PROCEEDINGS OF THE SAWTOOTH SOFTWARE CONFERENCE

November 2022

ARCHETYPAL ANALYSIS AND PRODUCT LINE DESIGN

YICHUN MIRIAM LIU[1]
OHIO STATE UNIVERSITY
PETER KURZ[2]
BMS MARKETING RESEARCH + STRATEGY
GREG M. ALLENBY[3]
OHIO STATE UNIVERSITY

ABSTRACT

Product line design is challenged by the diversity of demand in the market and the wide variety of product features available for sale. Some consumers have broad experience in the activities associated with a product category and others engage narrowly and rely on products in more limited ways. The number of product features and their levels is often large and difficult to characterize in a low-dimensional space. Evaluating marketing opportunities when there exist many usage contexts and product features requires the integration of information on what and when features are demanded, and by whom. We propose an archetypal analysis that combines data on the context of consumption, alternative product usage and feature preferences useful for product line design and management.

Keywords: Grade of Membership, Conjoint Analysis, Market Segmentation, Heterogeneity

1. INTRODUCTION

Product line management requires the integration of information on heterogeneous consumer preferences and usage contexts to design and communicate the best array of offerings to consumers. Most products are effective across a range of consumption contexts where attributes vary in their importance. Product offerings that are preferred in one context of use may not be as preferred in another, and some consumers may participate in a wide array of usage context while others may use a product in more limited ways. The variety of demand conditions affecting the successful use of products requires models of heterogeneity that go beyond simply allowing for individual differences in models of preference.

At the heart of product line management is finding promising opportunities in person-situation interactions (Dickson, 1982). An ideal product line is one that offers at least one attractive product variant to each consumer in the market. Consumers who prefer one variant may not be interested in another for many reasons, such as their expertise in applying or using the product, the types of problems the product helps to solve, and features demanded when using the product. A challenge in product line management is in finding the minimum set of offerings from which each consumer might be satisfied.

[1] LIU.7421@OSU.EDU
[2] P.KURZ@BMS-NET.DE
[3] ALLENBY.1@OSU.EDU

The search for productive interactions among persons and situations is complicated by the number of product features and consumption contexts associated with products. Even simple products, like super glue, come in a variety of features and benefits (e.g., invisible repairs, sets in seconds, size, and price) that can be applied to fix the wood frame or leather straps. The interaction among product features alone can be too numerous for an analysis that directly incorporates interaction terms into a model specification. In this paper we explore a new approach—archetypal analysis—to finding productive person-situation interactions through the use of the distribution of heterogeneity, and demonstrate its use in product line design.

Productive person-situation interactions for product lines are found in the tails of a distribution of heterogeneity, not the mean of the distribution. The mean of the distribution (e.g., Figure 1) is useful for locating one offering for sale that would appeal to the average consumer, but is not useful, by itself, in the design of an array of product offerings attractive to different subsets of individuals. An ideal distribution of heterogeneity would be one in which respondents have high positive preference for some product features but a dislike for others. In this case, the optimal product line would segment the distribution of heterogeneity along the line of preferences. Of course, the distribution of heterogeneity is never this cleanly delineated.

In this paper we explore the use of archetypal analysis to identify productive person-situation interactions for product lines. An archetypal analysis employs a mixed membership model of heterogeneity with exemplars, or pure types by which each respondent is characterized. Figure 2 presents the archetypal description of heterogeneity on the simplex with five exemplars (A1 to A5) and each point represents a respondent. The location of each point indicates the mixture of archetypal characterization of each respondent. We develop a model that combines information on 54 consumption contexts, the use of 17 related products and preferences for 74 product features to identify product line opportunities for an industrial adhesive, and show that a product line designed for these archetype respondents is optimal in providing consumers with a set of utility-maximizing offerings.

