \max_{M \in M} \sum_i \Pr(\left. u_i(\beta M) \geq u_i'(\beta M) \forall i' \in N \cup \{0\} \right. \text{ and } u_i(w_2) \geq u_0). \quad (2)
\]

Note that we have implicitly assumed above that the customer uses the same error term realization \( \varepsilon_i \) to compute both \( u_i(\beta M) \) and \( u_i(w_2) \). In other words, we are assuming that the idiosyncratic error term \( \varepsilon_i \) is sampled once and used for both the purchase and the return decision.

We justify the assumption that the customer uses the same error term both for the purchase and return decision as follows. First note that the error term typically captures the impact on utilities of all the factors that have not been explicitly modeled; in our case that would comprise all the factors not in the collection of features \( J \). So, a sufficient condition for our assumption is the following: all (or most) of the features whose part-worths may be altered by physical evaluation have been included in \( J \) and the remaining factors affecting utilities do not change from the purchase to the return decision. The sufficient condition is reasonable in practice because the collection of features \( J \) whose part-worths may be altered can be reasonably obtained through user studies. In addition, we expect any ‘pure’ idiosyncratic factors such as mood, daily weather, etc. that may change from the purchase to the return decision to have minimal impact on the decision because of our focus on products that are not pure impulse-buys, so that customers visit the offline store before making the purchase decision.
Finally, solving the decision problem in (1) in a tractable manner necessitates further assumptions on the structure of the constraint set $\mathcal{M}$ and the structure of the choice model.

We consider constraint sets $\mathcal{M}$ that are defined by capacity constraints over a full-factorial product universe. Specifically, we assume that the firm is constrained to offer at most $C$ products in the offline store. In addition, the product universe consists of all possible combinations of features i.e., the set of feature vectors $\{x \in \{0,1\}^J : \sum_{k=1}^{L_k} x_{k\ell} = 1, \text{ for } k = 1,2,\ldots,K\}$. Thus, there are a total of $n = \prod_k L_k$ products in the universe. Capacity is the most significant operational constraint that a firm faces and hence most existing work on assortment optimization studies operational decision problems with capacity constraints (Gallego and Topaloglu 2012, Rusmevichientong et al. 2010). The assumption of full-factorial product universe on the other hand may be justified as follows. The assumption is valid for many firms that both design and sell the products through their own retail channels. Timbuk2 and Crate & Barrel are two such examples. The firm may adopt a make-to-order strategy in order to avoid carrying inventory for the combinatorially many products. In addition, in cases in which certain feature combinations are not feasible, the features may be re-encoded to result in a full-factorial product universe. For example, suppose cars differ only on two attributes: Brand with two levels, Toyota and Porsche, and Body also with two levels, Sedan and Pickup Truck. Porsche does not make Pickup Trucks. So, we can encode the two attributes as a single Brand-Body attribute with three levels: Toyota-Sedan, Toyota-Pickup Truck, and Porsche-Sedan. Such re-encoding allows us to retain the full-factorial assumption.

As for the choice structure, taking a cue from the existing body of work on assortment optimization, we start with the assumption that choices happen according to a multinomial logit (MNL) model, so that the error terms $\varepsilon_i$ are i.i.d. Gumbel distributed. We then extend our results to the nested logit and Gaussian mixtures of the MNL model.

2.1. Discussion of the model

We highlight the following key aspects of our model.

Independent part-worth updates: Our specified form for $\beta^M$ implicitly relies on the assumption that physically evaluating a feature $j$, does not affect the customer’s valuation partworth of any features other than $j$. For example, evaluating a particular design pattern does not affect customer’s partworth for other patterns, or for other attributes, such as size. While this is a simplifying assumption, it is reasonable for features that capture unrelated aspects of the product, but may be limiting for other types of attributes, such as fabric and weight, or design and ease of cleaning or durability. Since the exact relationship between product attributes depends on the product category, we make the independence assumption for the analysis in this paper. The model may
be naturally extended to incorporate dependence among various attributes, but doing so would significantly complicate estimation and optimization.

**Interpreting changes in part-worths:** The key assumption underlying our model for channel interaction effects is that physically evaluating a product changes the utilities of (other) products. Since we assume linear-in-parameter specification $\beta^T x_i$ for the inherent utility of the product, it is natural to capture the change in utilities through changes in the part-worth parameter $\beta$. Note that, while it is most natural to interpret the features as individual product attributes such as Color, Size, Material, etc., the linear-in-parameter specification is more general and easily accommodate other factors (such as interaction terms). This generality of the linear-in-parameters allows us to abstract away the underlying process that leads to the change in utilities.

**Utility changes due to attribute uncertainty:** One way in which the utilities can change upon physical evaluation is through the resolution of uncertainty in product attributes. For instance, suppose the online description of a bag says ‘Red’, but even with a photograph, it may be difficult for the consumer to tell if it is a Bright Red or a Dark Red. A consumer who knows with certainty his valuations for Bright Red $\beta_B$ and Dark Red $\beta_D$ may be modeled as having a pre-evaluation part-worth that is a combination of $\beta_B$ and $\beta_D$, which upon physical evaluation changes to either $\beta_B$ or $\beta_D$ depending on whether color of the bag is Bright Red or Dark Red.

**Implications for assortment optimization:** Our assumptions that changes in utilities are captured through changes in utility part-worths and that part-worth updates are independent of each other imply that expected sales from both online and offline channels is determined only by the subset of features that are offered in the offline store, not the specific product configurations themselves. As a result, assortment optimization can be performed in the feature as opposed to the product space.

Furthermore, the optimal assortment under our model can be structurally very different from the one under the traditional approaches. For instance, the optimal assortment may include products that are not purchased by any of the consumers. Such products still generate sales by exposing consumers to new, previously undervalued, attribute levels. The following example illustrates the insight. Consider the purchase of messenger bags. For simplicity, suppose that the bags vary only on color (Black, Orange) and the type of back panel (Airmesh, Regular). Suppose there are two consumer segments, and both have an outside option with utility 1. The partworths prior to physical evaluation for both segments are all zeros, so the utilities of all products based on the initial partworths are zero resulting in no sales if there is only the online channel. However, suppose upon physical evaluation, the first segment likes Orange (post evaluation partworth +2) but dislikes
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Levels</th>
<th>Segment 1 Pre-Eval</th>
<th>Post-Eval</th>
<th>Segment 2 Pre-Eval</th>
<th>Post-Eval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exterior Design</td>
<td>Black</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Orange</td>
<td>0</td>
<td>+2</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>Back Panel</td>
<td>Regular</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Airmesh</td>
<td>0</td>
<td>-2</td>
<td>0</td>
<td>+2</td>
</tr>
</tbody>
</table>

Table 1: If capacity is equal to 1, then the unique optimal in-store assortment that captures 100% market share is an Orange bag with Airmesh Back panel, even though this not purchased by any consumers.

Airmesh (post evaluation partworth -2). Conversely, the second segment of consumers dislikes Orange (post evaluation partworth -2) and likes Airmesh (post evaluation partworth +2). These parameter values are summarized in Table 1. Assuming that the retailer can offer only one product in the offline store, it is easy to see that the optimal offering is an Orange bag with an Airmesh back panel. With this offering the online channel sees no sales, but the online channel captures the entire market. Note that with only the offline channel, the optimal offering is different and can capture only one of the two segments.

3. Model validation and parameter estimation: Conjoint Study

In this section, we focus on model validation and estimation issues. Our goal is to provide evidence for the phenomenon that attribute partworths change upon physical evaluation and demonstrate how the parameters of our model can be estimated. For that, we conducted a marketing conjoint study to collect primary data on consumer preferences from a sample of participants. We present the details of the study and the results we obtained. The results from our conjoint study (a) provide strong empirical evidence to support the assumptions of our model and (b) illustrate the key steps a practitioner may take in order to implement our model.

We rely on primary data instead of secondary data for fitting our model. Traditionally, choice modeling in Operations Management (OM) has relied on secondary data for model estimation. Existing works on model estimation have relied on secondary data, typically in the form of purchase transactions that are either aggregated over all customers (for e.g. see Farias et al., 2013; Vulcano et al., 2010, Vulcano et al., 2012) or collected at the individual level (such as IRI panel data; e.g. see Guadagni and Little, 1983, Gupta, 1988, Jagabathula and Vulcano, 2014, Lattin and Bucklin, 1989). The use of secondary data in the form of purchase transactions is popular because secondary data are relatively easy to collect (especially with the ready availability of point-of-sale (POS) systems) and capture ‘revealed preferences’ (as opposed to stated preferences) of customers in organic settings. As a result, the data are typically predictive of the purchase behavior of...
customers. Of course, such secondary data can often be “noisy” due to the presence of endogenous factors (such as stock outs and promotions) and censored observations (see [Vulcano et al., 2012] and the references there in).

Despite the above advantages of secondary data, they are unsuitable for our purposes because they lack key information. Specifically, a key component of our model is how consumer preferences change upon physically evaluating a set of products. In order to capture these changes, the data must contain information linking purchase behavior of the same consumer across different channels. In particular, we require data that: (a) capture individual customer visits and purchases to both the physical and the online channel, and (b) contain “many” purchases for each customer for reliable estimation of individual preferences. Unfortunately, in many practical applications, purchase transactions are only available separately for each sales channel. As a result, they do not capture information linking the purchase behavior of the same customer across multiple channels. In addition, many practical applications with potentially high discrepancies in utilities before and after physical evaluation correspond to infrequently purchased products, such as furniture, cars, luggage, etc. resulting in only a few observations for each customer.

To overcome the above challenges, we turn to primary data and collect preference data from a group of participants. The primary data collection tool we use is Conjoint analysis – a common approach taken in marketing research. Conjoint analysis was introduced to marketing by [Green and Rao, 1971] and has since been the method of choice for quantitative preference measurement. It is typically applied to help firms design products and/or predict market shares [Netzer et al., 2008]. It also has the advantage of being widely used by practitioners. However, traditional conjoint techniques do not directly apply to our setting (as discussed below), so we adapt them to our problem. To the best of our knowledge, this paper is the first to apply conjoint methods to problems involving assortment optimization.

Broadly, the study involves measuring participants’ preferences before and after physical evaluation. Next, we describe the details of the study, and present the results and conclusions.

3.1. Design of the study
A conjoint analysis allows us to collect detailed preference information for a group of participants. In a typical conjoint study, the experimenter shows the participants a set of products, and the participants provide their evaluations by rating, ranking, or choosing from the set of products shown. These evaluations are then used to back out individual-level attribute partworths by fitting linear or choice-based models. Typically, products are shown to participants either on a computer, along with descriptions of the features of each of the products. Some recent exceptions that use physical prototypes for measuring participants’ preferences are [Luo et al., 2008] and [She and MacDonald]
A conjoint task with product descriptions shown on a computer allows us to estimate utility partworths of participants when they purchase online. Similarly, a task in which participants evaluate physical prototypes allows us to estimate their utility partworths upon physical evaluation. Since our goal is to estimate the changes in utility partworths before and after physical evaluation, we adapt the traditional conjoint study by conducting two within-subject conjoints: one evaluating product descriptions on a computer followed by the other evaluating physical products.

Conducting the study requires us to specify the following key components: a universe of products that vary of a set of attributes of interest, a ‘study design’ consisting of a subset of products from the universe for which participant evaluations are collected, and the details of the tasks presented to each of the participants. Next, we specify these details.

**Product and attributes.** We chose Timbuk2 messenger bags for our study. These bags are suitable for our purposes because they are fully configurable, vary on discrete attributes, and are small enough for us to be able run the study with physical bags in the behavioral lab. In addition, Timbuk2 squarely fits the setting being considered in this paper: the firm must decide which subset of products to offer in its brick-and-mortar stores given that their website offers a “very large” selection of fully configurable products made to order. Timbuk2 also satisfies the three key assumptions of our model. First, the Timbuk2 brand differentiates itself from other brands in the category such that it does not face direct substitution. Second, it targets young customers, most of whom tend to be willing to purchase online after visiting a physical store, which is consistent with our channel indifference assumption. Third, the online selection is strictly broader than the offline selection.

Timbuk2’s custom order option allowed us to create bag configurations for a balanced study design. Specifically, we chose six bag features for the conjoint study:

- Exterior design: Black, Blue, Reflective, Colorful (illustrated in Figure 1)
- Size: Small (10 × 19 × 14 in), Large (12 × 22 × 15 in)
- Price: $120, $140, $160, $180
- Strap Pad: Yes, No
- Water bottle pocket: Yes, No
- Interior compartments: Empty bucket with no dividers, Divider for files, Padded laptop compartment

The above features comprise the set of configurable features on Timbuk2’s website, suggesting that they are important drivers of customer purchase decisions.

