Towards High-Resolution Multi-Channel Marketing for Physical Stores

A Thesis submitted to attain the degree of

DOCTOR OF SCIENCES OF ETH ZURICH

(Dr. sc. ETH Zurich)

Presented by

EDWARD HO

MSc. ETH Robotics, Systems and Controls, ETH Zürich

Born on 31.03.1983

Citizen of Canada

Accepted on the recommendation of

Prof. Dr. Elgar Fleisch, ETH Zürich (advisor)

Prof. Dr. Thomas Rudolph, University of St. Gallen

Dr. Alexander Illic, University of St. Gallen

2014
Acknowledgements

First and foremost, I want to thank Prof. Elgar Fleisch for giving me the opportunity to work with industry changing projects and for his support in supervising my research. His personal inspiration and vision of the Internet of Things goes far beyond the narrow scope of this piece of work, and would stick with me for years to come. I would also like to thank Dr. Alexander Ilic for becoming my co-supervisor, for his valuable feedback and for teaching me by example about entrepreneurship, storytelling, and transforming whole industries – these are lessons I will take with me long after my time at ETH Zürich. I thank also Prof Thomas Rudolph for kindly providing his support as co-supervisor and for taking the time to give his marketing insight about my research. A big thank-you to our retailer partner, who wishes to remain anonymous, for providing the data set for which this thesis is not possible without.

Furthermore, I would like to thank all of my colleagues at ETH and HSG for the collaboration, discussions and pleasant time we had. I am grateful to Peter Meyer, who encouraged my research from the very beginning and through my most difficult times, and gave me practical and hard advice about research. Thanks also to Tobias Kowatsch and Flavius Kehr, who supported and encouraged my research every step of the way, and who inspired the psychology and information systems sections of my thesis. I would like to mention Runhua Xu for the many discussions we had about research, career aspirations, and everything in between. I am very grateful for Denis Vuckovac for providing helpful feedback on the thesis. I am also grateful to Cristina Kadar, Stefan Mau, Remo Frey, Efe Aksüyek and Klaus Fuchs, for our many discussions about data science and careers towards the end of my PhD. A big thank you also to my colleagues, past and present for creating the great working environment: Claire-Michelle Loock, Dirk Volland, Dominic Wörner, Elisabeth-Vetsch Keller, Fabian Wahle, Felix Lossin, Felix Wortmann, Florian Michahelles, Gundula Heinatz, Irena Pletikosa Cvijikj, Jan Gugenheimer, Kay Noyen, Kristina Fluchter, Marcus Köhler, Markus Weinberger, Paul Rigger, Stefanie Turber, Stephan Karpischek, Thomas von Bomhard, Thorsten Staake, Verena Tiefenbeck and Vojkan Tasic. Thanks also to my project partners in IoT-A, especially Alexander Salinas, Martin Bauer, Martin Fiedler, Jourik de Loof, Rob van Kranenburg, and Werner Liekens.

Finally I want to express my deepest gratitude towards my family for their unconditional support, particularly my parents, my brother and especially Ramona, who have probably heard lots from my side about the thesis in the last few years.
Overview of Contents

1. Introduction ............................................................................................................................... 1
2. Related Work ........................................................................................................................... 13
3. Towards High Resolution Identification of Variety-Seeking Behavior ................................. 35
4. Leveraging Social Learning to Market Lesser-Known Products: The Sales Velocity Effect on Retailing ...................................................................................................................... 64
5. Towards High Resolution Evaluation of Customer Profitability ........................................ 99
6. General Discussion and Conclusions ..................................................................................... 122
7. Bibliography ........................................................................................................................... 145

Appendix: Questions and Measures of Chapter 4 Studies .......................................................... 165
## Table of Contents

1. Introduction .............................................................................................................................. 1
   1.1 Problem Statement & Research Questions ...................................................................... 3
       1.1.1 How Retailers Could Segment Customers in Multichannel Grocery Retailing ....... 5
       1.1.2 How Retailers Could Leverage Marketing Methods in One Channel to Enhance Performance in Another ........................................................................................................... 7
       1.1.3 The Need for High Resolution Customer Relationship Management ............... 8
   1.2 Research Methodology ..................................................................................................... 9
   1.3 Structure of the Thesis .................................................................................................... 10
   1.4 Statement of Prior Published Work ................................................................................ 12

2. Related Work .......................................................................................................................... 13
   2.1 Towards One-to-One Behavioral Segmentation of Customers ...................................... 13
       2.1.1 Identifying Customer Preferences by Psychometric Questionnaires .......... 14
       2.1.2 Identifying Customer Preferences for Recommender Systems ..................... 17
       2.1.3 Marketing and IS Motivations for Studying Variety Seeking Behavior ............ 19
       2.1.4 The Unit of Analysis for Variety Seeking ................................................................. 20
       2.1.5 Approaches in Enumerating Variety Seeking .......................................................... 20
   2.2 Leveraging Social Learning Marketing for Boosting Sales .............................................. 22
       2.2.1 Observational Learning in the Context of Social Learning Marketing .............. 22
       2.2.2 The Sales Velocity Research Gap ............................................................................. 25
   2.3 Customer Relationship Management in the Pre-Smartphone Age ................................ 27
       2.3.1 Research in Customer Relationship Management Profitability ......................... 27
       2.3.2 Models of Customer Lifetime Value (CLV) .............................................................. 30
       2.3.3 Spatial Resolution for CLV Models and Their Timely Application .................... 32

3. Towards High Resolution Identification of Variety-Seeking Behavior .................................... 35
   3.1 Introduction .................................................................................................................... 35
   3.2 Research Framework ...................................................................................................... 36
       3.2.1 Probabilistic Model Fit for Variety of Goods Purchased ................................. 37
3.2.2 Customer Variety Seeking Heterogeneity Across Product Categories .......................... 39
3.2.3 Customer Typologies ................................................................................................. 39
3.2.4 Time Effects on Consumer Typologies and Overall Variety Seeking ......................... 40
3.3 Method and Results ........................................................................................................ 40
3.3.1 Dataset for Estimation ............................................................................................. 40
3.3.2 Product Categories Definitions ................................................................................ 41
3.3.3 Probabilistic Model Fit for Variety of Goods Purchased ......................................... 42
3.3.4 Customer Variety Seeking Heterogeneity Across Product Categories ....................... 45
3.3.5 Customer Typologies ............................................................................................... 50
3.3.6 Time Effects on Consumer Typologies and Overall Variety Seeking ....................... 52
3.4 Discussion ....................................................................................................................... 57
3.4.1 Applicability at Different Granularity Levels of Categories ..................................... 57
3.4.2 From Category Level Variety Seeking to Overall Variety Seeking ........................... 58
3.4.3 Contribution to Research ........................................................................................ 58
3.4.4 Contribution to Retailers: An Appropriate High Resolution Segmentation Scheme for Physical Grocery Stores .................................................................................................. 59
3.4.5 Accuracy of results .................................................................................................. 61
3.5 Conclusion and Future Research ..................................................................................... 62
4. Leveraging Social Learning to Market Lesser-Known Products: The Sales Velocity Effect on Retailing .......................................................................................................................... 64
4.1 Introduction .................................................................................................................... 64
4.2 Research Framework ...................................................................................................... 65
4.2.1 The Sales Velocity Effect .......................................................................................... 65
4.2.2 The Role of Numeric Framing .................................................................................. 65
4.2.3 The Negative Sales Velocity Effect .......................................................................... 66
4.3 Overview of Studies ........................................................................................................ 67
4.4 Study 1: Establishing the Sales Velocity Effect ................................................................ 67
4.4.1 Task .......................................................................................................................... 67
A.2.2 Extent of Hedonism and Frequency of Purchase (used in Study 2 and 3 of Chapter 4):

................................................................. 166
List of Figures

Figure 1: Depiction of customer-retailer interactions with a retailer and the underlying retailer processes as modified from the MCMD framework ................................................................. 4

Figure 2: Structure of the thesis ........................................................................................................ 12

Figure 3: Customer classification and corresponding marketing tactic, depending on CLV dimensions (temporal dimension included), as modified from Wang & Hong (2006) .................................................... 33