Figure 1:
Traditional Model of Heterogeneity

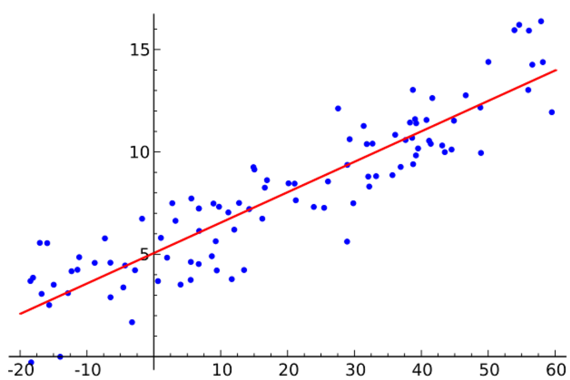
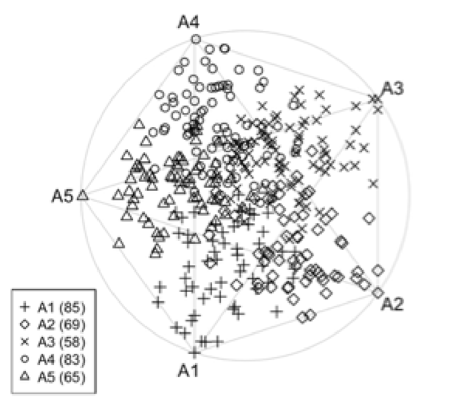


Figure 2:
Archetypal Description of Heterogeneity



The organization of the paper is as follows. We develop our model structure for studying person-situation interactions in Section 2. In Section 3 we describe an industrial study for construction adhesives that includes a series of conjoint studies and questions designed to understand consumption contexts and alternative products used. Empirical results are presented in Section 4, and in Section 5 we discuss product line design based on archetypal analysis. Managerial summaries are offered in Section 6.

2. Model Development

We begin with a description of a standard Grade of Membership (GoM) model (Blei et al., 2003) for analyzing scaled response data used to collect information on consumption contexts, and then discuss the development of multiple GoM models coupled with multiple discrete choice conjoint models to capture the richness of demand for products and product features across consumption contexts. We use the GoM models to describe patterns in consumption contexts and the types of products used by consumers. The conjoint models provide insight into the specific features desired. The combination of these two models results in an archetypal analysis useful for product line design.

2.1 Grade of Membership Model (GoM)

The GoM model belongs to a class of mixed membership models used to summarize high-dimensional multivariate data (Airoldi et al., 2014). It assumes that each individual *n* can belong to multiple clusters that are characterized by exemplars, or archetypes *K* (Blei et al., 2003; Pritchard et al., 2000; Woodbury et al.,1978). Kim and Allenby (2022) and Dotson et al. (2020) have previously applied the GoM model to characterize consumer preferences in choice models. We extend their work by integrating multiple GoM models and multiple choice models that provide greater flexibility in representing heterogeneity and preferences across a large number of scaled response questions and product features. Figure 3 presents the concept of GoM model analysis utilizing consumption contexts. We assume that there are *N* respondents and each respondent *n* provides responses to four discrete survey questions related to consumption contexts that respondents have involved (e.g., *yes/no* for *Build custom furniture, Install kitchen devices, Fix or repair existing furniture,* and *Assemble small broken objects*). The response patterns of all respondents can be characterized by two archetypes (*K=2*)—A1 (Professional) and A2 (Rookie)—where Professional Archetype is represented by involving in all consumption contexts and two contexts representing Rookie Archetype. The membership probability describes the probability of an individual belonging to each archetype (e.g., an individual 80% belongs to A1 and 20% belongs to A2) and is constrained to be non-negative and sum to 1.

Figure 3: Modeling Consumption Contexts

Contexts	A1 (Professional)	A2 (Rookie)
C1. Build custom furniture	Yes	No
C2. Install kitchen devices	Yes	No
C3. Fix or repair existing furniture	Yes	Yes
C4. Assemble small broken objects	Yes	Yes

80% You 20%

membership probability

Figure 4: Modeling Products Used

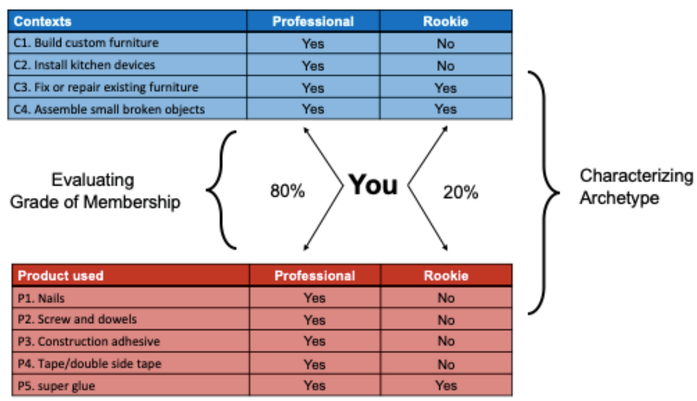
Product used	A1 (Professional)	A2 (Rookie)
P1. Nails	Yes	No
P2. Screw and dowels	Yes	No
P3. Construction adhesive	Yes	No
P4. Tape/double side tape	Yes	No
P5. super glue	Yes	Yes