**Product configurations.** We now specify the subset of products from the above universe that were included in the study design. Note that our product universe is described by six design
variables or attributes. The universe consists of a total of $4^2 \times 3 \times 2^3 = 384$ (4 levels of exterior design, 4 levels of price, 3 levels of interior compartments, and 2 levels each of size, strap pad, and water bottle pocket) feature combinations. Since it is infeasible to ask each participant to evaluate all the 384 options, we must pick an ‘optimal’ study design in order to minimize the number of responses per participant that are needed to reliably estimate utility partworths. Existing work (e.g. Huber and Zwerina (1996), Toubia and Hauser (2007)) has proposed several techniques to pick optimal designs. All of these techniques rely on optimizing certain statistical criteria. Of the several existing optimality criteria, we choose the widely used ‘D-optimal’ design criterion. Specifically, let $X$ denote the ‘design matrix’, in which each row corresponds to a product that is included in the study design, each column corresponds to a feature (or attribute-level combination), and an element in row $i$ and column $j$ is the value of feature $j$ for product $i$. Then, the D-optimal design chooses the design matrix $X$ that maximizes the determinant of the information matrix $X^\top X$, thus minimizing a measure of the variance of the parameter estimates (see Huber and Zwerina (1996), Toubia and Hauser (2007) for details). The quality of a D-optimal design is measured using the D-efficiency metric, equal to the determinant of the information matrix $X^\top X$ of the chosen.

Figure 1  Images of the 4 Exterior design options that were used. These are the same images that were used in the actual task.
design matrix normalized so that the full-factorial design matrix (consisting of all possible feature combinations) has a D-efficiency of 1. Since the determinant of the information matrix is being maximized, the higher the D-efficiency, the better the design.

For our study, we created a D-optimal design consisting of 20 bags. Our design had a D-efficiency metric of 0.97, which is considered sufficiently high for reliable estimation. The 20 bags that were included in the study are presented in Table 5 in Appendix B.

Participants. We recruited participants from a university subject pool. They were paid $7 for their participation and entered in a raffle to win an incentive aligned prize (discussed below). Each participant completed two tasks in the same season, which took about 25 minutes on average. In total, 82 participants completed both tasks. We used a within subjects design in which each participant completed both the web-based and paper-and-pencil task, which are described below.

To ensure incentive compatibility, participants were told by the experimenter that they would be entered in a raffle for a chance to win a free messenger bag. Were they to win, their prize would be a bag that was configured to their preferences, which the researchers would infer from the responses they provided in the study. This chance of winning a bag provides incentive to participants to take the task seriously and respond truthfully with respect to their preferences Ding (2007). We followed the instructions used by Ding et al. (2011) and told participants that, were they to win, they would be given a messenger bag plus cash, which represented $180. The cash component should eliminate any incentive for the participants to provide higher ratings for more expensive items, in order to win a more expensive prize.

Study procedure: description of the tasks. We invited participants to complete a two-part task: a web-based conjoint survey, and a paper-and-pencil survey providing evaluations of physical products. Both parts were ratings-based conjoint tasks, in which participants rate each bag individually with respect to how likely they would be to purchase the bag. A five point scale was used to rate the bags: Definitely not buy, Probably not buy, May or may not buy, Probably buy, Definitely buy. We chose to conduct a ratings-based conjoint rather the more common choice-based conjoint. In a choice-based conjoint, participants are presented with a series of choice sets and asked to choose one product from each set. Conducting such a choice-based conjoint is logistically much harder when the choice tasks involve evaluating physical products. Hence, for the purposes of logistical simplicity, we carried our the ratings-based conjoint study.

Online task. The web-based conjoint was conducted using Sawtooth Softwares CVA tool. The top image in Figure 2 provides a screen shot of the task. Of the six features, five were represented with text, and one, Exterior design, was represented with an image.
The task proceeded as follows:

1. The experimenter informed the respondent that there would be two parts, one on the computer, one in an adjacent room with paper-and-pencil, and described the incentive-aligned prize lottery.

2. Initial screens ensured privacy and described the basic study.

3. The next six screens introduced the features one at a time and included a brief description of each feature.

4. Participants rated one bag as a warm up exercise. They were informed that this response would be discarded.

5. Participants then provided ratings for the 20 bags on the five-point scale described above.

**Paper and Pencil Study.** After completing the online study, participants were escorted to a different room, where the same set of 20 bags were laid out on a conference room table, as shown in Figure 3. The prices were displayed on stickers on a tag attached to the bag. Each bag had an index card next to it displaying a number indexing the bag, and the bags were laid out in order 1 through 20. All participants saw the bags in the same order. The experimenter walked the participant through all the bag features, showing each feature on a sample bag. Participants were then asked to complete the paper- and-pencil survey, in which they provided ratings for each of the bags (see Figure 3).
The experimenter asked them to take their time and examine all the bags and rate them with respect to how likely they would be to purchase such a bag. Participants were also reminded of the incentive aligned lottery.

Figure 3 Photo of the actual task faced by respondents. The 20 bags were laid out on a conference room table and labeled 1 through 20. Participants could look at each bag and provided their evaluations on a paper and pencil survey.

3.2. Results

We analyzed participant responses to investigate differences at the population and the individual levels. We first present results from the population level model, then the individual level model.

**Population-level model.** At the population level, we fit the following linear model separately to the two datasets: online and paper-and-pencil.

\[ r_{pi} = \alpha_z + \sum_j \beta_j x_{ij} + \varepsilon_{pi}, \]

where \( r_{pi} \) is the rating provided by participant \( p \) for bag \( i \) and \( z \) denotes the dataset – online or paper and pencil. We assigned the ratings from 1 to 5 corresponding to the response scale Definitely not buy, Probably not buy, May or may not buy, Probably buy, Definitely buy, respectively. Price is modeled as a single continuous attribute with one coefficient and the remaining attributes are modeled as categorical variables. We used dummy coding for each of the categorical variables and set the levels Black, Small, No strap pad, No waterbottle pocket, and Empty bucket with no dividers to zero for attributes Exterior design, Size, Strap pad, Waterbottle pocket, and Interior compartment, respectively. Note that since the attributes are modeled as categorical variables, the coefficients corresponding to the levels set to zero are not identified. Their combined effect is included in the intercept term. In total, there are 10 coefficients and one intercept term estimated for each dataset. We term the coefficients estimated from the online dataset “Pre-Eval” to convey that they correspond to utility partworths before physical evaluation of the bags. Similarly, we term the coefficients estimated from the paper and pencil dataset “Post-Eval” to convey that they correspond to utility partworths estimated after physical evaluation of the bags. The results of the
Table 2 There are statistically significant differences between pre-evaluation (online) and post-evaluation (paper & pencil) partworths for several of attribute-level combinations. Results are based on the 62 participants who completed the online task first, followed by the offline task. The levels with no coefficients were set to zero in dummy encoding.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Level</th>
<th>Pre-eval</th>
<th>Post-eval</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exterior Design</td>
<td>Reflective</td>
<td>−0.23**</td>
<td>−0.50***</td>
<td>+0.27*</td>
</tr>
<tr>
<td></td>
<td>Colorful</td>
<td>−1.05***</td>
<td>−0.68***</td>
<td>+0.37**</td>
</tr>
<tr>
<td></td>
<td>Blue</td>
<td>−0.08</td>
<td>−0.06</td>
<td>+0.02</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>Large</td>
<td>0.30***</td>
<td>−0.21**</td>
<td>−0.51***</td>
</tr>
<tr>
<td></td>
<td>Small</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>$120, $140, $160, $180</td>
<td>−0.27***</td>
<td>−0.20***</td>
<td>+0.07*</td>
</tr>
<tr>
<td>Strap pad</td>
<td>Yes</td>
<td>0.49***</td>
<td>0.29**</td>
<td>−0.20*</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water bottle pocket</td>
<td>Yes</td>
<td>0.45***</td>
<td>0.20***</td>
<td>−0.25**</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interior Compartments</td>
<td>Divider for files</td>
<td>0.42***</td>
<td>0.50***</td>
<td>+0.08</td>
</tr>
<tr>
<td></td>
<td>Crater laptop sleeve</td>
<td>0.71***</td>
<td>0.85***</td>
<td>+0.14</td>
</tr>
<tr>
<td></td>
<td>Empty bucket/no dividers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td>2.57</td>
<td>2.58</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *p < 0.1, **p < 0.05, ***p < 0.001

We make the following observations. Focusing on the partworths in the column “Pre-Eval,” it is clear that as expected, participants have a negative price coefficient. Furthermore, all else being equal, participants on an average prefer Black to the other Exterior Designs. For example, participants on an average gave 1.05 points less rating to Colorful Exterior Design than Black. In the Pre-Eval stage, participants also prefer Large to Small, Water Bottle Pocket to no Water Bottle Pocket, and a Strap Pad to no Strap Pad. The partworths under the “Post-Eval” column can be interpreted in a similar fashion.

Difference between pre-evaluation and post-evaluation partworths. We then examined the differences between the pre-evaluation and post-evaluation partworths first at the population level, and then at the individual level.
The last column in Table 2 contains the discrepancy between the “Pre-Eval” and the “Post-Eval” partworth for each feature. In order to determine the features for which the differences between pre and post evaluation partworths are statistically significant at the population level, we fit the following model to the data pooled from the two studies:

\[ r_{piz} = \alpha + \sum_j \beta_j x_{ij} + \delta z + \sum_j \delta_j z x_{ij} + \epsilon_{piz}, \]

where \( z \) is a boolean variable taking the value 0 for the data from the online study and 1 for the data from the paper-and-pencil study. The coefficients \( \delta_j \) capture the difference in partworths for feature \( j \).

At the population level, three feature coefficients differ with statistical significance at \( p < .05 \):

1. **Size** (\( p < 0.001 \)): The difference in the Size coefficient is the most statistically significant (significant at \( p < 0.001 \)). The coefficient corresponding to Size flips in sign: online, people prefer the Large size, but once they see it in person, they prefer the Small size. Note that, as is typical in online shopping experiences, we included the dimensions in inches of the size in addition to the label (‘Large’ or ‘Small’) (see Figure 2 top). Thus, despite the fact the participants were provided with precise dimensional information, we are seeing a statistically significant change in attribute partworth for Size.

2. **Colorful exterior design** (\( p < 0.05 \)): The Colorful pattern was more appealing in person than on the screen. While the post-evaluation coefficient of Colorful is still negative, meaning that, in general, people prefer Black over Colorful, it is significantly higher than the pre-evaluation coefficient. In other words, Colorful is an under-valued attribute.

3. **Water bottle pocket** (\( p < 0.05 \)): The coefficient for Water bottle pocket decreased significantly upon physical evaluation. One probable cause is that the Water bottle pocket is on the inside of the bag and would be hard to reach were the bag to be full, so participants who initially expected to derive utility from the Water bottle pocket may upon examine realize that they would rarely use the pocket.

Three more features are significant at \( p < 0.1 \): Reflective exterior design, which is less appealing in person than online; Price, whose partworth becomes less negative, meaning consumers are less price sensitive in person; and Strap pad, whose partworth decreases, which means people value it less in person than online.

**Individual-level model.** In order to investigate the individual-level differences we fit a linear model to the data collected for each participant. Specifically, for each participant, we pooled her responses from the online and the paper-and-pencil tasks and fit the following two models:

\[ r_{iz} = \alpha + \sum_j \beta_j x_{ij} + \delta z + \sum_j \delta_j z x_{ij} + \epsilon_{iz}, \] (3)

where $z$ is a boolean variable taking the value 0 for the data from the online task and taking the value 1 for the data from the paper-and-pencil task Note that the model in (4) is nested within the model in (3). Hence, an F-test (Anova test) can determine if the two models (3) and (4) differ with statistical significance. If the F-test reveals a statistically significant difference, then we can conclude that the more complex model in (3) is a better fit and hence the models for the online and paper-and-pencil datasets must be different. We found that the models in (3) and (4) are statistically significantly different for $45\%$ of participants (28 out of 62) at $p < 0.05$ and $48\%$ of the participants (30 out of 62) at $p < 0.1$.

**Conclusions.** Based on the above results, we draw the key conclusion that there is a statistically significant difference between the utility partworths when evaluating products online and offline. The key question now is which partworths the customer will use for making purchases online after the store visit. Would the customer revert back to the online partworths or would the offline partworths persist? In order to answer this question, we asked another group of participants to complete the tasks in reverse order: first the physical evaluation task, followed by the online evaluation task. For this group, our individual-level analysis shows that the models are statistically significantly different ($p < 0.05$) for none of the participants (0 out of 20). The result from this second group of participants provides evidence that the partworth of a feature used for the purchase decision depends only on whether the customer has been exposed to the feature in a physical product. It does not depend on the channel in which the purchase decision is made. Once the customer has been exposed to the feature, he will apply the new partworths to both his online and offline purchasing decisions. This simplifies our optimization problem, because we do not have to take into account in which order the customer visits the two channels.

### 4. Assortment Optimization: Without Returns

In this section, we present our results for the problem of determining the subset of products to offer in the offline store in order to maximize the expected sales across both the online and offline channels. We focus on the setting without product returns. Section 5 deals with the issue of product returns.