Figure 4: Frequency distribution of customers who bought n-number of products ......................... 44

Figure 5: Variety index of a customer given the number of unique products bought in a year ............... 46

Figure 6: Sample of four customers and their distribution of variety indices ........................................ 48

Figure 7: Distribution of Gini Coefficients which show the extent of diversity of variety indices ............. 49

Figure 8: (a) A histogram of the customer’s average variety index. (b) A plot of the customer’s mean variety index and the degree of consistency, indicated by the Gini coefficient of the variety indices .......... 50

Figure 9: Four examples of customers with their overall variety seeking typology and their variety indices in the top five categories. For convenience, variety seeking behavior is marked as L=low, M=medium, and H=high depending on the index value .................................................................................................................. 51

Figure 10: Consistency over time of mean variety seeking ..................................................................... 54

Figure 11: Consistency over time of variety seeking consistency ............................................................ 55

Figure 12: Distribution of mean consistency of variety seeking exhibited within product categories .......... 56

Figure 13: Mapping of customer overall variety seeking and consistency towards retailer strategy ......... 60

Figure 14: A product’s average sales per week and the associated sales rank. For that same product at the given sales rank, the triangles then indicates the average rank change of the product, and the dashed box the mid tail .................................................................................................................................. 69

Figure 15: Average sales and maximum rank change distributions for select product categories .......... 69

Figure 16: Components in a product description; the descriptions were made as similar as possible, and parameters such as the current rank, past rank and choice set size were controlled for in a pre-test .... 72

Figure 17: Components in a product description; two generic profiles Q and R are presented ............... 86

Figure 18: Embedding sales velocity into an advertisement or featured product (a) or also within a recommendation (b) .......................................................................................................................... 91

Figure 19: A schematic showing how sales velocity (c) can complement top selling products (b); they form separate features which show different products in the long tail .................................................................................................. 93

Figure 20: Schematic which shows how sales velocity can be implemented as a category-grouped list of products sorted by the change in sales rank within that category ................................................................................................. 95

Figure 21: Two customers who have the same RFM at the store-level, and therefore, same CLV, but different category level RFM, and therefore, different CLV at the category level ................................................................. 102

Figure 22: Concept of the category level CLV ........................................................................................ 104

Figure 23: For milk, the actual and model based number of transactions (a) and actual and modeled distribution of purchases (b) .................................................................................................. 109

Figure 24: Histogram of Gini coefficients which show the extent of diversity of CLV indices ............... 111
Figure 25: Typology of marketing strategies given a person’s Gini coefficient of CLV indices and overall CLV index ................................................................................................................................................ 114

Figure 26: Customers’ reactions to promotions and the effect on CLV ........................................................ 117

Figure 27: Individual level reaction to promotions (a) and the effect on (b) average spending per transaction and (c) CLV ........................................................................................................................................ 118

Figure 28: Depiction of customer-retailer interactions with a retailer and the underlying retailer processes as modified from the MCMD framework, and the corresponding contributions of the thesis chapter contributions .................................................................................................................................... 124

Figure 29: Overall and category level variety indices of four customers (a) and the corresponding typologies and retailer promotion strategy (b) ................................................................................................... 127

Figure 30: Sales velocity embedded as a featured product (a), within a recommendation (b) or as a list of products sorted by sales velocity (c) .................................................................................................. 132

Figure 31: Link between the MCMD framework and the concrete analytical and marketing steps to take, as guided by the thesis chapters ........................................................................................................................................ 138
List of Tables

Table 1: Classifying customer bases, as modified from Fader and Hardie (2009) ......................................................... 31
Table 2: Summary statistics of best fitting distributions using different category levels. .................................................. 43
Table 3: Sample of variety seeking categories. Note that $\lambda_{\text{Poisson}} = \mu_{\text{variety}}$ ...................................................................................................................... 45
Table 4: Best fitting distribution mapping in each quarter, compared against an estimation on a whole year of data .................................................................................................................................................... 52
Table 5: Product characteristics for “hit”, “mid-tail” and “long-tail” products for a physical retailer ...................... 70
Table 6: Summary of paired Mann-Whitney U statistics (given by U) for the likelihood of purchase. The * indicates significant p-values after a Bonferroni correction was applied .............................................. 77
Table 7: Summary of paired Mann-Whitney U statistics (given by U) for the sales rank change. The * indicates significant p-values after a Bonferroni correction was applied .............................................................................................................. 78
Table 8: Groups and number of participants in Study 2 ........................................................................................................ 83
Table 9: Summary of manipulation checks; the * indicates significant p-values (p<0.05) for the corresponding test ..................................................................................................................................................... 84
Table 10: Results for the Mann-Whitney U tests and descriptive statistics for the likelihood of purchase. The * indicates significant p-values of the Mann-Whitney U test ................................................................................. 85
Table 11: Summary of paired Mann-Whitney U statistics for the likelihood of purchase. The * indicates significant p-values of the Mann-Whitney U test ..................................................................................................................................................... 88
Table 12: Interpurchase times for five selected product categories ................................................................................. 111
Table 13: CLV and CLV indices computed by aggregation on vs. direct purchases .................................................. 112
Table 14: Three examples of customers with their CLV indices computed from their purchases in four categories, as well as their CLV indices computed from all 128 categories .......................................................... 113
Table 15: Cross-tabulation table of number of customers in different segments of CLVsummed and Gini coefficients, with number of product categories visited in brackets .................................................................................................................. 115
Table 16: Insights from the Wang and Hong Framework combined with CLV-Consistency Framework of this Thesis ............................................................................................................................................... 136
Abbreviations

A  Accessibility
AIC  Akaike Information Criterion
ANOVA  Analysis of Variance
ANOVA  Analysis of Variance
BIC  Bayesian Information Criterion
CLV  Customer Lifetime Value
CRM  Customer Relationship Management
CV  Coefficient of Variance
DERT  Discounted Estimated Residual Transactions
EDLP  Everyday low price stores
FMCG  Fast Moving Consumer Goods
G  Gini Coefficient
H  Hypothesis
HiLo  Stores with overall high pricing, but temporarily deep discounts
H(n)  Kruskal-Wallis statistic
IS  Information Systems
I  Poisson’s Lambda for Variety Seeking
M  Expected Average Revenue per Transaction
m  Mean
MCMD  Multichannel Customer Management Decision
MLE  Maximum Likelihood Estimation
n  Number of subjects
NBD  Negative Binomial Distribution
OL  Observational learning
p  p-Value
PCV  Past Customer Value
PoS  Point of Sale
Q1  First annual quarter
Q2  Second annual quarter
Q3  Third annual quarter
Q4  Fourth annual quarter
r  Effect size
RFM  Recency, Frequency and Monetary Value
RQ  Research Question
RS  Recommender Systems
s  Standard Deviation
SOP  Share of Purchase
SOV  Share of Visits
SOW  Share of Wallet
T  Skillings-Mack Statistic
t  Time Index
T  Trend
t(n)  T-statistic
U  Mann-Whitney U statistic
V  Volatility
α  Cronbach's Alpha
Abstract

It is widely accepted in E-commerce that highly granular marketing – in the form of timely and personalized offers - is more effective than applying the same marketing policy to all customers.
For physical grocery retailers to also benefit from such highly granular marketing, previous literature have identified three obstacles which they must overcome; the stores need the capability to (1) collect data about customer transactions, (2) transform the data into insights and (3) operationalize the results. In the past, with customer loyalty cards and the customer relationship management systems, it was already possible to collect transaction data; the increasing ubiquity of smartphones and tablets make it for the first time economically and technologically feasible to also operationalize multi-channel, personalized marketing for physical grocery retailing, such as through apps and recommendation systems.

In spite of the identified technological and economic readiness enabled by these emergent technologies, analytical and methodological challenges remain. Existing methods which give retailers insights about which customers to target – such as models of customer’s lifetime value (CLV) and psychometric measures of customer variety seeking – ignore that customers have heterogeneous buying preferences for different product categories, since traditional mass marketing assumes that customers have uniform purchasing preferences across categories. Additionally, a further challenge is how to leverage the marketing opportunities offered by smartphones – such as eschewing costly price discounts in favor of recommendations and social marketing - to act on such insights.