80% You 20%

membership probability

In the same vein, Figure 4 demonstrates a GoM analysis with binary response for five product used survey questions (e.g., *yes/no* for using *Nails*, *Screw and dowels*, *Construction adhesive*, *Tape/double side tape*, and *Super glue*). The membership profile is now represented by different tools with probabilities for the two archetypes, e.g., A1 (Professional) and A2 (Rookie). In such example, the Professional Archetype is characterized by utilizing all tools, whereas the Rookie Archetype is represented only using super glue. The membership probability of an individual belonging to each product used archetype is constrained to be non-negative and sum to 1 (e.g., an individual 80% belongs to A1 and 20% belongs to A2). Figure 5 shows the concept of integrating scaled response data for consumption contexts and product used GoM analysis. We assume a common membership vector representing the location of the respondent within the convex hull of indicator vectors corresponding to the archetypes described by the membership profiles utilizing contexts and product used.

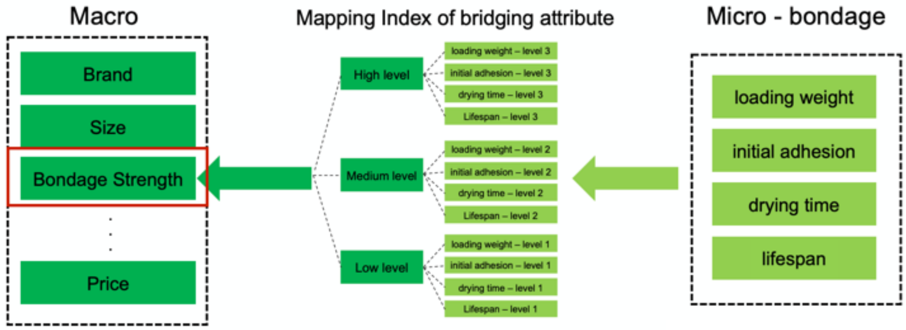
Figure 5: Integrating Multiple GoM



2.2 Integrated Conjoint Model

The second part of our proposed model incorporates choice data from multiple conjoint exercises. Our approach to integrating data from multiple conjoint exercises is through summary attributes (e.g., bridging attributes) of the features that link the datasets. We integrate data from a general, macro conjoint study with data from focused, micro conjoint studies by assuming that part-worth parameters are common to both model specifications. The utility for the linking term is added to the utility specification and can be identified with data from the micro conjoint exercise separately as shown below in our empirical application. Figure 6 presents the concept of integrating a macro conjoint exercise that includes a bundled set of micro-features (e.g., bondage) through the defining of mapping index.

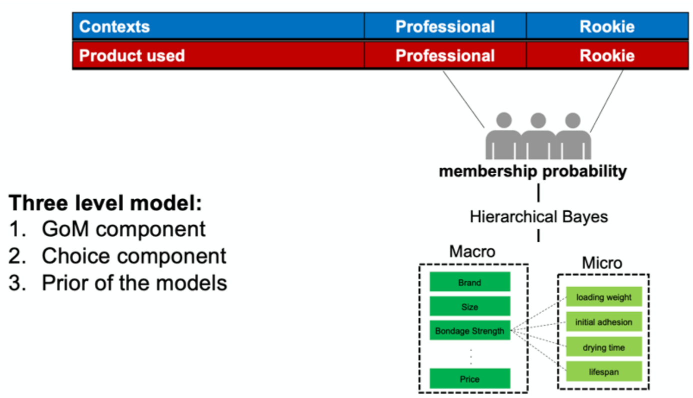
Figure 6: Integrating Multiple Conjoint Exercises



2.3 Integrated GoM and Conjoint Models

The integration of multiple GoM models and multiple choice models is achieved using a hierarchical Bayes specification. Membership probabilities from the GoM models are introduced into the hierarchical Bayes model utilizing a multivariate regression model as covariates in the upper level of the model, or the distribution of heterogeneity. Figure 7 demonstrates this concept of integration, where the archetypal covariates are represented by both consumption contexts and product used to describe the distribution of heterogeneity. Hence, the proposed model is a three level model that consists of GoM components, choice components as well as the prior of models.