Recall from Section 2 that our goal is to solve the following decision problem:

$$
\arg\max_{M \in \mathcal{M}} \phi(M) = \arg\max_{M \in \mathcal{M}} \left(1 - \Pr_\epsilon \left( u_0 \geq u_i(\beta^M) \; \forall \; i \in N \right) \right),
$$

at $p < 0.1$ significance, the models were statistically significantly different for $10\%$ (2 out of 20) of the participants.
where \( M \) denotes the collection of feasible in-store assortments; \( u_i(\beta) = \beta^\top x_i + \varepsilon_i \) is the utility assigned to product \( i \) with \( x_i \) denoting the feature vector of product \( i \) and \( \varepsilon_i \) denoting the idiosyncratic error term; and \( \beta^M \) is the utility part-worth vector of the consumer after evaluating the subset \( M \) of products offered in the store. We assume that the part-worth vector is updated as follows:

\[
\beta^M_j = \begin{cases} 
  w_{2j}, & \text{if } j \in S_M, \\
  w_{1j}, & \text{otherwise,} 
\end{cases}
\]

(6)

where \( S_M \) denotes the subset of features \( \{ j \in J : x_{ij} = 1 \text{ for some } j \in M \} \) that are offered as part of the offer set \( M \).

As mentioned in Section 2, we assume that the product universe is full-factorial i.e., it consists of all possible combinations of features: \( \{ x \in \{0,1\}^J : \sum_{k=1}^L x_{k} = 1 \} \). We focus on the capacity constraints of the firm. We start with the un-capacitated setting in which the firm is not constrained in the number of the products it can offer in the brick-and-mortar store. In the capacitated setting, on the other hand, we impose a constraint on the number of products the firm can offer in the brick-and-mortar store.

Solving the decision problem in (5) for a general choice model is computationally challenging. Hence, taking a cue from the existing body of assortment optimization literature, we start with focusing our attention on the single-class MNL model. Some of our results and insights extend to more general choice models. We discuss these extensions at the end of this section.

When the underlying choice model is the single-class MNL model, the probability of purchase \( \phi(M) \) is given by

\[
\phi(M) = \frac{\sum_i \exp \left( \sum_j \beta^M_j x_{ij} \right)}{1 + \sum_i \exp \left( \sum_j \beta^M_j x_{ij} \right)},
\]

(7)

so that the decision problem in (5) becomes

\[
\arg \max_{M \in \mathcal{M}} \sum_i \exp \left( \sum_j \beta^M_j x_{ij} \right)
\]

Due to the monotonicity of the function \( y/(1 + y) \) with respect to \( y > 0 \), we can reformulate the assortment optimization problem as

\[
\arg \max_{M \in \mathcal{M}} \sum_i \exp \left( \sum_j \beta^M_j x_{ij} \right)
\]

(8)

In the case when the product universe is full-factorial, the optimization problem becomes separable in the attributes. Specifically, when the product universe is full factorial, we can write

\[
\sum_i \exp \left( \sum_j \beta_j x_{ij} \right) = \sum_i \prod_j \exp(\beta_j x_{ij})
\]
\[
\sum_i \prod_k \left( x_{k1} \exp(\beta_{k1}) + x_{k2} \exp(\beta_{k2}) + \cdots + x_{kL_k} \exp(\beta_{kL_k}) \right)
= \sum_{(\ell_1, \ell_2, \ldots, \ell_K) \in \mathcal{L}_1 \times \mathcal{L}_2 \times \cdots \times \mathcal{L}_K} \exp(\beta_{1\ell_1}) \exp(\beta_{2\ell_2}) \cdots \exp(\beta_{K\ell_K})
= \prod_{k=1}^K \left( \exp(\beta_{k1}) + \exp(\beta_{k2}) + \cdots + \exp(\beta_{kL_k}) \right),
\]

where the second equality follows because only one of \( x_{k\ell}, \ell = 1, 2, \ldots, L_k \) is non-zero; \( \mathcal{L}_k \) denotes the set of indices \( \{1, 2, \ldots, L_k\} \) for any \( k = 1, 2, \ldots, K \); and the third equality follows because the product universe is full-factorial. With the above simplification, the problem becomes separable in the attributes by taking the logarithm, resulting in the following optimization problem:

\[
\text{arg max } M \subset \mathcal{M} \sum_{k=1}^K \log \left( \sum_{\ell=1}^{L_k} \exp(\beta_{kM}^{\ell}) \right). \tag{9}
\]

Note that the optimization problem in (9) is equivalent to the optimization problem in (8) because of the monotonicity of the logarithm function.

Because the product universe is full factorial, we solve the optimization problem in (9) in two steps: (i) find the optimal in-store feature vector \( a^M \), and (ii) determine a set of products \( M \) to offer in the store that “covers” the optimal in-store feature vector. To that end, we first note the following:

\[
\exp(\beta_{jM}^{\ell}) = a_j^M \exp(w_{2j}) + (1 - a_j^M) \exp(w_{1j}) = \exp(w_{1j}) + a_j^M (\exp(w_{2j}) - \exp(w_{1j})),
\]

where the first equality follows from (6). Now letting \( W_k = \sum_{\ell=1}^{L_k} \exp(w_{1k\ell}) \) and \( d_{k\ell} \) denote the “discrepancy” \( \exp(w_{2k\ell}) - \exp(w_{1k\ell}) \), we can reformulate the optimization problem in (9) as follows:

\[
\text{arg max } a \in \{0, 1\}^J : a \in \mathcal{A}_M \sum_{k=1}^K \log \left( W_k + \sum_{\ell=1}^{L_k} a_{k\ell} d_{k\ell} \right), \tag{10}
\]

where \( \mathcal{A}_M \) is defined as the collection of binary vectors \( \{ a \in \{0, 1\}^J : a = a^M \text{ for some } M \in \mathcal{M} \} \).

We have thus reduced the decision problem from the product space to the feature space. Once we determine the optimal feasible in-store feature vector \( a^* \in \mathcal{A}_M \), it can be mapped to a feasible in-store assortment \( M^* \in \mathcal{M} \) as shown below. But before that, we state our optimality results.

**Theorem 4.1 (Optimal assortment for full-factorial, uncapacitated MNL).** Suppose the product universe is full factorial and there is no capacity constraint. Further suppose that for every attribute \( k \), there is at least one attribute level \( \ell \) such that \( d_{k\ell} \geq 0 \). Then, the optimal solution to (10) is the feature vector \( a^* \) given by \( a^*_{k\ell} = 1 \) if and only if \( d_{k\ell} \geq 0 \) for \( 1 \leq \ell \leq L_k \) and \( k = 1, 2, \ldots, K \).
Proof of Theorem 4.1 We show in Lemma 4.2 that when the firm faces no capacity constraints, the collection $\mathcal{A}$ of feasible in-store feature vectors is described by the set 
$$\left\{ a \in \{0,1\}^J : \sum_{\ell} a_{k\ell} \geq 1 \forall k \right\}.$$ 
In other words, the in-store feature vector $a$ can be any boolean vector taking at least one non-zero value for every attribute $k$. As a result, the optimization problem in (10) separates in the decision variables, so that the optimal in-store feature vector $a^*$ is such that $a^*_{k\ell} = 1$ whenever $d_{k\ell} \geq 0$ and $a^*_{k\ell} = 0$ whenever $d_{k\ell} < 0$.

The result of the theorem now follows. □

The result of Theorem 4.1 is intuitive: the optimal assortment offered in the store is such that the under-valued features (for which $w_{2k\ell} \geq w_{1k\ell}$) are physically exposed to the customers and the over-valued features (for which $w_{2k\ell} \geq w_{1k\ell}$) are physically “hidden” from the customers. We show below (in Lemma 4.2) that capacity needed to capture the maximum possible market share is no more than the maximum number of under-valued levels in each attribute. As a result, even though there may be an exponential (in $K$) number of products in the product universe, the required capacity to achieve the maximum market share will only scale in the maximum number of levels in the attributes.

In the statement of Theorem 4.1 we make the technical assumption that for every attribute $k$, there is at least one attribute-level $\ell$ such that $d_{k\ell} \geq 0$. The technical assumption was made to simplify the exposition of the result. The above result may be extended to the capacitated settings while relaxing the technical assumption. It, however, requires additional notation.

For each attribute $k$, without loss of generality, assume that the levels are indexed such that $d_{k1} \geq d_{k2} \geq \cdots \geq d_{kL_k}$. Let $m_k$ denote the number of under-valued levels of attribute $k$, i.e., $m_k \triangleq \# \{1 \leq \ell \leq L_k : d_{k\ell} > 0\}$. Furthermore, assume that the firm will offer at least one product in the offline store. Under our model, it is certainly possible for the optimal solution to not offer any product in the offline store. In such a case, the firm would compare the expected sales from not offering any products to the maximum possible expected sales of offering at least one product and choose the solution that results in the maximum expected sales. Given this, with the intention of keeping the statements of the results simple and without loss of generality, we assume that the firm is required to offer at most one product in the offline store. We then have the following result:

Lemma 4.1 (General optimality result for full-factorial, uncapacitated MNL).
Suppose the product universe is full factorial and there is no capacity constraint. Then, the optimal solution to (10) is the feature vector $a^*$ given by $a^*_{k\ell} = 1$ if and only if $1 \leq \ell \leq \max\{m_k,1\}$ for $k = 1, 2, \ldots, K$. 


The proof of Lemma 4.1 is similar to that of Theorem 4.1 and is presented in Appendix A. Lemma 4.1 generalizes the result of Theorem 4.1 by allowing all the levels of some attributes to be over-valued ($d_{k\ell} < 0$). In that case, since we must offer at least one attribute level in each attribute $k$, we cannot “hide” all the levels of a particular attribute. Lemma 4.1 states that it is then optimal to offer the “least” over-valued level i.e., the attribute level with the least negative value of $d_{k\ell}$.

The result for the un-capacitated setting extends to the capacitated setting as follows:

**Theorem 4.2 (Optimal assortment for full-factorial, capacitated MNL).** Suppose the product universe is full factorial and we can offer at most $C$ products in the offline store. Then, the optimal solution to (10) is the feature vector $a^*$ given by $a^*_{k\ell} = +1$ if and only if $\ell \leq \min\{C, \max\{m_k, 1\}\}$ for any attribute $k$.

We present the details of the proof of Theorem 4.2 in Appendix A. Theorem 4.2 extends the result of Lemma 4.1 to the case when the firm is constrained by a fixed capacity, measured in terms of the size of the offer set. The result states that for each attribute $k$, it is optimal to offer the top at most $C$ levels according the values of the “discrepancies” $d_{k\ell}$. In other words, the discrepancy measures the significance of each attribute level with respect to determining the optimal offline assortment.

The above results characterize the optimal solution in terms of the optimal in-store feature vector. We now discuss how a feasible in-store feature vector can be mapped to a feasible offline assortment. In particular, letting $A_C$ denote the collection of feasible in-store feature vectors when the firm can offer at most $C$ products in the offline store, we can establish the following result:

**Lemma 4.2 (Feasible in-store feature vectors).** Suppose the product universe is full-factorial and the firm can offer at most $C$ products in the offline store. Let $M_C$ denote the feasible set of products $\{M \subseteq N : 1 \leq |M| \leq C\}$. Correspondingly, let $A_C$ denote the set of feasible in-store feature vectors $a$ that are achievable by $M \in M_C$ i.e., $\{a \in \{0, 1\}^J : a = a^M \text{ for some } M \in M_C\}$. We must then have that $A_C = \{a \in \{0, 1\}^J : 1 \leq \sum_{\ell=1}^{L_k} a_{k\ell} \leq C \forall k\}$. Further, given an in-store feature vector $a \in A$, there could be multiple assortments $M \in M_C$ such that $a^M = a$. The smallest cardinality assortment $M$ such that $a^M = a$ has cardinality $m^* = \max_k \sum_{\ell=1}^{L_k} a_{k\ell}$.

We present the details of Lemma 4.2 in the appendix. The proof is constructive in the sense that we provide an explicit construction of the collection of products that achieve every feasible in-store feature vector $a \in A_C$. The result of Lemma 4.2 shows that when the product universe is full-factorial and capacity is not constrained, feasible feature vectors are given by boolean vectors $a \in \{0, 1\}^J$ such that $\sum_{\ell=1}^{L_k} a_{k\ell} \geq 1$ for all $1 \leq k \leq K$. Similarly, when the product universe is full-factorial and capacity is $C$, feasible in-store feature vectors are given by boolean vectors $a \in \{0, 1\}^J$ such that $1 \leq \sum_{\ell=1}^{L_k} a_{k\ell} \leq C$. 


4.1. Extensions of the results to general choice models.

The results in Theorem 4.1 and Theorem 4.2 can be extended to more general choice models. We discuss extensions to a variant of the nested logit model and the Gaussian mixture of MNL models.