The thesis addresses these three main problems as follows: it proposes data-driven models which allow retailers to identify at high resolution - for each individual customer and for each in the categories which they shop – (1) the customer’s inclination for variety seeking and the (2) CLV in these categories. The data-driven models were estimated and evaluated on a year of point of sale data from a physical grocery retailer partner, consisting of 150,000 unique receipts covering a total of two million transactions. Complementing these methods, the thesis also (3) proposes and tests a social learning marketing method – dubbed sales velocity - which utilizes the opportunity offered by smartphone screens to provide alternative marketing methods to costly price discounts. The sales velocity effect was tested in this thesis in three online experiments.
The results in the thesis first demonstrate that variety seeking can be mathematically estimated within individuals at the level of the product categories where they shop, thus replacing the current cumbersome method of estimation via psychometric questionnaires. The result offers an objective mathematical comparison of variety seeking between individuals and between categories and is readily implementable. The result confirmed within-customer heterogeneity in variety seeking between categories. Secondly, the thesis demonstrates that current CLV models ignore the strong within-customer between-category heterogeneity in CLV. It was also shown that existing CLV models overestimate a person’s overall lifetime value, which obfuscates the fact, shown in the thesis, that even though a customer may not have a low CLV overall, he may have low or declining CLV in select categories. Taken together, the thesis showed that by ignoring heterogeneity, existing variety seeking and CLV models lack the necessary high resolution required for personalized marketing. Finally, the thesis showed that social learning in the form of sales velocity increases the likelihood of a purchase and helps retailers promote mid-tail products. Sales velocity is robust across four highly different product categories and stronger for hedonic and fast moving goods. The studies reveal practical guidelines on a field deployment of sales velocity, which can be integrated in a wide range of marketing tools, replacing price cuts.

For practitioners, the thesis presented and enabled the structured analytical steps, which the retailer can take to craft a high resolution marketing strategy. These steps include: a method of estimating variety seeking for every customer in every product category without the cumbersome method of psychometric measures; a method for evaluating and monitoring a customer’s CLV in a specific category and a method of social marketing that can help retailers promote lesser known products.

As a key contribution to theory, these results address analytical gaps that impede high resolution marketing. It confirmed within-customer heterogeneity and extended it to the previously unconsidered domain of variety seeking and CLV. The results complement previous studies which estimate variety seeking from psychometric questionnaires. The thesis’ CLV-category model also complements the current CLV-temporal framework. Furthermore, in developing the sales velocity concept, the observational learning literature was augmented with key insights from intertemporal choice theory. Finally, the results contribute to the methodological research in multi-channel marketing, which have identified analytics as one of the key research gaps.
Zusammenfassung


Verkaufsgeschwindigkeitseffekt wurde in drei Online-Experimenten für vier Produktkategorien erfolgreich getestet.


Als wichtigen Beitrag zur Theorie helfen diese Ergebnisse analytische Lücken anzusprechen, die das hochauflösende Marketing behindern. Sie bestätigen die Heterogenität von Kundenseite und
erweitern dieses Konzept auf das nicht herangezogene Gebiet der Abwechslungsinclination und des CLV. Die Ergebnisse ergänzen frühere Studien zur Abwechslungsinclination, die auf psychometrischen Fragebögen basierten, und ergänzen die CLV-Kategorie Modelle auf den aktuellen CLV-Zeit Rahmen.
1. Introduction

It is widely accepted in online E-commerce that highly granular marketing – in the form of timely and personalized offers - is more effective than applying the same marketing policy to all customers (Changchien et al. 2004; Jiang and Tuzhilin 2006). Highly granular marketing, such as personalized coupons or recommendations, enjoys improved relevancy to the customer’s individual needs (Sheth and Parvatlyar 1995), leading to greater satisfaction for customers (Gummesson 2002; Liang et al. 2007; Xiao and Benbasat 2007). It is well established that satisfied customers are more loyal and more likely to repeat purchases at the retailer (Mittal and Kamakura 2001; Olsen 2002; Shankar et al. 2003; Venkatesan and Farris 2012). An additional benefit for the retailer is that with a highly granular marketing approach, since individual customers have different long term value to the retailer, the retailer can target specifically customers of value according to their business strategy (Kumar et al. 2006), either to further develop loyal segments or prevent potential switchers from defecting to another retailer (Gedenk et al. 2010; Wang and Hong 2006). Increasing the effectiveness of a promotion through highly granular marketing would be particularly useful for high volume, low profit industries such as physical grocery retailing, where the profit margins are low, typically in the 1% to 3% range in the USA (Singh et al. 2006).

In the domain of physical grocery retailing, for the retailers to also benefit from highly granular marketing, the retailers need the capability to (1) collect customer transaction data, (2) transform the data into insights and (3) act on the results with individual level marketing (Arora et al. 2008; Neslin et al. 2006; Oliver et al. 1998; Sheth et al. 2000). In the past, with loyalty cards and customer relationship management (CRM) systems, the retailers had the means to collect the data; however with only direct mail available, they lacked a timely method for operationalizing the results (Gedenk et al. 2010). The standard practice is thus one of mass-marketing, such as via weekly discounts (Ailawadi et al. 2001; Gedenk et al. 2010) or non-reactive forms of rewarding customer purchases, such as loyalty card bonus point collection schemes (Kumar and Shah 2004; Kumar et al. 2006). In spite of the mass-marketing norm, there is evidence, as follows, that customers are already accustomed to multichannel retailing for product informational search, and thus a natural extension would be personalized marketing via smartphones. Past research noted that the previously separate channels of E-commerce and physical stores are converging (Enders and Jelassi 2000), into what are now multichannel “bricks-and-clicks” stores (Gulati and Garino 2000); i.e. stores that have both an online and offline
presence. The multiple channels enabled customers to obtain product information in one channel and make purchases in another (Dholakia et al. 2010). In particular, customers often browse in one channel and buy in another (Verhoef et al. 2007) – with most of the purchases still ending with an in-store purchase. The other two patterns – (1) a purely online search and purchase and (2) “showrooming“, where one examines an item in-store and purchases it online at a lower price - are not common as common (Verhoef et al. 2007). The online channel enables lower search costs for finding information about products and enabling information search even when the store is closed, while in-store, customers are able to experience and evaluate products with all senses and buy the product immediately (Neslin and Shankar 2009; Zhang et al. 2010). These multichannel interactions have been shown to benefit firms - even a simple non-customized multichannel marketing campaign – a physical store having an informative website – can increase in-store sales even in the absence of a customization campaign (Avery et al. 2012; Verhoef et al. 2007). Furthermore, customers which utilize different channels have more contacts with the firm and provide higher revenues, higher share of wallet, have higher past customer value and have a higher likelihood of being active than other customers. (Kumar and Venkatesan 2005; Neslin and Shankar 2009; Thomas and Sullivan 2005). Thus, given the proven economic benefits and the customer acceptance of a non-customized “online information search, in-store purchase” shopping pattern, a personalized extension with “smartphone individualized marketing, in-store purchase“ shows promise for retailers.

Additionally, with the increasing ubiquity of smartphones and tablets, it becomes for the first time economically and technologically feasible to operationalize personalized marketing also for physical grocery retailing via information systems such as retailer’s CRM – for example, by pushing personalized offers through their mobile apps directly to individual customers, or by personalizing the customer’s purchase decision tools (ex. Product search tools and recommendation agents) (Yadav and Pavlou 2014). The mobility and the personal nature of the mobile device distinguish it from other electronic devices such as the television and the personal computer; it enables direct, personalized and timely marketing, in forms such as mobile advertising and couponing, and decision aids (Shankar et al. 2010). This enables the benefits of highly granular marketing as proven in E-commerce - such as increased relevance of offers (Changchien et al. 2004; Jiang and Tuzhilin 2006) - to also reach in-store customers at a low cost. Additionally, smartphones open up forms of marketing previously unavailable in physical stores: in place of coupons and advertisements, which incur a high cost to the retailer for deployment and redemption (Shankar and Balasubramanian 2009), mobile devices enable personalized
recommendations, which are also effective in influencing customer decisions (van der Heijden 2006; Kowatsch and Maass 2010; Lee and Benbasat 2010). Recent studies show that retailers having branded apps can benefit from a positive persuasive impact, increasing interest in the brand and also the brand's product category (Bellman et al. 2011). These branded apps and websites could thus also act as a channel for personalized marketing. Accordingly, not only is it technologically feasible to deploy highly granular marketing, but smartphones make it economically feasible to deploy personalized offers, and open up forms of marketing previously not possible in physical stores, such as recommendations.