Figure 7 Proposed Integrative Model



3. The Data

We examine the proposed model using a dataset from an industry study in which respondents provide detailed information on products used in home repair projects and their preference for various products and features. We refer to these repair projects as usage contexts. The goal of the survey was to understand how consumers currently use a broad array of existing products, and to understand how adhesive products could substitute for the products currently used. The questionnaire is developed based on consultation with market research experts from a large manufacturer of construction adhesives in the United States. Respondents were qualified for inclusion in our study if they have performed at least one project where they needed construction adhesives and have bought construction adhesives within the last 12 months.

Respondents were asked to indicate the products they have used in these repair projects in the last 12 months from a cross-table that lists 54 repair projects and 17 related products used to join material together. Examples of home projects include fixing furniture, installing flooring and repairing rain gutters. These projects constitute alternative contexts for the use of nails, screws, and other types of products used in construction. Data on preference for construction adhesive product features were collected from multiple conjoint exercises (e.g., one macro conjoint and two micro conjoint exercises). The macro conjoint exercise was used to understand the importance of various brands names, price and summary attributes such as bondage (low, medium and high) and field of use (wide, medium and narrow). The micro conjoints are used to measure preferences for the features that comprise the summary attributes plus other features that might possibly be included in the product formulation. There are a total number of 74 attribute levels in the analysis. Respondents provided responses to twelve choice tasks in the macro conjoint exercise and eight choice tasks in each of two micro conjoint exercises. Data from 480 respondents were available for analysis. Figure 8 and Figure 9 displays screen shots of the Macro, Bondage micro, and Field of Use micro conjoint exercises, respectively.

Figure 8:
Screen Shot of Macro Conjoint

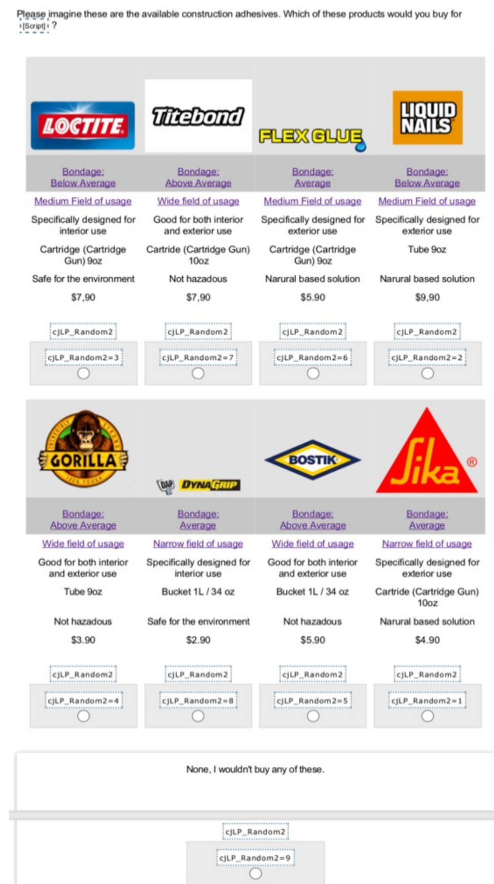
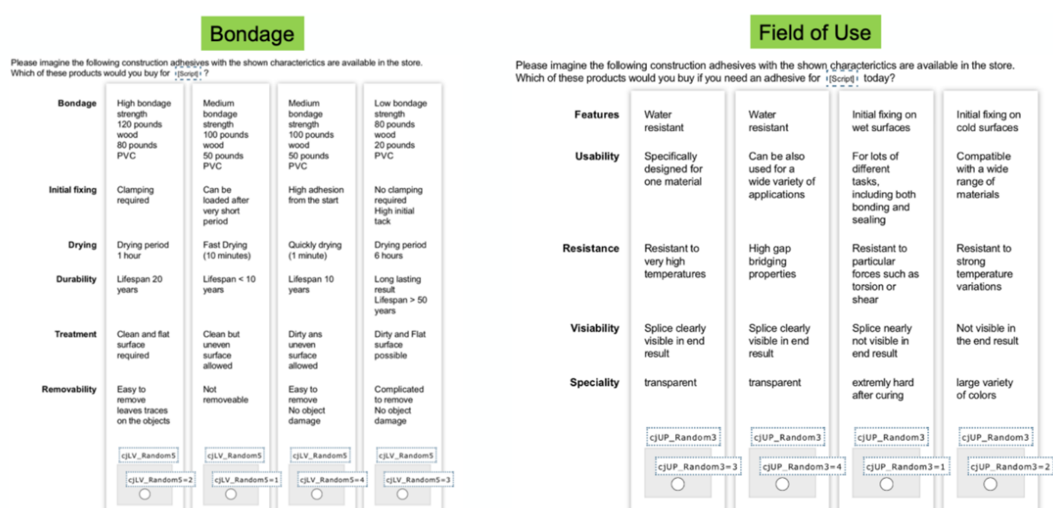