**Nested logit model.** The nested logit model alleviates a well-known shortcoming of the MNL model: the Independence of Irrelevant Alternatives (IIA) property, which specifies that the ratio of the choice probabilities of two different products is independent of the offer set. A consequence of the IIA property is that if a product $j$ is removed from the offer set, then the customers purchasing $j$ will get distributed among the remaining products proportional to their respective market shares. The IIA property fails to hold in many practical applications (see Ben-Akiva and Lerman (1994)), which limits the applicability of the MNL model. For instance, a common application is to model the market share a firm captures from its offer set using a choice model in which the ‘no-purchase’ alternative is assumed to capture the rest of the market. If one uses an MNL model, then $\exp(\mu_0)$ captures the rest of the market, and it is typically the case in many applications that $\exp(\mu_0) \gg \sum_{j \in N} \exp(\mu_j)$ (where $\mu_j$ represents the mean utility of product $j$), so that a typical firm captures only a “small” fraction of the market. Now due to the IIA property, the MNL model specifies that if a single product is removed from the firm’s offering, then a significant fraction (proportional to $\exp(\mu_0)$) of customers purchasing the product move to the rest of the market. This significant loss in market share due to the dropping of a single product is contrary to what is usually observed in practice and is an artifact of the IIA property of the MNL model. One can avoid this issue, by considering a simple extension of the MNL model: the ‘no-purchase’ alternative (rest of the market) is in a nest of its own and the remaining products are in a different nest. This variant of the nested logit model alleviates the above issue while still retaining the tractability of the MNL model.

We consider the following variant of the nested logit model for which the problem of determining the revenue maximizing assortment is known to be polynomially solvable (Davis et al. 2014). The product space is assumed to consist of two ‘nests’: one containing the no-purchase alternative and the other containing all the other products. We normalize the inherent utility of the no-purchase alternative to 0. Under this variant of the nested logit model, the probability that the customers will purchase upon physically evaluating the subset of products $M$ is given by

$$
\phi(M) = \frac{\left( \sum_i \exp \left( \sum_j \beta_j^M x_{ij} \right) \right)^\gamma}{1 + \left( \sum_i \exp \left( \sum_j \beta_j^M x_{ij} \right) \right)^\gamma},
$$

where $\beta_j^M$ represents the utility of product $j$ in nest $M$.
where $0 \leq \gamma \leq 1$ is the nest dissimilarity parameter. The decision problem of determining the subset of products $M$ that maximizes the probability of purchase becomes

$$
\arg \max_{M \in \mathcal{M}} \frac{\left( \sum_i \exp \left( \sum_j \beta^M_{ij} x_{ij} \right) \right)^\gamma}{1 + \left( \sum_i \exp \left( \sum_j \beta^M_{ij} x_{ij} \right) \right)^\gamma}.
$$

(11)

Due to the monotonicity of the function $y^\gamma/(1 + y^\gamma)$ with respect to $y > 0$, the above optimization problem reduces to solving the following optimization problem

$$
\arg \max_{M \subset \mathcal{M}} K \sum_{k=1}^K \log \left( \frac{L_k}{\sum_{\ell=1}^L \exp(\beta^M_{k\ell})} \right).
$$

(12)

Note that the optimization problem in (12) is the same as the optimization problem (9), which we solved for the single-class MNL model. Using the sequence of transformations described above for the single-class MNL model, the optimization problem in (12) can be shown to be equivalent to

$$
\arg \max_{a \in \{0,1\}^J : a \in \mathcal{A}_M} \sum_{k=1}^K \log \left( W_k + \sum_{\ell=1}^L a_{k\ell} d_{k\ell} \right),
$$

(13)

where $\mathcal{A}_M$ is defined as the collection of binary vectors $\{ a \in \{0,1\}^J : a = a^M$ for some $M \in \mathcal{M} \}$. The optimization problem in (13) is the same as the optimization problem (10) we solved for the single-class MNL model. Hence, when the product universe is full-factorial, the results of Theorem 4.1, Lemma 4.1, and Theorem 4.2 extend immediately to the above variant of the nested logit model. For completeness, we state these results as the following theorems:

**Theorem 4.3 (Optimal assortment for full-factorial, uncapacitated NL variant).**

Suppose the product universe is full factorial and there is no capacity constraint. Further suppose that for every attribute $k$, there is at least one attribute level $\ell$ such that $d_{k\ell} \geq 0$. Then, the optimal solution to (13) is the feature vector $a^*$ given by $a^*_{k\ell} = 1$ if and only if $d_{k\ell} \geq 0$ for $1 \leq \ell \leq L_k$ and $k = 1, 2, \ldots, K$.

**Lemma 4.3 (General optimality result for full-factorial, uncapacitated NL variant).**

Suppose the product universe is full factorial and there is no capacity constraint. Then, the optimal solution to (13) is the feature vector $a^*$ given by $a^*_{k\ell} = 1$ if and only if $1 \leq \ell \leq \max \{m_k, 1\}$ for $k = 1, 2, \ldots, K$.

**Theorem 4.4 (Optimal assortment for full-factorial, capacitated NL variant).**

Suppose the product universe is full factorial and we can offer at most $C$ products in the offline store. Then, the optimal solution to (13) is the feature vector $a^*$ given by $a^*_{k\ell} = +1$ if and only if $\ell \leq \min \{C, \max \{m_k, 1\} \}$ for any attribute $k$. 


Gaussian mixture of MNL models. This model is another popular extension, which captures heterogeneity of consumer preferences. We focus on the following specific variant. Each customer samples the pre and post physical evaluation part-worths $w_1$ and $w_2$ according to a multivariate normal distribution:

$$
\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}, \sigma^2 \begin{bmatrix} I_J & 0 \\ 0 & I_J \end{bmatrix} \right),
$$

where $I_J$ is the $J \times J$ identity matrix and $0$ is the $J \times J$ matrix of 0s. Once the pre and post physical evaluation part-worths are sampled, the customer then makes choices according to the part-worth vector $\beta^M$ defined in (6) when the subset of products $M$ is offered in the offline store.

For any feature $j$, let $f_j^\nu$ denote density of the normal distribution with mean $\nu_{ij}$ and standard deviation $\sigma_j$. Similarly, let $g_j^\nu$ denote the density of the normal distribution with mean $\nu_{2j}$ and standard deviation $\sigma_j$. We can now write the probability of making a purchase as

$$
\phi(M) = \int_{\beta} \frac{\sum_i \exp \left( \sum_j \beta_j x_{ij} \right)}{1 + \sum_i \exp \left( \sum_j \beta_j x_{ij} \right)} \prod_{j: a_j^M = 0} g_j(\beta_j) \prod_{j: a_j^M = 1} f_j(\beta_j) d\beta,
$$

Our goal then is to determine the subset of products that maximizes the probability of purchase, which requires solving the following optimization problem:

$$
\arg\max_{M \in \mathcal{M}} \int_{\beta} \frac{\sum_i \exp \left( \sum_j \beta_j x_{ij} \right)}{1 + \sum_i \exp \left( \sum_j \beta_j x_{ij} \right)} \prod_{j: a_j^M = 0} g_j(\beta_j) \prod_{j: a_j^M = 1} f_j(\beta_j) d\beta. \quad (14)
$$

We replicate the results for the single-class MNL model. For that, we borrow notation from the single-class MNL model by re-defining $d_{k\ell} = \Delta \exp(\nu_{2k\ell}) - \exp(\nu_{1k\ell})$ for any attribute-level combination $k, \ell$. We assume that the levels are indexed such that $d_{k1} \geq d_{k2} \geq \cdots \geq d_{kL_k}$. We say a feature is under-valued (over-valued) if $d_j \geq 0$ ($d_j < 0$). Further, let $m_k$ denote the number of under-valued levels of attribute $k$, i.e., $m_k \Delta \# \{1 \leq \ell \leq L_k: d_{k\ell} \geq 0 \}$. We now have the following result:

**Theorem 4.5 (Optimal assortment for full-factorial, uncapacitated Gaussian MNL).**

Suppose the product universe is full factorial and there is no capacity constraint. Furthermore, suppose that for each attribute $k$, there is at least one level $\ell$ such that $\nu_{2\ell} \geq \nu_{1\ell}$, so that $m_k \geq 1$. Then, the optimal solution to (14), which maximizes the probability of purchase, is such that the in-store feature vector $a^M = a^\ast$, where $a^\ast$ is given by $a^\ast_{k\ell} = 1$ if and only if $1 \leq \ell \leq m_k$ for $k = 1, 2, \ldots, K$.

**Proof of Theorem 4.5** We prove a stronger result. Let $M$ be the current in-store assortment, and let $a^M$ denote the in-store feature vector corresponding to $M$. Let $F$ denote the set $\{ j \in \mathcal{J}: a^M_j = 1 \}$ of features that are offered in the store. Suppose there exists a feature $\ell \in \mathcal{J}$ such that $\ell \notin F$ and $\nu_{2\ell} > \nu_{1\ell}$. We then claim that adding the feature $\ell$ to the in-store assortment will increase the
expected sales. To see that, let $F^+$ denote the set of features $F \cup \{\ell\}$. Note that since the product universe is full factorial, there always exists a product $i$ such that the subset $M \cup \{i\}$ results in the set of features $F^+$ offered in the store. We now focus on the change in the probability that a customer does not purchase from adding feature $\ell$:

\[
\Delta = \int_{\beta} \left[ \prod_{j \in F^+} g_j(\beta_j) \prod_{j \notin F^+} f_j(\beta_j) \frac{1}{1 + \sum_i \exp\left(\sum_j \beta_j x_{ij}\right)} \right] \mathrm{d}\beta - \int_{\beta} \left[ \prod_{j \in F} g_j(\beta_j) \prod_{j \notin F} f_j(\beta_j) \frac{1}{1 + \sum_i \exp\left(\sum_j \beta_j x_{ij}\right)} \right] \mathrm{d}\beta
\]

\[
= \int_{\beta} \left[ \prod_{j \in F} g_j(\beta_j) \prod_{j \notin F} f_j(\beta_j) \frac{g_\ell(\beta_\ell)}{1 + \sum_i \exp\left(\sum_j \beta_j x_{ij}\right)} \right] \mathrm{d}\beta - \int_{\beta} \left[ \prod_{j \in F} g_j(\beta_j) \prod_{j \notin F} f_j(\beta_j) \frac{f_\ell(\beta_\ell)}{1 + \sum_i \exp\left(\sum_j \beta_j x_{ij}\right)} \right] \mathrm{d}\beta
\]

where the first equality follows from the fact that $F^+ = F \cup \{\ell\}$. Now, in order to simplify notation, let $X(\beta_{-\ell})$ denote the term

\[
X(\beta_{-\ell}) \overset{\Delta}{=} \prod_{j \in F} g_j(\beta_j) \prod_{j \notin F^+} f_j(\beta_j),
\]

where we use the notation $\beta_{-\ell}$ to denote the vector of all $\beta$s excluding $\ell$. In other words, $X(\beta_{-\ell})$ does not depend on $\beta_\ell$. Now define $\nu^*$ as $(\nu_{2\ell} + \nu_{\ell\ell})/2$. Since $\nu_{2\ell} > \nu_{\ell\ell}$ and $f_\ell(\cdot)$ and $g_\ell(\cdot)$ denote normal density functions with means $\nu_{\ell\ell}$ and $\nu_{2\ell}$ respectively, it must be that $f_\ell(\beta_\ell) > g_\ell(\beta_\ell)$ for $\beta_\ell < \nu^*$ and $f_\ell(\beta_\ell) < g_\ell(\beta_\ell)$ for $\beta_\ell > \nu^*$. We can now write

\[
\Delta = \int_{\beta_{-\ell}} \left[ \int_{\beta_\ell \geq \nu^*} \frac{(g_\ell(\beta_\ell) - f_\ell(\beta_\ell))X(\beta_{-\ell})}{1 + \sum_i \exp\left(\sum_j \beta_j x_{ij}\right)} \mathrm{d}\beta_\ell - \int_{\beta_\ell \leq \nu^*} \frac{(f_\ell(\beta_\ell) - g_\ell(\beta_\ell))X(\beta_{-\ell})}{1 + \sum_i \exp\left(\sum_j \beta_j x_{ij}\right)} \mathrm{d}\beta_\ell \right] \mathrm{d}\beta_{-\ell}.
\]

Focusing on the inner integral, we simplify notation as follows. Let $Y(\beta_\ell, \beta_{-\ell})$ denote the expression $1 + \sum_i \exp\left(\sum_j \beta_j x_{ij}\right) = 1 + \sum_i \exp\left(\beta_\ell x_{\ell\ell} + \sum_{j \neq \ell} \beta_j x_{ij}\right)$. Then, due to the monotonicity of the $\exp(\cdot)$ function, we must have that

\[
Y(\beta_\ell, \beta_{-\ell}) \geq Y(\nu^*, \beta_{-\ell}) \text{ for } \beta_\ell \geq \nu^* \text{ and } Y(\beta_\ell, \beta_{-\ell}) \leq Y(\nu^*, \beta_{-\ell}) \text{ for } \beta_\ell \leq \nu^*.
\]

(15)

The inner integral can now be written as

\[
\int_{\beta_\ell \geq \nu^*} \frac{(g_\ell(\beta_\ell) - f_\ell(\beta_\ell))X(\beta_{-\ell})}{Y(\beta_\ell, \beta_{-\ell})} \mathrm{d}\beta_\ell - \int_{\beta_\ell \leq \nu^*} \frac{(f_\ell(\beta_\ell) - g_\ell(\beta_\ell))X(\beta_{-\ell})}{Y(\beta_\ell, \beta_{-\ell})} \mathrm{d}\beta_\ell
\]

\[
\leq \int_{\beta_\ell \geq \nu^*} \frac{(g_\ell(\beta_\ell) - f_\ell(\beta_\ell))X(\beta_{-\ell})}{Y(\nu^*, \beta_{-\ell})} \mathrm{d}\beta_\ell - \int_{\beta_\ell \leq \nu^*} \frac{(f_\ell(\beta_\ell) - g_\ell(\beta_\ell))X(\beta_{-\ell})}{Y(\nu^*, \beta_{-\ell})} \mathrm{d}\beta_\ell
\]

\[
= \int_{\beta_\ell} \frac{X(\beta_{-\ell})}{Y(\nu^*, \beta_{-\ell})} \int_{\beta_\ell} (g_\ell(\beta_\ell) - f_\ell(\beta_\ell)) \mathrm{d}\beta_\ell
\]

\[
= 0,
\]
where the inequality follows from (15) and last equality follows from the fact that \( \int_{\beta_L} f(\beta_L) d\beta_L = \int_{\beta_L} g(\beta_L) d\beta_L = 1 \). It now immediately follows that \( \Delta \leq 0 \).