1.1 Problem Statement & Research Questions
As noted earlier, for grocery retailers to also benefit from highly granular marketing, the stores need the capability to (1) collect data about customer transactions, (2) transform the data into insights and (3) operationalize the results (Arora et al. 2008; Oliver et al. 1998; Sheth et al. 2000). In spite of the identified technological and economic readiness, and the promise offered by smartphones and tablets in enabling finer grained marketing, analytical and methodological challenges remain (Neslin and Shankar 2009). As identified in the multichannel customer management decision (MCMD) framework (Neslin and Shankar 2009), a multichannel strategy consists of multiple stages: analyzing the customers, developing a multichannel strategy, designing channels, implementing the marketing and evaluating the results. The operationalization of the MCMD framework in the context of smartphone-enabled marketing is depicted in Figure 1.
In the context of the MCMD framework for smartphone-enabled marketing, there remains a research gap in analyzing customers, conducting direct marketing and evaluating the results at a high granularity. These gaps, which are the central foci of this thesis, arises as a result of smartphones enabling marketing timely and personalized marketing which has been ignored in traditional mass marketing. Traditional marketing from physical stores typically consist of mass-marketed flyers or coupons mailed out weekly, on the assumption that customers as a population have uniform purchasing preferences across product categories (Kumar 2010). However, it is also known that there is heterogeneity in customer buying preferences for different product categories (Allenby and Rossi 1998; Leszczyc and Bass 1998; Lim et al. 2005). Due to this heterogeneity, the current mass market practice of giving all customers the same promotion (typically a price cut) entails wasted potential, since not every customer will be interested in a mass-marketed offer, or even worse, the mass market offer may motivate certain customers to buy when an offer is given (Mela et al. 1997; Zeelenberg and Putten 2005). Both of these outcomes could be avoided with a highly granular, personal marketing campaign; an online channel can recommend or promote the right product from the right category given customers’ individual category preferences and the firm’s CRM strategy towards that customer (Albadvi and Shahbazi 2009), thus maximizing the benefit to the customer and the firm. This requires, however, an appropriate analytical method of customer behavior in-store. Past models which segment customers (needed for identifying customer groups for personalization) (Bawa
1990; Kim et al. 2006; Konus et al. 2008; Trivedi 1999; Verhoef et al. 2002) and past models which evaluate the profitability of customers (needed for marketing evaluation) (reviewed in Fader & Hardie (2009) are at the level of customers, designed for the mass market age, and thus do not examine customers’ category level preferences and their category-level responses to marketing attempts. Given customers’ heterogeneous tastes between a retailer’s different product categories, and the technological means to act on these tastes, there is a need for models and methods which can identify, market to, and evaluate these customers. Since new technologies have already enabled efficient data collection, the focus of this thesis lies enabling the transforming of data to insights, operationalizing the results, and enabling its evaluation. All of these three aspects require methods to enable their in-store reality. At a high level, the main research question is thus:

*How can high-resolution multi-channel marketing be enabled for physical grocery stores?*

This question can be broken down into sub-issues, discussed in the following subsections. It is the goal and contribution of this thesis to address the following issues with the aim that the presented solution can be transformed into insights and operationalized into marketing results. In this manner, this thesis aims to be actionable, with an emphasis on providing methods or results which could be easily implemented by practitioners.

### 1.1.1 How Retailers Could Segment Customers in Multichannel Grocery Retailing

The first core issue concerns how the retailer should segment customers in a multichannel environment (Dholakia et al. 2010; Konus et al. 2008; Neslin and Shankar 2009; Payne and Frow 2005). Segmentation schemes help retailers identify customers with heterogeneous preferences, and thus form the basis of finer grained marketing (Dickson and Ginter 1987). At the finest granularity – that is a segment that consists of only one customer – there is complete marketing personalization, which is now enabled by E-commerce and smartphone technology (Arora et al. 2008; Changchien et al. 2004). This thesis will thus address segmentation at a single customer level as a minimal level of granularity.

Irrespective of what granularity of segmentation, there is the question of what measures should the segmentation be based on. That is, on what behavioral or psychological basis should the segmentation scheme be based on? A good segmentation scheme requires that the segments be measurable, accessible, differentially responsive, actionable and substantial (Kotler and Keller 2006). Segmentation schemes can be divided between those which are questionnaire driven or data driven. The tradition of psychometric questionnaire measures in physical retailing
marketing research is well established; frameworks and measures exist for different dimensions of customers’ shopping orientation, such as their price-quality orientation (Fisher et al. 2002; Garretson et al. 1998; Kerin et al. 1992), their price sensitivity (Dodds et al. 1991; Wakefield and Inman 2003), their openness to useful vs. hedonic goods (Voss et al. 2003) and their extent of variety seeking (Baumgartner and Steenkamp 1996; Van Trijp et al. 1996). The majority of these frameworks and measures have been empirically verified against purchase intentions and in some cases, real purchase data (Baumgartner and Steenkamp 1996) to evaluate their predictive validity. Similarly, recent research in information systems (i.e. mobile shopping aids) have incorporated the influence of customer beliefs on behavior (van der Heijden 2006; Kamis et al. 2008; Lee and Benbasat 2010), as measured by questionnaire-based constructs. Recent segmentation research for multichannel research are also questionnaire based, and focus on a customer’s different interactions between channels, i.e. how often they use a channel, their intrinsic preference for a channel, and their response to marketing actions between channels (Dholakia et al. 2010; Konus et al. 2008; Neslin et al. 2006); the channel itself becomes the unit of analysis. In spite of the ubiquity of questionnaires in both research and practice, the implementation problems with regards to achieving true personalization are recognized, as evident by the body of research dealing with non-participation: these range from survey length reduction (Bergkvist and Rossiter 2007; Childers and O. C. Ferrell 1979) to methods of analyzing unanswered questions (Bosnjak et al. 2005; Porter 2004). As noted by Nulty (2008), participation rate is roughly 30% for online surveys. Thus they are not easy for practitioners to implement with minimal overhead, and has been in the past only possible to sample from the overall population, rather than identify the preferences of every single customer and address them individually.

Meanwhile, data driven schemes are typically based on Point of Sale (PoS) loyalty card data widely available for the retailer. These schemes segment the customers’ purchasing behavior within a physical store and have the advantage of being easily measurable, accessible, differentially responsive with respect to a segment-of-one point of view, but being designed before the smartphone-multichannel age, they segment customer behavior at a store level (Bawa 1990; Kim et al. 2006; Konus et al. 2008; Trivedi 1999; Verhoef et al. 2002) – that is, an individual’s overall purchases within a store. However, even within a store and within a single customer, it is acknowledged that customers can behave differently and have different preferences at the level of product categories, due to heterogeneity in product preferences (Allenby and Rossi 1998; Leszczyc and Bass 1998; Lim et al. 2005). Thus, one challenge is to
develop models which account for customer behavior and preferences at the level of categories, which are known to vary within a person and thus should be accounted for targeting in a high resolution marketing scheme.

A data-driven segmentation approach is preferable to a questionnaire approach, since it can then be operationalized into CRMs which retailers already possess. It can then be automated on widely available CRM data, and marketing actions such as recommendations can then be automatically delivered through smartphones. Addressing this, the research question becomes:

\[ \text{RQ1: How can a relevant high resolution segmentation scheme be developed for multichannel marketing for physical grocery stores?} \]

“High resolution” means in this context an individualized or one-to-one segmentation scheme where even a customer’s different preferences between product categories can be identified.