Figure 9: Screen Shots of the Micro Conjoint Exercises



4. The Results

The proposed model is applied to the data describing consumption contexts, product usage and feature preferences to understand what and when product features are demanded. We fitted alternative numbers of archetypes in the proposed model and found marginal difference in model fit between four and six profiles, and the four archetype solution was specified for further analysis as it provided the most distinguishing meaning of each archetype.

We segment individuals into one of the four groups with the highest membership probability and visualize the distribution of heterogeneity on the simplex as shown in Figure 10. Each point represents a respondent, and the four extreme points are the archetypes. The location of each point indicates the mixture of archetypal characterization of each respondent, with points closer to the corners indicating a higher probability of belonging to a particular archetype. The size of each segment is reported in parenthesis in the figure and are about equally sized.

Figure 10: Simplex Plot of Membership Probabilities

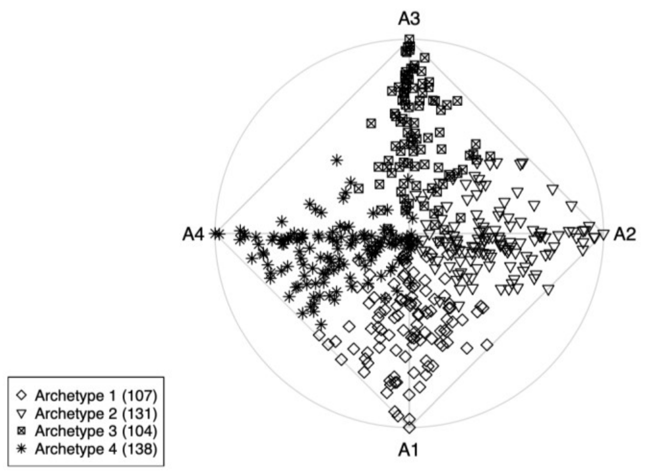


Table 1 presents the summary of archetypal characterization and the preference for product features for each archetype. Repair projects are shown on the top of the table, products used are listed at the middle section and the preferences for construction adhesive attributes are listed at the bottom. Reported are the archetype the respondent engages in the project or uses the indicated products with probabilities that are relatively high for each of the items listed. The results suggests each archetype involves different repair projects, utilizes different tools, and prefers different product features of construction adhesives. For instance, Archetype 3 (A3) describes respondents who are professional or well experienced regarding in-home projects. Most of the usage contexts are estimated with high probability. Especially, they are distinguished by installation tasks within the kitchen/bathroom area. Respondents in this profile use most of the products for joining material listed and are less price sensitive for the construction adhesive.