We have thus shown that if a feature \( \ell \) is such that \( \nu_2^\ell > \nu_1^\ell \), adding the feature to the offline store will increase the expected sales. Hence, the optimal in-store feature vector \( a^* \) must contain all the features such that \( \nu_2^\ell > \nu_1^\ell \).

Following a symmetric argument, we can show that \( \Delta \geq 0 \) if \( \nu_2^\ell < \nu_1^\ell \), so that removing feature \( \ell \) for which \( \nu_2^\ell < \nu_1^\ell \) will increase expected sales. As a result, the optimal feature vector \( a^* \) should not include any features \( \ell \) for which \( \nu_2^\ell < \nu_1^\ell \).

The rest of the theorem follows from the result of Lemma 4.2 because there is at least one level in each attribute such that \( \nu_2^\ell > \nu_1^\ell \). \( \square \)

The above theorem states that for the uncapacitated setting, the result of the single-class MNL model extends to the Gaussian mixture of MNL models. In particular, if we re-define an ‘over-valued’ (‘under-valued’) feature as the one with \( \nu_2^j < \nu_1^j \) (\( \nu_2^j > \nu_1^j \)), then the optimal in-store feature vector must display all the undervalued features and ‘hide’ all the over-valued features.

We require the minor technical condition that every feature contains at least one level that is undervalued. Otherwise, we will have to perform the following exhaustive search. Let \( a \) denote the in-store feature vector such that \( a_{k\ell} = 1 \) for all features that are undervalued i.e., \( \nu_2^k > \nu_1^k \) and \( a_{k\ell} = 0 \) otherwise. Let \( \mathcal{K} \) denote the collection of attributes \( k \) such that \( a_{k\ell} = 0 \) for all \( \ell = 1, 2, \ldots, L_k \). Finally, let \( \mathcal{A} \) denote the collection of in-store feature vectors defined as

\[
\left\{ a' \in \{0,1\}^J : a'_{k\ell} = a_{k\ell} \forall k \notin \mathcal{K}, 1 \leq \ell \leq L_k \quad \text{and} \quad a_{k\ell k} = 1, a_{k\ell} = 0 \forall k \in \mathcal{K}, 1 \leq \ell_k, \ell \leq L_k, \ell \neq \ell_k \right\}
\]

We then have to exhaustively search through all the feature vectors in \( \mathcal{A} \) to obtain the optimal in-store feature vector. Note that the cardinality of the collection \( \mathcal{A} \) is equal to \( \prod_{k \in \mathcal{K}} L_k \). Hence, the search can be performed efficiently whenever the size of the set of attributes \( \mathcal{K} \) is “small”.

In summary, we have shown that when there are no returns, it is optimal for the firm to carry under-valued features and “hide” over-valued features. This result holds when there is no capacity constraint across the MNL model and variants of the NL and Gaussian mixture of MNL models. When there is a capacity constraint, then it is optimal for the firm to carry the top at most \( C \) levels in each attribute according to the discrepancy values \( d_{k\ell} = \exp(w_{2k\ell}) - \exp(w_{1k\ell}); \) this result is true for the MNL model and a variant of the NL model.

5. Assortment Optimization: With Returns

We now consider the setting in which customers may return a product after making a purchase. As mentioned in Section 2, our model allows us to endogenously model product returns. Specifically, we assume that the customer will return the product if the post-evaluation utility of a product that
is ordered online is less than that of the outside option. As expected, the assortment optimization problem becomes significantly more complicated in the presence of returns. As above, we focus on the setting with a full-factorial product universe and consider the uncapacitated and capacitated cases separately.

More precisely, our goal is to solve the following decision problem:

$$\arg\max_{M \in \mathcal{M}} \phi_R(M) = \arg\max_{M \in \mathcal{M}} \sum_i \Pr\left( u_i(\beta^M) \geq u_{i'}(\beta^M) \forall i' \in N \cup \{0\} \text{ and } u_i(w_2) \geq u_0 \right), \quad (16)$$

where $\phi_R$ captures the expected sales net returns.

We first consider the setting in which the firm is not constrained by capacity. Our goal is to determine the offline assortment that maximizes expected sales, while minimizing returns. Intuitively, one may argue that in the case when returns are allowed, the incentives of the firm and the customer are aligned: both want the customer to identify the ‘best’ product for the customer. Otherwise, the customer may either not purchase (if the products predominantly ‘undervalued’ due to lack of information) or purchase and return (if the products are predominantly ‘overvalued’ due to lack of information). This intuition bears out in the uncapacitated case. Specifically, we can prove the following strong result: if there is no capacity constraint, it is always optimal for the firm to offer all the features in the store, irrespective of the underlying choice structure (error term distribution). In other words, it is in the best interest of the firm to provide ‘full information’ to the customer, which is always possible when there is no capacity constraint. This result continues to hold under all assumptions for the error terms in the utility model! We state this result formally in the following theorem.

**Theorem 5.1 (Optimal assortment for full factorial, uncapacitated, and with returns).** Suppose the product universe is full-factorial and there is no capacity constraint. Then, the optimal solution to (16), which maximizes the probability that a customer will purchase, but not return, is such that the in-store feature vector $a^M = a^*$, where $a^*$ is given by $a^*_k = 1$ for all $1 \leq k \leq K$ and $1 \leq \ell \leq L_k$.

*Proof of Theorem 5.1* Let $s(a)$ denote the sales net returns and $\beta_a$ denote the part-worth vector when the in-store feature vector is $a$. Then, we prove the result by arguing that $s(a) \leq s(a^*)$, where $a^*$ is the full-information in-store feature vector so that $a^*_j = 1$ for all features $j$.

First, we claim that in the full-information case, there are no returns. To see this, note that a customer purchases and returns product $i$ only if $u_i(\beta_a) \geq u_0$, but $u_i(w_2) < u_0$, where $a$ is the in-store feature vector. When $a = a^*$, it must be that $\beta^*_{a^*} = w_2$ so that $u_i(\beta^*_{a^*}) = u_i(w_2)$. Thus, it can never be the case that $u_i(\beta^*_{a^*}) \geq u_0$ and $u_i(w_2) < u_0$ simultaneously. As a result, there are no
returns in the full information case and the probability that a customer will make a purchase is given by

\[ s(a^*) = \Pr_{\varepsilon} \left( \bigcup_i E_i \right), \quad \text{where } E_i \text{ is the event } u_i(w_2) \geq u_0. \] (17)

Now consider any other in-store feature vector \( a \). Product \( i \) will be purchased and not returned only if \( u_i(\beta_a) \geq u_0 \) and \( u_i(w_2) < u_0 \). Let \( E'_i \) denote the event that \( u_i(\beta_a) \geq u_0 \) and \( u_i(w_2) \geq u_0 \). It is then easy to see that \( E'_i \subseteq E_i \). We can now write

\[ s(a) = \Pr_{\varepsilon} \left( \bigcup_i E'_i \right) \leq \Pr_{\varepsilon} \left( \bigcup_i E_i \right) = s(a^*), \]

where the first inequality follows from the fact that \( E'_i \subseteq E_i \) and the last equality follows from (17).

We have thus shown that the expected sales observed at any in-store feature vector \( a \) is bounded above by the expected sales at full-information. The result of the theorem now follows. \( \square \)

We note that the above result is true for any general random utility choice model. Unfortunately, even with assumption that customers make choices according to a single-class MNL model, the optimization problem with a capacity constraint remains complicated. The MNL assumption, however, allows us to derive a closed-form expression for the probability that product \( i \) will be purchased and not returned.

Specifically, let \( \beta \) denote the part-worth vector that the customer uses to make the purchase decision after visiting the store; we drop the dependence of \( \beta \) on the in-store assortment \( M \) to keep the notation simple. Then, the customer purchases and does not return product \( i \) if and only if

\[ u_i(\beta) \geq u_i'(\beta) \text{ for all } i' \in N \cup \{0\} \text{ and } u_i(w_2) \geq u_0(w_2). \]

We may simplify the above expression as follows:

\[ \varepsilon_i \geq \varepsilon_i' + (\beta^T x_i' \neq \beta^T x_i') \text{ for all } i' \in N \cup \{0\} \text{ and } \varepsilon_i \geq \varepsilon_0 - w_2^T x_i, \]

which further simplifies to

\[ V_i + \varepsilon_i \geq \max_{i' \in N} \{ \varepsilon_i' + V_i' \} \text{ and } V_i + \varepsilon_i \geq V_i + \max \{ \varepsilon_0 - w_2^T x_i, \varepsilon_0 - V_i \}, \]

where we have used \( V_i \) to denote \( \beta^T x_i \). Note that in writing the above expression, we have made use of the fact that the customer samples the error terms once and uses the same error term both for the purchase decision and the return decision. We can further simplify the above expression to get

\[ V_i + \varepsilon_i \geq \max_{i' \in N} \{ \varepsilon_i' + V_i' \} \text{ and } V_i + \varepsilon_i \geq \varepsilon_0 + (\beta - w_2) x_i \]
where $x^+$ denotes $\max\{x, 0\}$ for any number $x$. The probability that the customer purchases product $i$ and does not return it is given by

$$\Pr_x \left( V_i + \varepsilon_i \geq \max_{i' \in N} \{ \varepsilon_{i'} + V_{i'} \} \right) \text{ and } V_i + \varepsilon_i \geq \varepsilon_0 + \left( \beta - w_2 \right)^\top x_i. \quad (18)$$

The probability expression above is similar to the probability expression that appears in the computation of MNL choice probability with appropriately defined inherent utilities. Hence, invoking the result for the expression of MNL choice probability [Ben-Akiva and Lerman 1994], we can write that the probability that the customer purchases product $i$ and does not return it as

$$\exp(V_i) \exp \left( \left( \beta - w_2 \right)^\top x_i \right) + \sum_{i' \in N} \exp(V_{i'}) \exp \left( \left( \beta - w_2 \right)^\top x_{i'} \right).$$

The expected sales net returns under the above model can then be written as

$$\phi_R(M) = \sum_{i \in N} \left( \frac{\exp(\beta^\top x_i)}{\exp \left( \left( \beta - w_2 \right)^\top x_i \right) + \sum_{i' \in N} \exp(\beta_{i'}^\top x_{i'})} \right). \quad (19)$$

When there is a capacity constraint, the intuitive result of Theorem 4.2 that the optimal in-store assortment will carry the top at most $C$ non-negative levels according to $\exp(w_{2k}) - \exp(w_{1k})$ in each attribute $k$ is no longer true. To see that consider the following counter-example:

**Example 5.1 (Optimal to offer over-valued features when returns are allowed).**

Consider the setting in which there are two attributes, each with two levels. In the full-factorial, setting we will have a total of 4 products. For concreteness, suppose the attributes are Color and Size and products can either be Red or Blue and either Small or Large. Suppose that the pre and post evaluation part-worth vectors are as follows: $w_1 = (0, 0, 0, 0)$ and $w_2 = (1, -5, 2, 0)$, where the first two components correspond to the colors Red and Blue respectively, and the next two components correspond to sizes Large and Small respectively. We assume that the inherent utility of the outside option is normalized to 0. Now suppose we can offer at most one product in the offline store. We can do an exhaustive search:

<table>
<thead>
<tr>
<th>Offered product</th>
<th>$\beta$ vector</th>
<th>sales net returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>(0, 0, 0, 0)</td>
<td>0.45</td>
</tr>
<tr>
<td>{Red-Large}</td>
<td>(1, 0, 2, 0)</td>
<td>0.75</td>
</tr>
<tr>
<td>{Red-Small}</td>
<td>(1, 0, 0, 0)</td>
<td>0.69</td>
</tr>
<tr>
<td>{Blue-Large}</td>
<td>(0, -5, 2, 0)</td>
<td><strong>0.89</strong></td>
</tr>
<tr>
<td>{Blue-Small}</td>
<td>(0, -5, 0, 0)</td>
<td>0.67</td>
</tr>
</tbody>
</table>

We see that offering the Blue-Large product maximizes the sales net returns. The Blue-Large product corresponds to the in-store feature vector $(0, 1, 1, 0)$, which means that we are offering the
level Blue for which $\exp(w_{2k\ell}) - \exp(w_{1k\ell}) = \exp(-5) - \exp(0) = -0.993$. Thus, the optimal in-store feature vector contains an attribute-level combination such that $d_{k\ell} < 0$ while excluding a level $l'$ with $d_{k\ell'} > 0$. This result is counter to the intuition of the result of Theorem 4.2.