1.1.2 How Retailers Could Leverage Marketing Methods in One Channel to Enhance Performance in Another

The second core issue concerns how to leverage the advantages of one channel to enhance the performance of another; frameworks have been proposed, but empirical evaluations have been lacking (Neslin and Shankar 2009; Thomas and Sullivan 2005; Verhoef et al. 2007). In this context, this would be how retailers can leverage the possibilities provided by smartphones, tablets and websites to drive in-store purchases. For example, E-commerce retailers (where both marketing and purchases are online) have utilized to great effect lists of top ranked products to promote product sales; the higher the sales rank, the more likely customers buy that product (Chen, Wang, et al. 2011; Duan et al. 2009; Tucker and Zhang 2011). This influence to buy, based on observing what others bought is known as observational learning OL, a form of social learning (Bikhchandani et al. 1998). This use of social learning as marketing is an example of leveraging the strengths particular to that channel; the ranked lists, which need to be rapidly computed and updated, are easy to implement online, but difficult to implement in print in physical stores. Furthermore, while discounts incur a per-use cost to the retailer, the marketing method of social learning persuades on information alone and do not carry this disadvantage. Now, with smartphones, social learning has the potential to benefit physical stores, presented as information when customers browse products, or as information in targeted ads.

Given its aforementioned benefits already found in the online-to-online E-commerce channel, accordingly, this thesis focuses on investigating how to leverage the social marketing method of
observational learning for benefiting physical stores. Additionally, because unprecedented, the potential impact of leveraging this social learning method for physical grocery retailing is evaluated. Therefore, the question is:

*RQ2: How can social learning marketing be developed for benefiting physical grocery stores?*

### 1.1.3 The Need for High Resolution Customer Relationship Management

The previous two research questions covers how to segment customers for a multi-channel marketing campaign in physical stores, and it also examined how to leverage social marketing in this setting. Closing the loop on this would be evaluating the effectiveness of the multichannel marketing strategy (Neslin and Shankar 2009; Zhang et al. 2010), in particular in evaluating how they affect the long-term profitability of a customer to a retailer (Verhoef and Venkatesan 2010). For evaluating the long-term profitability of individual customers for a given retailer, a popular and widely accepted evaluative metric used in CRMs is the customer lifetime value (CLV) or the lifetime value (LTV) metric (Fader and Hardie 2009; Gupta et al. 2006; Jain and Singh 2002; Kumar et al. 2006). It is defined as the present value of the future cash flows associated with the customer (Pfeiferis et al. 2004), factoring the cost of retention. As a forward facing metric that is computed from - but distinct from - historic profitability, CLV thus holds promise as an evaluation tool of marketing actions frequently and rapidly deployed via online channels. In line with this idea, recently Kumar (2010) proposed frameworks – but not an actual model - which incorporate CLV for evaluating the effectiveness of multichannel marketing. There is also a lack of CLV models which factor in the finer grained marketing enabled by smartphone online marketing - current CLV models are applied at the level of the customer’s total purchases in the retailer (Fader and Hardie 2009), which therefore ignore customer’s heterogeneity in buying preferences for different product categories, and by extension, different lifetime value towards a retailer’s different product categories. Thus, in addition to the overall change in a customer’s profitability to the retailer, a CLV analysis should factor in improvements in a customer’s profitability corresponding to the different product categories where the promotion was given, and be scalable with time. Hence, the final research question becomes:

*RQ3: How can the effectiveness of high resolution marketing actions be evaluated, given customers’ heterogeneous preferences and the capability of retailers to act on these preferences in a timely and precise manner?*
1.2 Research Methodology

To answer the research questions raised in this thesis, three quantitative empirical studies have been conducted.

The first study proposes a data-driven customer segmentation approach that can identify variety seekers, a customer base shown to be valuable in previous questionnaire-based research and of relevance to physical grocery retailing. The model allows identification of the customers’ overall extent of variety seeking as well as their specific variety seeking at a category level. The study will show that the approach is easy to operationalize and provides insights at a level of granularity not available to past questionnaire based research, thus enabling the large-scale deployment of personalized marketing measures in mobile commerce in physical stores. The model is estimated using a year of Point-of-Sale data from a European retailer partner.

Closely linked to the idea of the first study, given that customers seek variety in their purchases in the domain of grocery retailing, one question is how to promote products to this group, using the strength of the online channel to boost in-store sales. Accordingly, the second study proposes a method of leveraging a previously ignored dimension of observational learning, namely the “sales velocity”. Sales velocity describes how quickly a product is gaining in popularity, and like traditional observational learning in E-commerce, can be deployed and presented at low cost to customers. The study presents the theory that lead to the development of the idea of sales velocity, and consists of three customer choice online experiments to establish and confirm its effect across different product classes and in different circumstances. The study will show the effect of sales velocity and provide insight under what conditions it is effective as a marketing tool. The study also shows using a year of Point-of-Sale data from a European retailer partner for which products sales velocity can market effectively, and guidelines for implementing sales velocity in the retailers’ channels are also presented.

Finally, the third study presents a method that would allow physical retailers to evaluate the effectiveness of their marketing strategies at a high granularity, particularly suitable for the increased time and product resolution afforded by smartphone-delivered marketing. In particular, the study builds on research that quantifies and models the long-term profitability of a customer, by extending it to a product category level per person. The choice of customer long-term profitability as the evaluation metric allows retailers to determine to which customers would give the best return on their marketing investment. Using this method, the model is able to show which customers are profitable for different product categories, thus allowing the
evaluation of the effectiveness of highly grained marketing action at the category level. The results also enable retailers in identifying product categories where customers have the risk of defecting, thus enabling intervention measures. The result is further able to characterize how consistently a person is profitable across categories, which, combined with knowledge of their overall long-term profitability, results in typologies for retailers’ loyalty development strategies. Finally, the study will show that by applying the model at regular time intervals, it is possible to characterize whether a marketing action is leading to an improving or decreasing trend in customer profitability.

In summary, this dissertation provides methodologies which, taken together, help enable high-resolution multichannel marketing. It provides and evaluates (1) a method for transforming high-resolution customer behavioral data into insights about individual customers, (2) a method for leveraging these insights into social marketing and (3) a method for evaluating these high-resolution marketing actions. These methods are derived from three quantitative empirical studies. After every study, conclusions and recommendations for practice are given, which aim to help practitioners and researchers implement high resolution marketing schemes with already existing technology. The insight from this thesis aims to fill the methodological and analytical gaps.

1.3 Structure of the Thesis

The structure of the presented dissertation is summarized in Figure 2. The introductory Chapter 1 motivates the objectives and research questions addressed in the thesis. It summarizes the research processes and the studies used to address the research questions, and outlines the thesis structure. Chapter 2 presents the state of the art in analysis and marketing regarding the three open problems investigated in this thesis. The subsequent three chapters (Chapter 3, 4 and 5) then present three studies which each examine an open problem in enabling high resolution marketing today, and presents and evaluates a solution for each case. Chapter 3 presents and evaluates a data-driven method for segmenting customers based on their extent of variety seeking. The model is estimated on a year of real-world transaction data from a grocery retailer partner. Chapter 4 presents a form of social learning marketing, the sales velocity, and establishes its effects and limitations in three customer choice online experiments. Chapter 4 also shows for which products could sales velocity market effectively, by analyzing a year of transaction data from a grocery retailer partner. Chapter 5 presents a model that would allow physical retailers to evaluate the effectiveness of their marketing strategies at a high granularity. The contribution builds on research that quantifies and models the long-term profitability of a
customer, and the model is estimated from a year of real-world transaction data from a grocery retailer partner. Chapter 6 concludes this dissertation by summarizing the findings of the three studies and answering the research questions raised in Chapter 1.

Chapter 1: Introduction
Problem and Research Questions
Research Methodology
Structure of the Thesis

Chapter 2: Related Work
Behavioral Segmentation of Customers in FMCG
Leveraging Social Learning Marketing for Boosting Sales
Customer Relationship Management in the Pre-Smartphone Age

Chapter 3: Towards High Resolution Identification of Variety Seeking Behavior

Methodology: Probabilistic modeling estimated on real-world transaction data

Q1: How can a relevant high resolution segmentation scheme be developed for multichannel marketing for physical grocery stores?