Table 1: Summary of Archetypal Characterization and Preferences

Archetype 1:	Archetype 2:	Archetype 3:	Archetype 4:

• Repair projects,
• Tasks conducted on the wall/ceiling
• Combining different materials
• Other installation tasks
	• Small number of home repair projects
	  • Install lamp
	  • Install decorative elements outside
		• well experienced in home projects
		• installation tasks within the kitchen/bathroom area.
			• Small number of home repair projects
			  • Design crafts
			  • Small repair works

• Nails,
• Wood screws,
• Screw and anchors
• White glue/wood glue,
• Instant adhesive/super glue
	• A small set of joining material
	  • Nails
	  • Wood screws
		• Almost all solutions	• A small set of joining material.
			  • Screws and anchors

• Not satisfied with current brand • Wide application • High bondage • Not visible • Long-lasting • Easy to remove • Small-sized containers	• Price sensitive • Narrow field of use • Medium bondage	• Less price sensitive • Prefers all brands • Specific range of application • Medium bondage • Large-sized containers	• Price sensitive • Wide field of use • High bondage

5. Discussion

5.1 Market Opportunity of Archetype 1

This paper proposes a model that integrates high-dimensional data on consumption contexts (i.e., repair projects), product usage and consumer preferences to understand market opportunities for product line design. We apply our model to data from a survey on home repair projects where multiple conjoint studies and survey questions are used to measure consumer preferences across 74 attribute levels, 54 projects and 17 existing products. We take a broad approach to measuring consumer preferences by allowing respondents to summarize their preferences across the projects in which they engage. Preferences are measured using conjoint studies that examine specific aspects of product formulation for adhesive offerings that are suspected to not be fully utilized by consumers. We find evidence of a market segment of individuals, characterized by Archetype 1, who engage in a variety of home repair projects but tend to favor traditional products for joining materials such as nails and screws, but not construction adhesives. A strength of our model is that it allows the characterization of what and when product features are demanded across a large number of alternatives.

We investigated current adhesive offerings in the market and found that the product "Loctite PL MAX Premium" comes closest to delivering on the preferred features. As shown in Figure 11, it is advertised to have high bondage strength, long durability, wide application and drying time within 20 minutes. The product is offered only in the form of a 9 oz cartridge with a Manufacturer Suggested Retail Price (MSRP) of $5.90. Detailed product descriptions can be found on the official web site.[4, 5]

However, some of the desired features listed in Table 1 are undersupplied by this product (i.e., drying time within 10 minutes, long lifespan of durability, not visible, easy to remove, and

[4] https://www.loctiteproducts.com/en/products/build/construction-adhesives/loctite_ pl_premiummaxconstructionadhesive.html

[5] https://www.loctiteproducts.com/en/products/build/construction-adhesives/loctite_ pl_premiummaxconstructionadhesive.html/2292244.html#variants-advertisement

small size container). Hence, one approach to better satisfy the demands of Archetype 1 is to introduce a new adhesive product that is an advanced version of the Loctite PL MAX Premium offering. Understanding the context of product use and alternative products that are already used is helpful in product design and communicating the advantages of specific offerings.

Figure 11: New Product for Archetype 1

(a) Bondage Strength, Durability, Specialty (b) Field of Use (c) Drying

5.2 Product Line Design

We investigate the use of archetypal analysis for the design of an entire product line and show that the collection of offerings constructed for each archetype individually creates a product line that is utility maximizing. We begin our analysis by specifying an ideal product for each of the archetypes in our data. Table 2 provides a list of desired product attributes for adhesive bondage and field of use for each archetype along with the corresponding archetypal product design and its cost. The archetypal product designs are based on the preferences for each product attribute and attribute levels. For each archetype, the attribute level with highest part-worth for each attribute is considered as the best element for its product configuration. Thus, a one represents the most preferred attribute levels and zero otherwise in the archetypal design matrix. “High bond,” for instance, is the most preferred attribute level for Archetype 1 among the three bondage attribute levels. Thus, we coded it as one, and zero for “med bond” and the reference attribute level “low bond.”

5.3 Consumer Welfare Evaluation

We evaluate consumer welfare for the archetypal offerings by taking into account private information held by the consumer at the time of choice. This information is represented as the error term in the choice model, whose value is not realized until the respondent is confronted with a choice, (i.e., $u_i = v_i + \varepsilon_i$). Consumer welfare is determined by the maximum attainable utility of a transaction (E[max$\{u_i\}$]), where the maximization operator is taken over the choice alternatives and the expectation operator is taken over error realizations.