We are not able to theoretically derive a complete characterization of the optimal in-store assortment when there is a capacity constraint, except in the case when all the features are under-valued. If all the features are under-valued i.e., $w_{1j} \leq w_{2j}$ for all $j \in J$, then $(\beta - w_{2})^\top x_i \leq 0$ for all $i$ (since $\beta_j$ is either $w_{1j}$ or $w_{2j}$), which implies that the expected sales net returns in (19) reduces to the expression for expected sales in (7) for the MNL model without returns. Hence, the result in Theorem 4.2 for the capacitated case readily applies. Specifically, we have the following result:

**Theorem 5.2 (Optimal assortment for capacitated, MNL with all feature undervalued).**

Suppose the product universe is full factorial and we can offer at most $C$ products in the offline store. Further, suppose that all the features are undervalued, so that $w_{1j} \leq w_{2j}$ for all $j \in J$. Then, the optimal solution to (16) is the feature vector $a^*$ given by $a^*_{k\ell} = +1$ if and only if $\ell \leq \min\{C, \max\{m_k, 1\}\}$ for any attribute $k$.

Due to the complexity of solving the capacitated assortment optimization problem analytically, we resort to understanding the structure of the optimal assortment through simulations, the details of which are presented in Section 6.

Furthermore, since the optimization problem is computationally challenging for several choice structures, we propose the following greedy heuristic. We demonstrate the accuracy of the greedy heuristic in our simulations, described in Section 6.

**Greedy heuristic**

**Input** Given a collection of product feature vectors $N = \{x_1, x_2, \ldots, x_n\}$; capacity constraint $C$; and function $\phi(M)$ mapping every subset of products to the probability that the customer will purchase, but not return, do:

**Step 1** Initialize the estimate of the optimal assortment $M$ to $\emptyset$.

For $t = 1, 2, \ldots$ do

\[
x^* = \arg\max_{x \in N \setminus M} \phi(M \cup \{x\})
\]

$\hat{M} \leftarrow M \cup \{x^*\}$

If $\phi(\hat{M}) \geq \phi(M)$ or $|\hat{M}| > C$, do

Break from the loop

EndIf

$M \leftarrow \hat{M}$

EndFor

**Output.** Estimate of the optimal in-store assortment $M$.

In the case when the product universe is full-factorial, the above Greedy heuristic can readily be modified to greedily add one feature at a time instead of one product at a time. In that case, we...
initialize the Greedy heuristic with the product $x^*$ such that $x^* = \arg\max_{x \in \mathbb{N}} \phi\{x\}$ and then add features greedily one at a time until we hit the capacity constraint. Once we have the optimal in-store feature vector, we determine the collection of products to cover the feature vector as described in Lemma 4.2.

6. Numerical Study

In this section, we present the results of the numerical studies we conducted to demonstrate the effectiveness of our approach. We conducted two simulation studies and a case study with the data collected on Timbuk2 messenger bags. The goal of the simulation studies is three-fold: (a) demonstrate the gain in net sales a firm can expect from accounting for channel interactions; (b) demonstrate the accuracy of the Greedy algorithm; and (c) gain insight into the firm’s operational decision on how much offline store capacity to invest in. The goal of the case study is to demonstrate how our framework applies to the real-world offline assortment problem faced by Timbuk2. The first two subsections describe the simulation studies and the last subsection describes the case study with data on Timbuk2 messenger bags.

6.1. Greedy performance and impact of joint channel optimization

The goal of the first experiment was to evaluate the performance of the Greedy heuristic and understand the impact of jointly optimizing both the online and offline channels. For this experiment, we focused on the case without product returns. The results are qualitatively similar when product returns are allowed.

**Setup.** We considered a market with 10 customer segments. The universe of products was described by two 4-level attributes. We assumed that the product universe was full factorial, so that there are a total of $4^2 = 16$ products in the universe. For each segment, the post physical evaluation partworths $w_{2j}$ for $j = 1, 2, \ldots, 8$ were generated as i.i.d uniform random variables over the interval $[-3, 1]$. Given the post physical evaluation partworths for each segment, we generated the pre-physical evaluation partworth vector $w_1$ by setting $w_{1j} = 0$ with probability $\xi$ and $w_{1j} = w_{2j}$ with probability $1 - \xi$. Thus, the parameter $\xi$ captures the discrepancy between the pre and post physical evaluation partworths. Intuitively, we expect the loss from our model to increase with increase in the discrepancy between the pre and post physical evaluation partworths.

We varied the discrepancy parameter $\xi$ from 0 to 1 in increments of 0.1 obtaining a total of 11 values. Then, for each value of the discrepancy, we generated 1000 instances of the market (with 10 customer segments) by randomly sampling the model parameters as described above. For each model instance, we imposed a capacity constraint of $C = 3$ and computed the optimal offline assortment under our model through an exhaustive search over $\binom{16}{3} = 560$ possible assortments.
Let $M^*$ denote the optimal assortment obtained through the exhaustive search. Then, we used the greedy heuristic described above, but modified as discussed to search in the feature space, to obtain an estimate $M^{\text{greedy}}$ of the optimal assortment. Finally, we computed the optimal assortment using the ‘classical’ approach where it is assumed that customers make purchases only from a single (offline) channel and will not go back to purchase from the online channel. We determine the optimal assortment again through an exhaustive search and we denote it by $M^{\text{classical}}$.

For each value of the discrepancy, we computed two metrics: the average classical gap and the average greedy gap, defined as

$$
\text{Average percent market share captured classical} = \frac{1}{1000} \sum_{\text{iter}=1}^{1000} \frac{F(M^{\text{classical}}_{\text{iter}})}{F(M^*_{\text{iter}})}
$$

$$
\text{Average percent market share captured greedy} = \frac{1}{1000} \sum_{\text{iter}=1}^{1000} \frac{F(M^{\text{greedy}}_{\text{iter}})}{F(M^*_{\text{iter}})}
$$

where $F(\cdot)$ maps each assortment to the expected sales (market share) captured from both the channels under our model and $M_{\text{iter}}$ refers to the corresponding assortment obtained in model instance iter.

The results of the experiment are presented in Figure 4. We draw the following conclusions from the results:

1. The percentage of optimal sales generated by the classical method decreases as the discrepancy between the consumer’s pre and post physical evaluation partworths increases. This implies that accounting for channel interaction becomes important for products that comprise experience attributes, whose discrepancy is large (Wright and Lynch Jr 1995).

2. The sales loss from using the classical method that does not capture the interaction between the two channels can be significant: of the order of 16% on average observed in our simulation study.

3. The Greedy heuristic provides good approximations to the optimal assortment: an average loss of about only 2% relative to the optimal observed in our simulation study.

6.2. Impact on operational decisions

The goal of the second simulation experiment was to gain insights into how a firm’s operational decisions are impacted by the interaction between online and offline channels. We focus on the firm’s capacity decision; specifically, our goal is to understand how the ‘discrepancy’ between the pre and post evaluation utility partworths impacts the marginal value on each additional unit of capacity in terms of the additional market share it allows the firm to capture. For this experiment,
Dzyabura and Jagabathula: Offline Assortment Optimization in the Presence of an Online Channel

Figure 4  Greedy heuristic provides good approximations and the loss due to ignoring the interaction between channels can be significant. The figure plots the loss in expected sales (captured market share) due to the use of the ‘classical’ approach and the greedy heuristic as a function of the discrepancy between the pre and post physical evaluation partworths captured by the parameter $\xi$ (the higher the $\xi$, the higher the discrepancy). Classical gap measures the fraction of the market share captured if we ignored the interaction between online and offline channels and focused only on the offline channel. Greedy gap measures the fraction of the market share captured if we used the greedy heuristic to estimate the optimal assortment.

we focus on the case with returns. Again, the results are qualitatively similar for the case without returns.

Setup. As above, we assume a universe of products described by two 4-level attributes, for a total of 8 features. We assume a full factorial product design, such that there was a total of $4^2 = 16$ products. We assume a single class MNL model, such that all consumers use the same pre- and post-evaluation partworths for the product attributes. The 8 feature partworths are sampled from a bivariate normal distribution for each instance of the simulation, according to:

$$\left( \begin{array}{c} w_{1j} \\
  w_{2j} \end{array} \right) \overset{d}{=} N\left( \left[ \bar{\omega}_1 \bar{\omega}_2 \right], \sigma \left[ \begin{array}{cc}
1 & \rho \\
\rho & 1 \end{array} \right] \right) \quad (20)$$

The utility of the no purchase alternative, $u_0$ was set to 0. In the results we report, the parameter $\bar{\omega}_2$ was set to $-1$ in order for the sales to vary depending on the assortments: if the means of post-evaluation utilities are too low, most consumers will not purchase, independent of the assortment; if the values are too high, most consumers will purchase, independent of the assortment. The
parameter $\sigma$ was set to 0.5. We vary the values of $\bar{w}_1$ and $\rho$, which control the sign and degree of discrepancy. Note that according to our generative model, the difference of pre and post-evaluation part-worths $w_{2j} - w_{1j}$ is normally distributed with mean $\bar{w}_2 - \bar{w}_1$ and standard deviation $2(1 - \rho)$.

For each set of parameter values, we generated 10,000 model instances. For each model instance, we imposed a capacity constraint $C$ and computed the optimal offline assortment under our model through an exhaustive search over $\binom{16}{C}$ possible assortments. Let $M^*$ denote the optimal assortment obtained through the exhaustive search. We computed the expected sales the firm captures under $M^*$. Because each attribute has 4 levels, we allow $C$ to take on values between 0 and 4; recall from Section 4 that in the context of our model, when the product universe is full-factorial, the firm can expose each customer to all the features with a capacity equal to the maximum of the number of levels for all the attributes.

The optimal expected sales averaged across all the model instances as a function of model parameters and capacity levels is summarized in Table 3. In discussing the behavior of the market in this model, we first focus on the balanced regime, in which the individual product features have an equal chance of being over or undervalued. In this regime, we set $\bar{w}_1 = \bar{w}_2 = -1$. We let $\rho$ take on

<table>
<thead>
<tr>
<th>$(w_1, \rho)$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(-2,0.00)$</td>
<td>0.282</td>
<td>0.595</td>
<td>0.691</td>
<td>0.729</td>
<td>0.743</td>
</tr>
<tr>
<td>$(-2,0.25)$</td>
<td>0.294</td>
<td>0.598</td>
<td>0.691</td>
<td>0.730</td>
<td>0.743</td>
</tr>
<tr>
<td>$(-2,0.50)$</td>
<td>0.303</td>
<td>0.592</td>
<td>0.687</td>
<td>0.727</td>
<td>0.742</td>
</tr>
<tr>
<td>$(-2,0.75)$</td>
<td>0.307</td>
<td>0.583</td>
<td>0.679</td>
<td>0.724</td>
<td>0.743</td>
</tr>
<tr>
<td>$(-2,1.00)$</td>
<td>0.308</td>
<td>0.560</td>
<td>0.664</td>
<td>0.717</td>
<td>0.745</td>
</tr>
<tr>
<td>$(-1,0.00)$</td>
<td>0.541</td>
<td>0.665</td>
<td>0.711</td>
<td>0.733</td>
<td>0.742</td>
</tr>
<tr>
<td>$(-1,0.25)$</td>
<td>0.579</td>
<td>0.682</td>
<td>0.718</td>
<td>0.736</td>
<td>0.743</td>
</tr>
<tr>
<td>$(-1,0.50)$</td>
<td>0.622</td>
<td>0.700</td>
<td>0.727</td>
<td>0.739</td>
<td>0.744</td>
</tr>
<tr>
<td>$(-1,0.75)$</td>
<td>0.668</td>
<td>0.717</td>
<td>0.733</td>
<td>0.740</td>
<td>0.743</td>
</tr>
<tr>
<td>$(-1,1.00)$</td>
<td>0.743</td>
<td>0.743</td>
<td>0.743</td>
<td>0.743</td>
<td>0.743</td>
</tr>
<tr>
<td>$(0,0.00)$</td>
<td>0.586</td>
<td>0.658</td>
<td>0.698</td>
<td>0.725</td>
<td>0.742</td>
</tr>
<tr>
<td>$(0,0.25)$</td>
<td>0.620</td>
<td>0.674</td>
<td>0.705</td>
<td>0.727</td>
<td>0.742</td>
</tr>
<tr>
<td>$(0,0.50)$</td>
<td>0.658</td>
<td>0.690</td>
<td>0.711</td>
<td>0.728</td>
<td>0.741</td>
</tr>
<tr>
<td>$(0,0.75)$</td>
<td>0.700</td>
<td>0.710</td>
<td>0.719</td>
<td>0.729</td>
<td>0.742</td>
</tr>
<tr>
<td>$(0,1.00)$</td>
<td>0.743</td>
<td>0.743</td>
<td>0.743</td>
<td>0.743</td>
<td>0.743</td>
</tr>
</tbody>
</table>

Table 3 Summary of the optimal expected sales averaged across 10,000 model instances at different capacity levels and values of $(w_1, \rho)$. 
Figure 5  The marginal value of capacity increases with the increase in the discrepancy between the pre and post physical evaluation partworths. Each line in the plot corresponds to a different value of the discrepancy, which is controlled by the parameter \( \rho \): the higher the \( \rho \), the lower the discrepancy. The figure plots the percentage of market share captured under the capacity constraint as a function of capacity.