Chapter 4: Leveraging Social Learning to Market Lesser-Known Products: The Sales Velocity Effect on Retailing

Methodology: Online empirical studies

Q2: How can social learning marketing be developed for benefiting physical grocery stores?

Chapter 5: Towards High Resolution Evaluation of Customer Profitability

Methodology: Probabilistic modeling estimated on real-world transaction data

Q3: How can the effectiveness of high resolution marketing actions be evaluated?

Chapter 6: General Discussion and Conclusions
Motivation and Summary of Thesis
Key Findings and Implications of the Thesis
Implications for Practitioners and Managers
Limitations and Future Research
1.4 Statement of Prior Published Work

Parts of this dissertation have already been published by myself and colleagues as scientific articles in peer-reviewed journals or in conference proceedings; other parts are currently being prepared for publication in additional outlets. As a result, some sections of this thesis correspond literally to work previously published by me or bear strong similarities.

While I am the first author of all of these documents and I declare that the majority of the content incorporated into the thesis has been written by myself, other coauthors have contributed to scientific papers with their reviews, changes, suggestions and edits.


Chapter 3 is an extension of the paper “Towards High Resolution Identification Of Variety-Seeking Behavior”, published in the 2014 European Conference on Information Systems, authored by myself and Alexander Ilic. The methodology, modeling and figures have been adapted to the thesis with minor modifications; the introduction has been re-written and new additional results are given, particularly in the analysis of customer variety seeking heterogeneity and its patterns over time, and the discussion is expanded.

Chapter 4 is adapted from the paper “The Sales Velocity Effect on Retailing”, published in the Journal of Interactive Marketing, authored by myself, Tobias Kowatsch and Alexander Ilic. The methodology, description of the sub-studies, results and figures were integrated with minor modifications, with the introduction, discussion and implications for retailers expanded on.

Chapter 5 is adapted from the work under review “Managing Customer Profitability – One Category at Time”, authored by myself, Elgar Fleisch and Alexander Ilic. The methodology, modeling and figures have been adapted to the thesis with minor modifications; the analysis, discussion and implications for retailers have been expanded.
2. Related Work

This chapter outlines the theoretical background of the thesis. The sections correspond to the three core problems addressed by the thesis, namely, in analyzing customers, conducting direct marketing and evaluating the results at a high granularity.

The first section 2.1 describes past attempts both online and offline on segmenting and addressing customers at a higher granularity, which historically has been leveraged to better address customers with customized marketing campaigns, products and recommendations. It begins by presenting literature on psychological segmentation, which has been used by marketing researchers and market research firms for determining common behavioral characteristics of customer groups. It then presents information on recommender systems, a technique popular in E-commerce with specific paradigms in identifying customer preferences. The remaining subsections present and discuss the attractiveness of the variety-seeking customer base for FMCG and methods for identifying them. In particular, the sections discuss how a data-driven variety seeking model would complement both psychometric segmentation research and recommender systems research.

The second section 2.2 focuses on social marketing, which has mostly been in E-commerce channels. In particular, the method of observational learning is introduced and the section discusses its advantages in comparison to alternative forms of social marketing, such as word-of-mouth (WoM). The section examines a research gap in observational learning and proposes, based on theory, an extension that can be appropriate for FMCG.

The third section 2.3 discusses customer management techniques in the pre-smartphone age. The section presents relevant issues to researchers and practitioners and introduces research in the area of customer relationship management systems (CRM). The research gaps in this area are introduced and particular focus is given to evaluative metrics in CRM. In particular, the customer lifetime value (CLV) metric is elaborated on and past models are presented. The section presents current gaps associated with CLV models.

2.1 Towards One-to-One Behavioral Segmentation of Customers

Segmentation schemes help retailers identify, from a population of customers with heterogeneous preferences, groups who nonetheless have similar characteristics within the group, and thus form the basis of finer grained marketing (Dickson and Ginter 1987). A good segmentation scheme requires that the segments be measurable, accessible, differentially responsive, actionable and substantial (Kotler and Keller 2006) – “actionable” meant that in the
past, an individualized coupon mailing campaign tended to be at the level of segmented groups, due to the cost of individualizing a coupon to every customer. With the advent of E-commerce portals, individualizing promotions becomes actionable (Changchien et al. 2004), allowing the operationalization of “one-to-one” or “segment of one” marketing (Arora et al. 2008) – that is, a completely individualized and matching offer given a person’s unique tastes (Dickson and Ginter 1987). It can be seen as an extreme form of marketing, where the target segment is of size one (Arora et al. 2008). Such higher granularity in marketing is widely considered to be more effective than mass marketing (Changchien et al. 2004; Jiang and Tuzhilin 2006; Sheth et al. 2000; Thomas and Sullivan 2005).

Regardless of the granularity of the segment, a segment has to be based on some concept of customer preference. Typically the basis of segmentation schemes were measured from psychometric measures administered by questionnaires, and more recently – as seen in E-commerce, past purchasing behavior from actual customers. The state-of-the-art in segmentation is in contrast to earlier marketing segmentation research (Haley 1968), where demographics information was popular; current research “strongly suggests that past purchases of consumers are better predictors of future purchase behavior than demographics” (Gupta et al. 2006, pg. 142) and in line with this, literature has shown that “demographics alone predict 1.2% of decisions to buy or not buy and only 0.3% of the variance in the number of purchases made by online buyer’s” (Bellman et al. 1999, pg. 37) and that similarly for offline, in-store environments the effect of observable demographics was shown to be insignificantly related to store choice (Rhee and Bell 2002) and for decisions in repeated purchases (Hoyer 1984).

As such, this section presents the different methods of estimating customer preferences in FMCG physical stores and in the E-commerce domain, with an emphasis on psychometric measures and E-commerce methods estimated from actual purchase data, rather than on demographics.

2.1.1 Identifying Customer Preferences by Psychometric Questionnaires

The tradition of psychometric questionnaire measures in physical retailing marketing research is well established; paradigms and measures exist for different dimensions of customers’ shopping orientation and on its subsequent effect on purchases. Accordingly, the ability to identify these customer preferences can help targeted marketing.

One well-cited paradigm is that of a customer’s price-quality orientation; past research has been able to identify customers who associate high prices with high quality, and thus a marketing
A campaign that appeals to these customers’ perception of quality can lead to higher prices paid for the product and vice versa (Fisher et al. 2002; Garretson et al. 1998; Kerin et al. 1992). For example, Garretson et al. (1998) developed a scale that measured American consumers’ attitudes toward national label brands, which are perceived as being more expensive and having higher quality than store-label brands, and are more appealing to customers who are identified as being deal prone. Similarly, Fisher et al. (2002) discovered that those who associate high prices with high quality have a stronger intention to purchase the national brand. Accordingly, by identifying those willing to pay more for quality, retailers are able to position their product assortment to target these groups; similarly, this knowledge can affect the marketing of such products.

In the opposite end of the spectrum, psychometric questionnaires have also been able to identify customers who are price sensitive – customers who actively search for the lowest price in shopping (Ailawadi et al. 2001; Lastovicka and Bettencourt 1999). Mittal (1994) found that demographics were a poor predictor to deal-proneness, thus motivating the need and insight of a psychometric scales for segmenting and predicting the preferences of customers. Lastovicka & Bettencourt (1999) show that price sensitivity, which they call frugalness, as a distinct personal characteristic identifiable by questionnaire measures, and building on this, Rick, Cryder, & Loewenstein (2008) show that customers identified as highly price conscious are averse to spending money on more expensive goods. Researchers have investigated how these identified characteristics have an effect on specific marketing methods: Lichtenstein et al. (1997) found that depending on the person, price sensitivity can manifest itself as discount proneness and/or receptiveness to coupons; although both are driven by price sensitivity, the effectiveness between these two marketing methods differs even within users, and the inclination to the different marketing methods can be identified by questionnaires. Further research have devised multi-item psychometric scales to measure specifically proneness to discounts (Wakefield and Inman 2003) and coupons (Colombo et al. 2003; Lichtenstein et al. 1997; Mittal 1994; Swaminathan and Bawa 2005). These results show how even customers who are similar by the psychometric characteristic of price sensitivity may differ even in their receptiveness to specific deals to latent aspects one level behind price sensitivity, measurable only by these psychometric questionnaires. Accordingly, identifying customers by price sensitivity and furthermore by their inclination towards certain promotion methods allows finer grain marketing.