For a logit demand model the maximum attainable utility for a given choice set for an individual can be shown to be equal to (Small and Rosen, 1981; Manski et al., 1981):

$$E[\max\{u_i\}] = ln \sum_{i=1}^{I} \exp(v_i) = ln \sum_{i=1} \exp(a_i' \alpha),$$

where I is the number of choice alternatives, a is the product attribute, and α are the part-worth estimates for a respondent. The effect of competitive offers in a product line, thus, are taken into account by considering respondent choices among the archetypal choices and an outside option.

Table 2: Archetypal Product Line Design

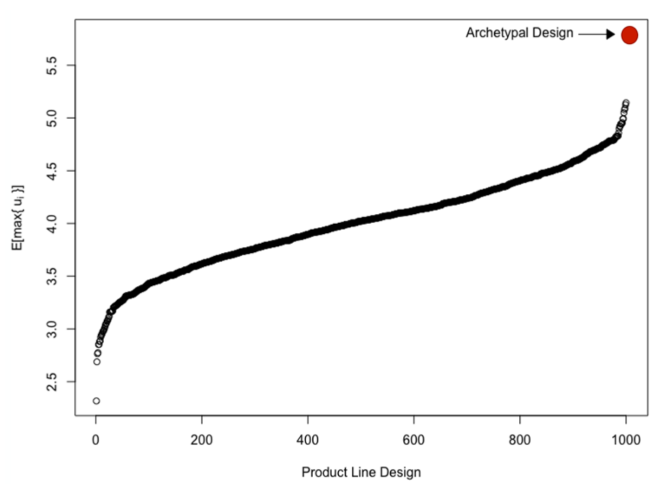
Archetype 1:	Archetype 2:	Archetype 3:	Archetype 4:
• Not satisfied with current brand • Wide application • High bondage • Not visible • Long-lasting • Easy to remove • Small-sized containers	• Price sensitive • Narrow field of use • Medium bondage	• Less price sensitive • Prefers all brands • Specific range of application • Medium bondage • Large-sized containers	• Price sensitive • Wide field of use • High bondage

---------- Archetypal Preferences ----------

Product 1	Product 2	Product 3	Product 4
• High bond • High adhesion • Dry in 1min • Lifespan > 50 years • Dirty uneven surface • Easy to remove • Water resist • Differ task • Resist to particular forces • Not visible • Transparent Cost $3.26	• High bond • No clamping required • Dry in 6hr • Lifespan 20 years • Clean flat surface • Easy to remove • Water resist • One application • Resist to high temp • Visible • Transparent Cost $2.12	• Med bond • Clamping required • Dry in 6hr • Lifespan < 10 • Clean flat surface • Complicate to remove • Wet surface • One application • Resist to High temp • Visible • Extremely hard Cost $1.83	• High bond • High adhesion • Dry in 1min • Lifespan > 50 • Clean uneven surface • Easy to remove • Water resist • Wide application • Resist to high temp • Not visible • Transparent Cost $2.92

We evaluate the archetypal product line to alternative product lines through a simulation study that compares the archetypal design (AD) to a randomly generated design (RD) of product attributes. Figure 12 presents the ranked maximum attainable utility for 1,000 randomly generated designs and the AD, indicating that the AD design generated the highest level of consumer welfare. Plotted in Figure 12 is the expected maximum utility for each design integrated over the distribution of heterogeneity. The point marked in red in the upper right of the figure corresponds to the AD and the remaining points are from the RDs. The AD maximized consumer welfare by ensuring that at least one of the AD products appeals to each respondent.

Figure 12: Ranked Expected Maximum Utility of Archetypal and Random Designs



5.4 Profit of Product Line Design

The AD products can also be evaluated in terms of profits by incorporating cost information from Table 2 into the analysis. For each of the 1,000 randomly generated RDs, we compute its marginal cost and make the additional assumption that the shelf price of an offering is set to twice its total marginal cost. Each individual's evaluation of product line Profit is calculated as:

$$Profit = \sum_{i=1}^{I} Pr_i \times (Price_i - Cost_i) = \sum_{i=1}^{I} \frac{\exp(v_i)}{\sum_{i=1}^{I} \exp(v_i)} \times (Price_i - Cost_i),$$