Figure 6  All the regimes converge to maximum potential sales at capacity of 4, as they can attain full information. Undervaluation hurts sales, and the curve is the steepest, meaning that the firm can capture many additional sales with each additional unit of capacity. The overvalued and balanced curves are similar. Overvaluation helps slightly for 0 capacity.

Balanced regime. Because the partworth means are equal, \( \rho \) controls the magnitude of the difference between \( w_1 \) and \( w_2 \). For each value of \( \rho \), we plot a curve tracking how sales grow with each additional unit of capacity (see Figure 5). We see that the higher the value of \( \rho \), the smaller the marginal value of each additional unit of capacity. In the extreme case, when \( \rho = 1 \), the pre- and post-evaluation partworths for all features are equal, so capacity does not generate any additional sales. This is the full information case, which we know from Theorem 4.1 to be an upper bound on sales. When discrepancy is higher (\( \rho \) is low), the firm benefits from providing information to consumers, and capacity is valuable. We see decreasing marginal returns in each additional unit of capacity for a given value of \( \rho \).

Predominantly undervalued and overvalued regimes. In these regimes, “most” features are either overvalued or undervalued, unlike the balanced case when the number of overvalued and
undervalued features were equal in expectation. To simulate these regimes, we set $\bar{w}_1 = -2$ for the undervalued regime and $\bar{w}_1 = 0$ for the overvalued regime.

Undervaluation hurts sales relative to the balanced regime for all capacity levels, as can be seen in Figure 6. The firm gets very low sales at 0 capacity, but at full capacity, sales reach the maximum level. Therefore, the sales vary a lot depending on capacity available. In this regime, the firm benefits a lot from opening a store and exposing the consumers to product attributes. This regime leads to the steepest increases in sales of the three regimes as capacity increases. In the extreme case, if all features are undervalued, or $w_{1j} < w_{2j}$ for all $j$, the problem reduces to the no returns case (Theorem 5.2).

The overvalued regime overall behaves similarly to the balanced regime. The sensitivity to capacity is slightly lower than in the balanced regime, because the firm captures a large portion of its potential market share even with with $C = 0$. In fact, at $C = 0$, the firm captures more sales in the overvalued regime than in the balanced regime. For higher levels of $C$, sales are slightly lower in the overvalued regime than in the balanced regime, but not as low as the undervalued regime.

Our results demonstrate that the firm’s decision to open a store or not and how much capacity to build (or purchase if the products are being shown at a third-party retailer) greatly depends on the ‘discrepancy’ between the pre and post evaluation partworths. Thus, channel interactions become significant for product categories that may experience high ‘discrepancies’ (such as experience goods; see [Wright and Lynch Jr 1995]). Such channel interactions are typically not accounted for and our simulation results demonstrate the gains that can be obtained from accounting for the channel interactions.

### 6.3. Offline assortment optimization for Timbuk2

We now demonstrate how our framework applies to the offline assortment optimization problem faced by Timbuk2. Specifically, we benchmark the optimal decision under our model against the optimal decision under the classical approach, which ignores the online channel. We show that (a) the optimal assortment is very different for both methods, (b) gain in sales can be substantial from accounting for channel interactions.

**Procedure:** We consider the universe of products in the full-factorial design presented in Section 3: 96 products comprising all possible combinations of 4 levels of Exterior Design (Reflective, Colorful, Blue, Black), 3 levels of Interior Compartment (Divider for files, Crater laptop sleeve, Empty bucket), and 2 levels each of Size (Small, Large), Strap pad (Yes, No), and Water bottle pocket (Yes, No). We ignore price for the purposes of this demonstration because of the difficulty in reliably estimating price coefficients using conjoint data (see the discussion in Section 7). Timbuk2 must
determine the subset of products from the above universe to offer in the offline store when the store faces a capacity constraint.

We assume consumers make choices according to an MNL model with online and offline utility partworths given in Table 2. The ratings data provided by the participants only allow us to identify relative preferences (because the scale is not specified) and thus are not sufficient to identify purchase probabilities. Identification of purchase probabilities requires the value of the market share captured by the firm (equivalently, the market share of the no purchase option). For this reason, we set the market share of the firm when returns are allowed to approximately 7%, which corresponds to decrements the intercept term by 10. The results are qualitatively robust to the choice of the market share value.

For the purposes of the demonstration, we also assume that Timbuk2 always offers the ‘default’ product consisting of attribute levels Black, Small, No strap pad, No water bottle pocket, and Empty bucket. The default product consists of all the attribute levels whose partworths were set to zero for the purposes of identifying the rest of the parameters (Table 2). The combined effect of the attributes in the default product is captured by the intercept term. As a result, we use the post physical evaluation value of the intercept term (2.58 under the column ‘Post-veal’ and row ‘Intercept’ in Table 2). Note that Timbuk2 may choose any product from the universe as default; given a default, we re-estimate the parameters of the model by setting the partworths of the attribute levels in the default product to be zero. Further, if there is no natural candidate for the default product, then we can construct sales functions \( \phi_i(\cdot) \) by setting each product \( i \) as the default. Then, we compute the expected sales for offline assortment \( M \) as \( \phi_i(M) \) for some \( i \in M \). Thus, the assumption that the firm always offers a default product is not restrictive.

Given this, we computed the optimal offline assortments \( M^{\text{returns}} \) and \( M^{\text{no returns}} \) under our model with and without returns, respectively. We also computed the optimal offline assortment \( M^{\text{classical}} \) under the classical model, which ignores the online channel and uses the offline partworths for all features from Table 2 to maximize offline store sales. All the optimal assortments were computed with the constraint that the default option is always included. Thus, we define capacity as the number of products that can be offered in addition to the default option. In order to compare the performance of the assortments, we use the following relative gain metric for the setting without returns:

\[
\text{Relative gain in sales} = \frac{\phi(M^{\text{no returns}}) - \phi(M^{\text{classical}})}{\phi(M^{\text{classical}})}
\]

and the following relative gain metric for the setting with returns

\[
\text{Relative gain in sales net returns} = \frac{\phi_R(M^{\text{returns}}) - \phi_R(M^{\text{classical}})}{\phi_R(M^{\text{classical}})},
\]
The gains that can be obtained from accounting for channel interactions can be substantial. The table reports the relative gains in sales and sales net returns over the classical method, which does not account for channel interactions. The table also reports the fraction of customers who return purchased products and the cardinality of the optimal offline assortments.

where $\phi(\cdot)$ denotes the expected sales function without returns, as defined in (7), and $\phi_R(\cdot)$ the expected sales function with returns, as defined in (19). Note that we assume that all of the assortments $M_{\text{returns}}$, $M_{\text{no returns}}$, and $M_{\text{classical}}$ always contain the default product.

**Results.** The results are summarized in Table 4. We draw the key conclusion that the gains in sales and sales net returns from accounting for channel interactions can be substantial. Specifically, we observe an average gain of 45% in the setting without returns and 5% in the setting with returns.

Further, the fraction of customers who return their purchased product can be substantially (about 2% points) higher under the benchmark model when compared to our model. This decrease in returns implies that the gains obtained by our model can be amplified if there is cost to returns.

In addition, note that when we increase the capacity from 3 to 4, $M_{\text{no returns}}$ and $M_{\text{returns}}$ do not change. In fact, $M_{\text{no returns}}$ does not change beyond capacity 2 (see the last three columns in Table 4). In contrast, $M_{\text{classical}}$ will keep growing to utilize the entire capacity. Now, note from Table 4 that as we increase the capacity from 3 to 4, the sales (with and without returns) decrease for $M_{\text{classical}}$. The decrease is due to overvalued features being included. This key observation suggests that investing in more capacity can lead to a decrease in sales (with and without returns) if channel interactions are ignored.

Finally, the optimal assortments $M_{\text{returns}}$ and $M_{\text{no returns}}$ can be very different from the optimal assortment $M_{\text{classical}}$. For instance, the optimal assortments for capacity value of 1 are

- $M_{\text{no returns}} = \text{Colorful, Small, No Strap Pad, No Waterbottle Pocket, Laptop Compartment}$
- $M_{\text{returns}} = \text{Colorful, Large, No Strap Pad, Waterbottle Pocket, Laptop Compartment}$
- $M_{\text{no returns}} = \text{Black, Small, Strap Pad, Waterbottle Pocket, Laptop Compartment}$.

We note the following:
1. As expected (from Theorem 4.2), the optimal assortment when there are no returns, excludes attribute levels (such as Strap pad and Large) that are overvalued and includes attributes levels (such as Colorful) that are undervalued.

2. On the other hand, the optimal assortment under the classical model offers the levels that have the highest partworth value in each attribute. For e.g. Black is included because the partworths of the other Exterior designs is negative. This in contrast with the fact that we also include Colorful under our model because it is an undervalued attribute. Indeed, in the verbal debrief, many participants mentioned to the experimenter that the Colorful design looked much better offline than online.

3. Although the capacity allows the Strap Pad to be shown, the optimal assortment with returns does not include the Strap Pad. Although intuition suggests (based on the fact that with no capacity constraint, providing full information is optimal; see Theorem 5.1) that including another feature provides more information and increases the objective, our results show that this is not the case.

7. Conclusions and Potential Future Directions

The paper focuses on the problem of determining the optimal assortment to offer in the offline channel when customers purchase both from the online and offline channels. The focus is on firms that both design and sell differentiated products that vary on multiple attributes. The products may also be infrequently purchased so that the there may be many novice customers in the category. While assortment optimization has been a mainstay of operations for several decades, we contribute to the existing work by considering a two-channel setting rather than single channel settings.

The key contributions of the paper include: (i) a model based on changing attribute partworths to capture channel interaction when customers visit multiple channels before purchasing; (ii) validation of the model’s key assumption and estimation of model parameters using a conjoint study; and (iii) theoretical results on the structure of the optimal assortment under popular choice models.

Our work lays the foundation for several interesting future research directions, some of which we detail below.

Channel Preference. Customers may exhibit preferences for one channel or the other. When the customers are indifferent between online and offline channels (the assumption made in this paper), the primary purpose of the offline assortment is to provide information to the customer. This still remains the case if the customer prefers the online channel to the offline channel. If however the customer prefers to purchase from the offline channel, then the purpose of the offline assortment is to meet demand as well as provide information.

The customer may also exhibit a continuum of channel preferences, so that the customer may experience disutility from purchasing from the less preferred channel. The model can be extended
to accommodate such preferences by adding a term to reflect channel preference to the utility function.

**Endogenous store visits.** The common assumption in operations is that the store visits of customers are exogenous (i.e., do not depend on the offered assortment). With the presence of the online channel and availability of information about which products are in the offline store, the store visit decision may become endogenous. In this case, a consumer may check the product availability and decide whether to go to the store by trading off the cost of going to the store with the value of the information they would obtain by visiting and examining the available products. Accounting for store visit endogeneity is a natural extension to pursue.

**External validity.** We used primary data collected from a marketing study to do model validation and parameter estimation. This was done because our model requires fine-grained individual level data, which is not typically available in secondary data. Such primary data collection methods however also have limitations that raise concerns about external validity of the obtained parameter values. Marketing researchers have used similar methods of preference elicitation for new product design and demand estimation, and have proposed various ways to help improve the external validity of the estimates. One successful technique is incentive alignment, which we use in our study. Incentive alignment involves compensating participants with prizes that will reflect their preferences, such that they behave more as they would when faced with a real choice. Another method that has been used in marketing that future research could explore is combining preference elicitation tools with empirical data, such as purchase data (e.g. Feit et al. (2010), Rao (2013)). Perhaps future research in operations can explore how secondary data can be used to improve the external validity of such tools when used for assortment decisions.