Another famous paradigm is the hedonic paradigm; a hedonic product is one that “relates to the multisensory, fantasy and emotive aspects of product usage experience” (Hirschman and
Holbrook 1982; Holbrook and Hirschman 1982) – that is, there is an affective component from consuming the product. Examples of such products in the area of grocery retail for example are chocolate, mixed nuts and bubble bath (Chandon et al. 2000). It has been shown in the benefit-congruency framework by Chandon et al. (2000) that the extent of perceived hedonism in the product affects how customers respond to the type of sales promotions (which they classified as utilitarian – ex. price cuts - or hedonistic like free gifts); the promotion is most effective when the promotion and product type are matched in hedonism. Therefore, a product’s extent of hedonism influences the type of promotion that is most effective for it. Since whether a product is hedonic or not is a matter of subjective perception from the customer, psychometric measures exist to evaluate this perception (Voss et al. 2003). Knowing how a customer perceives particular product classes would thus enable the optimization of the promotion type for that customer.

Research has also been conducted to define a customer’s extent of variety seeking (Baumgartner and Steenkamp 1996; Van Trijp et al. 1996). Variety seeking is defined as the “biased behavioral response by some decision making unit to a specific item relative to previous responses within the same behavioral category, due to the utility inherent in variation per se, independent of the instrumental or functional value of the alternatives of items” (Van Trijp et al. 1996). In other words, in addition to external circumstances (i.e. external situations like discounts, reviewed in McAlister and Pessemier (1982), or in stock-piling for future preference uncertainty, as reviewed in (Kahn 1995)), there is an innate personal characteristic that defines the extent of variety seeking, and that this degree of variety seeking is related to the product category. Variety seeking is particularly appropriate for grocery retailing since customers have strong habits (Hoyer 1984; Liu-thompkins and Tarn 2013; Wood and Neal 2009) and thus do not change their purchase pattern from week to week, until they have become satiated (Chintagunta 1999; McAlister and Pessemier 1982) and thus seek new tastes and experiences (Van Trijp et al. 1996). Identifying variety seekers would identify candidates for new product offers or for trying out product alternatives.

As can be seen, there are a varied number of psychometric paradigms on consumer behavior; the aforementioned review is by no means exhaustive, as the number of paradigms range in the hundreds, as reviewed in Bruner et al. (2005). The majority of these frameworks and measures have been empirically verified against purchase intentions and in some cases, real purchase data (Baumgartner and Steenkamp 1996) to evaluate their predictive validity. In spite of the ubiquity of questionnaires in both research and practice, the implementation problems of questionnaires are recognized, as evident by the body of research dealing with non-participation: these range
from survey length reduction (Bergkvist and Rossiter 2007; Childers and O. C. Ferrell 1979) to methods of analyzing unanswered questions (Bosnjak et al. 2005; Porter 2004). As noted by Nulty (2008), participation rate is roughly 30% for online surveys. As such, from an information systems point of view, there is a motivation for having a data-driven method of replicating select insights of the questionnaire measures, not only from the perspective of enabling high resolution marketing, but also for efficient data collection.

2.1.2 Identifying Customer Preferences for Recommender Systems

In E-commerce, a major class of customer purchase preference identification arises from the field of recommender systems (RS). Recommender systems are information systems that identify individual customers’ product interests and preferences, either explicitly or implicitly, and make a product recommendation accordingly (Xiao and Benbasat 2007). The highly personalized aspect of RS makes it a form of “one-to-one” or “segment of one” marketing (Arora et al. 2008) – that is, a completely individualized and matching offer given a person’s unique tastes (Dickson and Ginter 1987). Many RS attempt to offer customers products which they have not yet tried (the “cross-sell” outcome) (Adomavicius and Tuzhilin 2005; Schafer et al. 1999), an appeal to the customer’s desire for variety. RS have been deployed mainly in the area of E-commerce (rather than physical grocery stores) and generate recommendations based on the content of what customers have bought or what others have bought – this is the so called content-based vs. collaborative-based paradigm in RS (Adomavicius and Tuzhilin 2005; Xiao and Benbasat 2007).

Content-based systems has its roots in information retrieval and information filtering research (Baeza-Yates et al. 1999; Belkin and Croft 1992), which study algorithms which remove redundant information irrelevant to a given user. It relies on the customers’ purchase histories and the characteristics (represented as meta-data) of the bought products to infer the right recommendation - typically, recommending products with similar characteristics to ones bought frequently. Formally, for an item not yet bought by a customer, its utility for that customer is assigned based on the utilities of similar products already bought by that customer. When recommending a product, the system makes a recommendation that maximizes the utility to the customer. The attributes and characteristics (“content”) which form the basis of determining similarity between products can be data with predefined attributes or unorganized text (ex. In a product description), and thus one enabling research for content-based systems is extracting keyword features (Adomavicius and Tuzhilin 2005; Pazzani and Billsus 1997). Assuming the content is known, the problem becomes one of learning customer preferences as they iteratively
make product choices, and - in terms of recommending not-yet-seen products – optimizing the
match of recommended products. Thus, in contrast to the psychometric methods, which try to
recognize customer’s innate mental states and inclinations, a content-based recommendation
bases “preference” from purchase and consumption frequency of product attributes. Although
powerful, the content-based methods suffer from three recognized problems (Adomavicius and
Tuzhilin 2005). First, there is a dependency on how well the attributes are defined or are
available for automatic feature estimation; poor quality data or low data granularity regarding
products inhibit good estimation. Second is the problem of overspecialization; without some
randomness, a system cannot recommend items that are different from anything the user has
seen before – it is desirable for a recommendation system to introduce some degree of diversity
(Zhang et al. 2002). Finally, it suffers from the “cold start” or “new user problem”, where a user
has to rate a sufficient number of items before the recommendation system could sufficiently
understand the users’ preferences and thus present the user with reliable recommendations
(Adomavicius and Tuzhilin 2005; Rashid et al. 2002).

Meanwhile, collaborative-based systems (also known as collaborative filtering systems)
recommend products based on the purchasing histories of other customers who bought similar
products (for example, recommending a product that others buy frequently and is similar to
what you buy) (Shardanand and Maes 1995). It recommends products bought by people with
similar interests to that customer. Formally, for an item not yet bought by a customer, its utility
for that customer is assigned based on the utilities assigned (ratings) by similar customers. The
collaborative-based system overcomes the problem found in content-based systems of
recommending only products which are similar to those seen by that customer in the past since
the collaborative systems recommendation is based on customer similarity rather than item
similarity, they can recommend any item, not just items similar to the ones seen before by a
customer. This method has some similarities to the psychometric methods of segmented
marketing. In both cases, the commonalities between customers are the basis for a
recommendation or offer, but past psychometric methods make a top-down offer to that
segment based on prior theory as to what offer would appeal to this group, while the
collaborative filtering recommends a product without prior theory. While powerful,
collaborative-based systems are not without their flaws (Adomavicius and Tuzhilin 2005;
Balabanović and Shoham 1997). Similar to content-based recommender systems, in order to
determine user-to-user similarity, the system must learn the user’s preferences from the ratings
he gives, which would be scarce for a new user. Additionally, collaborative systems also suffer
from the new item problem; a new item that has not been rated by a substantial number of users would not be easily recommended.

Research which addresses the shortcomings found in content and collaborative based systems are the hybrid approaches, which combine collaborative and content-based methods (Albadvi and Shahbazi 2009; Balabanović and Shoham 1997; Burke 2002; Soboroff and Nicholas 1999; Zhou et al. 2010) in an attempt to leverage the strength of one method to cover for the weaknesses of the other. For example, the new item problem in collaborative recommender systems is solved when combined with a content-based system; as long as meaningful meta-data exist to describe the items, then even a new item not often rated can be recommended based on its content. Similarly, an item that is different content-wise from what a given customer typically buys can still be recommended if combined with a collaborative approach.