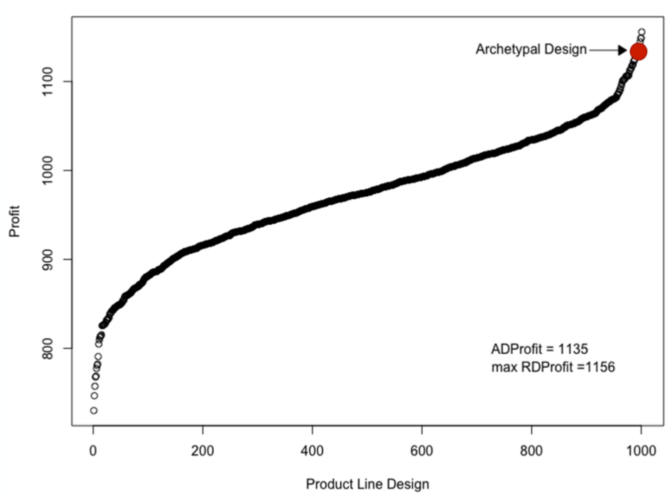
where $v_i = a_i'\alpha + \beta_p Price_i$, and $Cost_i = a_i' c_i$ where c is the cost of each attribute levels. The reported profit is the expected profit across products per respondent.

Figure 13 presents the ranked mean profit of the product line for 1,000 RDs and the AD evaluated across respondents. On the top right of the figure, marked in red, is the profit of the archetypal product line design. The result shows that majority of random designs have lower profits compared to the archetypal product line design. The results shows that the AD also generates near maximum profits.

6. Managerial Takeaways

Consumer preference for products is context-specific, and understanding the effect of consumption contexts is important for effective product development and communication. Consumers may not be aware that products are effective in some contexts, providing a potential source of untapped demand to sellers and a source of enhanced product solutions to buyers. When there exist many usage contexts of the product as well as many potential offerings, evaluating demand and identifying market opportunities requires the integration of information from multiple perspectives. Our proposed model provides a way to comprehensively examine data on the context of consumption, related product usage, and consumer preferences for product features. The managerial implication of our study is summarized as follows:

Figure 13: Expected Profit for Archetypal and Randomized Designs ($)



1. Archetypal analysis provides a rich description of heterogeneity. We show that people involved in different repair projects (usage contexts) have different preferences for product features and tend to use different products (products used) to achieve their goals.
2. Archetypal analysis helps to identify market segments that under-utilize existing offerings while simultaneously wanting specific combinations of product features that are not currently available in the market.
3. An archetypal representation of heterogeneity is useful for product line design. We find that products designed specifically for each archetype comprise a utility maximizing set of offerings.
4. We also find that the archetypal array is nearly profit maximizing assuming prices are set in proportion to marginal costs.
5. Our proposed model offers one way of dealing with the high dimensionality of consumption contexts while offering a parsimonious analysis of demand and market opportunity.

YiChun Miriam Liu

Peter Kurz

Greg M. Allenby

Reference

Airoldi, Edoardo M, David Blei, Elena A Erosheva, Stephen E Fienberg. 2014. *Handbook of mixed membership models and their applications*. CRC press.

Blei, David M, Andrew Y Ng, Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research* 3 993–1022.

Dickson, Peter R. 1982. Person-situation: Segmentation's missing link. *Journal of marketing* 46(4) 56–64.

Dotson, Marc R, Joachim Büschken, Greg M Allenby. 2020. Explaining preference heterogeneity with mixed membership modeling. *Marketing Science* 39(2) 407–426.

Kim, Hyowon, Greg M Allenby. 2022. Integrating textual information into models of choice and scaled response data. *forthcoming*, *Marketing Science*.

Manski, Charles F, Daniel McFadden, et al. 1981. *Structural analysis of discrete data with econometric applications*. Mit Press Cambridge, MA.

Pritchard, Jonathan K, Matthew Stephens, Peter Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2) 945–959.

Small, Kenneth A, Harvey S Rosen. 1981. Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society* 105–130.

Woodbury, Max A, Jonathan Clive, Arthur Garson Jr. 1978. Mathematical typology: a grade of membership technique for obtaining disease definition. *Computers and biomedical research* 11(3) 277–298.