**Behavioral decision making.** Finally, a natural next step is to gain a better understanding of online versus offline consumer information processing and decision making. It is possible that consumers use different processing, resulting in different decision rules, when examining products online and offline. For example, sensory attributes (such as aesthetics) are easier to process physically than utilitarian attributes (such as laptop compartment). In the offline settings, sensory attributes (such as aesthetics) may receive higher weights than utilitarian attributes (such as laptop compartment) because sensory attributes are easier to process physically. Such differential weighting could result in more non-compensatory decision making in the offline channel, where consumers make a choice based on just a few attributes, rather than making trade-offs among all the attributes. On the other hand, because all the attributes are presented to the consumer as text in the online channel, consumers may distribute their preference weights more evenly. Based on these considerations, our model can be enhanced through a better understanding of systematic biases specific to each channel.
Continuous attributes. Our paper has primarily focused on attributes that take unordered discrete values, such as color, exterior design, internal compartment, etc. However, there may be attributes with ordinal levels. Examples include discretized price, camera resolutions, fabric quality, etc. For such attributes, exposure to one of the levels may allow the customer to update partworths corresponding to other levels of the attribute. For instance, exposure to particular price level may allow the customer update the partworths corresponding to all price levels lower than the observed price level. Such interactions may readily be incorporated in our framework through an appropriate encoding of the attribute levels. For instance, for each price level, we can introduce a variable that takes value 1 if the price is less than that level and 0 otherwise. With such an encoding, exposing to a particular price level allows the customer to update all the partworths corresponding to price levels lower than the exposed price level.

References


Ding, Min. 2007. An incentive-aligned mechanism for conjoint analysis. *Journal of Marketing Research* 44(2) 214–223.


Appendix

A. Omitted proofs in Section 4

Proof of Lemma 4.1 We show in Lemma 4.2 that when the firm faces no capacity constraints, the collection \( \mathcal{A} \) of feasible in-store feature vectors is \( \{a \in \{0,1\}^J : \sum_{\ell=1}^{L_k} a_{k\ell} \geq 1 \forall k \} \). Thus, since \( a \) can be any boolean vector taking at least one non-zero value for every attribute \( k \), the optimization problem in (10) separates in the decision variables. As a result, for every attribute \( k \) such that \( m_k > 0 \), the optimal solution \( a^* \) is such that \( a^*_{k\ell} = 1 \) whenever \( d_{k\ell} > 0 \) and \( a^*_{k\ell} = 0 \) whenever \( d_{k\ell} < 0 \). On the other hand, for any attribute \( k \) such that \( d_{k\ell} < 0 \) for all \( \ell \), we set \( a^* \) in order to obtain the least negative value for \( \log (W_k + \sum_{\ell=1}^{L_k} a_{k\ell} d_{k\ell}) \); since the levels are indexed such that \( d_{k1} \geq d_{k2} \geq \cdots \geq d_{kL_k} \), we set \( a^*_{k1} = 1 \) and \( a^*_{k\ell} = 0 \) for any \( \ell > 1 \). Given our definition for \( m_k \), we can express the optimal feature vector \( a^* \) as \( a^*_{k\ell} = 1 \) for \( \ell \leq m_k \) and \( a_{k\ell} = 0 \) for \( \ell > m_k \) for both the cases.

Thus, for each attribute \( k \), we can have at most \( C \) non-zero elements in \( a \). It is easy to see that then the optimal solution includes the top at most \( C \) non-negative \( d_{k\ell} \) whenever there is at least one attribute level \( \ell \) such that \( d_{k\ell} \geq 0 \). Since we must offer at least one attribute level for each attribute \( k \), for attributes such that \( m_k = 0 \), it is optimal to offer the attribute-level with least positive value of \( d_{k\ell} \) i.e., to offer the attribute-level combination \((k,1)\), where recall that the attribute levels are indexed such that \( d_{k1} \geq d_{k2} \) for all \( 1 \leq \ell \leq L_k \).

The result of the lemma now follows. \( \square \)

Proof of Theorem 4.2 The proof of this result follows an argument similar to the proof of Lemma 4.1. Specifically, finding the optimal in-store feature vector requires solving the following optimization problem:

\[
\arg\max_{a \in \{0,1\}^J} \left( 1 \leq \sum_{\ell=1}^{L_k} a_{k\ell} \leq C \forall k \right) \sum_{k=1}^{K} \log \left( W_k + \sum_{\ell=1}^{L_k} a_{k\ell} d_{k\ell} \right).
\]

Thus, for each attribute \( k \), we can have at most \( C \) non-zero elements in \( a \). It is easy to see that then the optimal solution includes the top at most \( C \) non-negative \( d_{k\ell} \) whenever there is at least one attribute level \( \ell \) such that \( d_{k\ell} \geq 0 \). Since we must offer at least one attribute level for each attribute \( k \), for attributes such that \( m_k = 0 \), it is optimal to offer the attribute-level with least positive value of \( d_{k\ell} \) i.e., to offer the attribute-level combination \((k,1)\), where recall that the attribute levels are indexed such that \( d_{k1} \geq d_{k2} \) for all \( 1 \leq \ell \leq L_k \).

The result of the theorem now follows. \( \square \)

Proof of Lemma 4.2 The proof is constructive in the sense that for every feasible in-store feature vector \( a \), we exhibit a feasible subset of products that achieve it.

We proceed as follows. Let \( \mathcal{A}_C \) denote the set of all feasible in-store feature vectors when the capacity is \( C \).

We first argue that \( \mathcal{A}_C \subseteq \{a \in \{0,1\}^J : 1 \leq \sum_{\ell=1}^{L_k} a_{k\ell} \leq C \forall k \} \). For that, consider a subset \( M \subseteq N \) such that \( |M| \leq C \). Then, the in-store feature vector \( a^M \) corresponding to subset \( M \) must satisfy \( a^M_{k\ell} = \max_{x \in M} x_{k\ell} \).

We must thus have that

\[
\sum_{\ell=1}^{L_k} a^M_{k\ell} = \sum_{\ell=1}^{L_k} \max_{x \in M} x_{k\ell} \leq \sum_{\ell=1}^{L_k} \sum_{x \in M} x_{k\ell} = \sum_{\ell=1}^{L_k} \sum_{x \in M} x_{k\ell} = \sum_{x \in M} x_{k\ell} \leq C,
\]

where the first inequality follows from the fact that the maximum of a set of number is always upper bounded by their sum; the last equality follows from the fact that for each feasible feature vector \( x \) takes one and only one non-zero value among all the levels within the same attribute \( k \); and the last inequality follows from the fact that there are at most \( C \) products in the subset. We have thus show that \( \mathcal{A}_C \subseteq \{a \in \{0,1\}^J : 1 \leq \sum_{\ell=1}^{L_k} a_{k\ell} \leq C \forall k \} \).

We now show that \( \{a \in \{0,1\}^J : 1 \leq \sum_{\ell=1}^{L_k} a_{k\ell} \leq C \forall k \} \subseteq \mathcal{A}_C \). For that, consider a vector \( a \in \mathcal{A}_C \). Without loss of generality, suppose we re-index the attribute levels so that \( a_{k\ell} = 1 \) for some \( 1 \leq \ell \leq s_k \) and \( a_{k\ell} = 0 \).
Note that since $\sum_\ell a_{k\ell} \leq C$, it must be that $s_k = \sum_\ell a_{k\ell} \leq C$ for all $k$. Now consider the subset $M$ of products described by the feature vectors $x_1$, $x_2$, ..., $x_s$ such that $x_{ik\ell} = 1$ if and only if $\ell = i + ((i - 1) \mod s_k)$ where $s = \max_k s_k$. First note that the subset of products $M$ is feasible i.e., $|M| \leq C$ because $s_k \leq C$ for all $k$, which implies that $|M| = s = \max_k s_k \leq C$. Furthermore, note that each feature vector $x_i$ is feasible since there is exactly one non-zero attribute level for each attribute $k$. Now for any $k$, it follows from the definition that $x_{ik\ell} = 1$ for $\ell \leq s_k$ and $x_{ik\ell} = 0$ for all $i$. Therefore, $a^M_{ik\ell} = \max_{1 \leq i \leq s} x_{ik\ell} = 1$ if $\ell \leq s_k$ and 0 if $\ell > s_k$. It thus follows that $a^M = a$. Also note that since each product can contribute at most one non-zero entry to $a$ for each attribute $k$, the minimum number of products needed to cover $a$ is equal to $s$. Thus, the above construction of products is also the minimum cardinality assortment required to cover the optimal in-store feature vector $a$. □

B. Omitted details of the conjoint study

Table 5 presents the details of the Timbuk2 messenger bags that were included in the conjoint study.

<table>
<thead>
<tr>
<th>Task</th>
<th>Exterior Design</th>
<th>Size</th>
<th>Price</th>
<th>Strap Pad</th>
<th>Water bottle pocket</th>
<th>Interior pocket</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Black</td>
<td>Small</td>
<td>$160$</td>
<td>No</td>
<td>Yes</td>
<td>Empty</td>
</tr>
<tr>
<td>2</td>
<td>Blue</td>
<td>Small</td>
<td>$140$</td>
<td>Yes</td>
<td>Yes</td>
<td>Divider</td>
</tr>
<tr>
<td>3</td>
<td>Colorful</td>
<td>Small</td>
<td>$120$</td>
<td>Yes</td>
<td>No</td>
<td>Divider</td>
</tr>
<tr>
<td>4</td>
<td>Reflective</td>
<td>Small</td>
<td>$160$</td>
<td>Yes</td>
<td>No</td>
<td>Empty</td>
</tr>
<tr>
<td>5</td>
<td>Colorful</td>
<td>Large</td>
<td>$160$</td>
<td>No</td>
<td>Yes</td>
<td>Laptop Compt</td>
</tr>
<tr>
<td>6</td>
<td>Reflective</td>
<td>Small</td>
<td>$140$</td>
<td>No</td>
<td>No</td>
<td>Laptop Compt</td>
</tr>
<tr>
<td>7</td>
<td>Colorful</td>
<td>Large</td>
<td>$160$</td>
<td>No</td>
<td>No</td>
<td>Divider</td>
</tr>
<tr>
<td>8</td>
<td>Blue</td>
<td>Small</td>
<td>$120$</td>
<td>No</td>
<td>No</td>
<td>Laptop Compt</td>
</tr>
<tr>
<td>9</td>
<td>Black</td>
<td>Large</td>
<td>$120$</td>
<td>Yes</td>
<td>No</td>
<td>Divider</td>
</tr>
<tr>
<td>10</td>
<td>Colorful</td>
<td>Large</td>
<td>$140$</td>
<td>No</td>
<td>No</td>
<td>Empty</td>
</tr>
<tr>
<td>11</td>
<td>Colorful</td>
<td>Small</td>
<td>$180$</td>
<td>Yes</td>
<td>Yes</td>
<td>Laptop Compt</td>
</tr>
<tr>
<td>12</td>
<td>Blue</td>
<td>Small</td>
<td>$160$</td>
<td>No</td>
<td>Yes</td>
<td>Divider</td>
</tr>
<tr>
<td>13</td>
<td>Colorful</td>
<td>Small</td>
<td>$120$</td>
<td>Yes</td>
<td>Yes</td>
<td>Empty</td>
</tr>
<tr>
<td>14</td>
<td>Black</td>
<td>Small</td>
<td>$180$</td>
<td>No</td>
<td>No</td>
<td>Laptop Compt</td>
</tr>
<tr>
<td>15</td>
<td>Black</td>
<td>Large</td>
<td>$120$</td>
<td>Yes</td>
<td>Yes</td>
<td>Laptop Compt</td>
</tr>
<tr>
<td>16</td>
<td>Reflective</td>
<td>Large</td>
<td>$180$</td>
<td>No</td>
<td>Yes</td>
<td>Divider</td>
</tr>
<tr>
<td>17</td>
<td>Reflective</td>
<td>Large</td>
<td>$120$</td>
<td>Yes</td>
<td>Yes</td>
<td>Laptop Compt</td>
</tr>
<tr>
<td>18</td>
<td>Blue</td>
<td>Large</td>
<td>$180$</td>
<td>Yes</td>
<td>No</td>
<td>Empty</td>
</tr>
<tr>
<td>19</td>
<td>Blue</td>
<td>Large</td>
<td>$120$</td>
<td>No</td>
<td>Yes</td>
<td>Empty</td>
</tr>
<tr>
<td>20</td>
<td>Blue</td>
<td>Large</td>
<td>$160$</td>
<td>Yes</td>
<td>No</td>
<td>Laptop Compt</td>
</tr>
</tbody>
</table>

Table 5 The list 20 messenger bags that were included in the conjoint study.