In summary, given the content-collaborative paradigm, customer preferences are identified from item attributes of what they buy and/or from the preferences of similar customers. Current research in RS focuses on customer profile building and expanding the set of product characteristics which RS algorithms can infer customer preferences (Lops et al. 2011), to include also user-generated (i.e. collaborative) content like social tags (Milicevic et al. 2010; Shepitsen et al. 2008). It should be noted that these methods, unlike the psychometric based methods, make no prior theoretical assumption about a customer’s purchasing inclination – they try to estimate what products a customer might like, but as a methodological and research gap, they ignore these customer inclinations, such as in the area of variety seeking or price sensitivity. The contribution of a data-driven model which could identify inclinations could thus complement the current RS research on customer profile building.

2.1.3 Marketing and IS Motivations for Studying Variety Seeking Behavior

Out of the different possible questionnaire measures that is desirable to replicate in an information system (ex. such as an RS), the focus of this thesis is on variety seeking behavior. Given the large number of psychometric paradigms, it beyond the scope of the thesis to replicate them all in an information system. As discussed previously, variety seeking is a characteristic innate to a customer. This innate variety seeking has implications on other aspects of customer behavior that is of relevant to retailers. For example, it was shown that customers who pursued higher variety tend to increase overall consumption quantity (i.e. variety consumption has an additive rather than a substitutive effect) (Kahn and Wansink 2004; Read et al. 1995; Simonson 1990). Variety seekers also tend to be open to promotions (Ailawadi et al.
2001). As such, variety seekers form a potentially valuable group for individual segmentation and marketing initiatives delivered through smartphones. Identifying variety seekers would also be valuable to recommender systems (RS): it would complement current RS research which looks beyond the content-collaborative paradigm for additional input into a RS.

2.1.4 The Unit of Analysis for Variety Seeking

In terms of the unit of analysis for variety seeking, until recently, past marketing studies have focused on variety seeking as a characteristic trait, independent of product categories (reviewed in Steenkamp & Baumgartner (1992)). The study by Trijp, Hoyer & J. Inman (1996) was an early attempt in developing a psychometric scale that defined the customer’s extent of variety seeking with respect to a particular product category; however, since a typical grocery retailer can have several hundred distinct product categories, it is not yet possible to practically administer this questionnaire for every single product category for every single customer. Nor has it been practically possible in the past to be able to take individual action at the product category level with this knowledge. Indeed, past empirical studies of variety seeking have limited their focus on arbitrary and often merely convenient samples of product purchase data available to the researchers; as such their insights on individual levels of variety are arguably not generalizable across all categories. For example, the study by Kahn et al. (1986) examined sandwich bags, wraps, margarine, cereals and soft drinks while the study by Trijp, Hoyer & J. Inman (1996) focused on beer, coffee, hand rolled tobacco and cigarettes, and finally the study by Trivedi (1999) focused on hypothetical purchases of cola. These could hardly be called representative samples of a customer’s product space; indeed, there is an acknowledged research gap in investigating product-category influences on variety seeking (Michaelidou and Dibb 2009; Roehm and Roehm 2004; Tang and Chin 2007). This gap becomes important with smartphones, since it is now possible to analyze and take marketing action at the granularity of customers and product categories. In line with this, the approach in the thesis will characterize individuals’ degree of variety seeking for the different product categories in which he shops in.

2.1.5 Approaches in Enumerating Variety Seeking

This section discusses some existing approaches to numerically model variety seeking and their appropriateness for an information system context. This work is closest in application to the studies from Kumar & Trivedi 2006 and Trivedi (1999) (henceforth referred to as the “Trivedi model”, where a variety index was defined for individuals, which described the variation in products a person bought in a product category at a given time. This was then combined with
intensity – the number of purchases made in a product category – in order to identify customer typologies.

The Trivedi model is the state of the art in variety seeking research: in the search for models of variety seeking, a forward and backwards search was conducted for papers which cited the papers by Trivedi (1999) and Kumar & Trivedi (2006), and the models mentioned in the literature review by Kahn (1995). A further search was conducted for “variety seeking” and “variety seeking models” and other related keywords in the context of retailing, marketing and E-commerce; this search was applied within the “basket of eight” journals of information systems according to the Association for Information Systems¹, and within the top ten marketing journals defined in the analysis by Hult et al. (1997) and Steward & Lewis (2009). Articles were considered from 1991 to July 2014. It was found, among these basket of widely regarded journals, that variety seeking in the context of commerce has not been recently researched and thus there was a research gap in combining the insight of past marketing research on variety seeking (which was largely questionnaire based, and not studied at the category level) with today’s largely content-based recommender systems (which ignores individual differences in a consumer’s desire for variety).

However, in spite of its preeminence, there are two shortcomings with the Trivedi (1999) approach, which the thesis aims to overcome. The first is that the variety index comes from estimating a mathematical model that requires a customer’s stated preference of the different brands; while this could be solicited from a customer via technology (i.e. a smartphone survey) for some categories, it would be still difficult to scale this for all categories. The second is that the model was estimated in a controlled experimental setting for one product category (soft drinks), which runs counter to the objective of having a lightweight and generalizable description of customer variety seeking for different categories. The Trivedi model conceptually belongs to what McAlister & Pessemier (1982) notes as the class of models based on attribute satiation. The idea is that by purchasing and consuming a brand, the customer is “satiated” with that brand’s attributes, and therefore seeks out a much different brands. However, this class of model is not easily implemented in practice since customer perception of attributes is latent and unobservable to the retailer. Accordingly, the approach in the thesis for modeling and estimation aims to avoid the need for stated preferences or defined product category attributes.

¹ Association for Information Systems’ Senior Scholars’ Basket of Journals: http://ais.site-ym.com/?SeniorScholarBasket
A further class of models for variety seeking are the first order and second order models, which model variety seeking as a changing with time according to two competing processes of variety-seeking and inertia seeking, with a state dependency on the most recent purchase (Bawa 1990; Kahn et al. 1986)). These models differ from earlier zero-order models which assume an individual’s variety seeking as state-invariant (Bass et al. 1976, 1984). Although the high order models provide a finer granular description of individual variety seeking, in practice one has to assume a two brand market in order to assure enough degrees of freedom for estimating the model parameters (Bawa 1990; Kahn et al. 1986); as such their model is not appropriate for a multi-brand reality. Furthermore, Bawa (1990) showed that in practice, the descriptive power of the models did not differ much by assuming a higher order; the main advantage of a high order model was added research insight in the underlying mechanisms of variety seeking. As such, the approach in the thesis will be similar in spirit to the zero-order models.

2.2 Leveraging Social Learning Marketing for Boosting Sales

With the advent of smartphones, tablets and retailer branded apps, it becomes possible for retailers to leverage the online channel to drive in-store purchases. Recognizing the fact that customers are often influenced by the purchase choices of others when they make their own purchase choices (Chen, Wang, et al. 2011; Hanson and Putler 1996), retailers have leveraged E-commerce portals to facilitate popularity-based marketing tools such as a top ten sorted list of sold products. The underlying mechanism is denoted as Observational Learning OL and describes the observation of others’ actions without considering the underlying rationale (Bikhchandani et al. 1998). The customer subsequently infers some state of reality from the actions of others and takes the same action themselves.

This use of social learning as marketing is an example of leveraging the strengths particular to that channel; the ranked lists, which need to be rapidly computed and updated, are easy to implement online, but difficult to implement in physical stores, due to the limitations of timeliness of printed ads.

Accordingly, this section introduces the concept of observational learning in the context of social learning marketing, and explains the current theoretical streams and gaps for research in this field.

2.2.1 Observational Learning in the Context of Social Learning Marketing

The effectiveness of social learning for influencing customer behavior has been studied early in marketing research; for an example, Arndt (1967) found that customers who received positive
news from others about a new food product were much more likely to purchase it compared to those who got negative information from others, showing the direct effect of social learning. Brown and Reingen (1987) studied how the uptake of new service providers are facilitated between the links and communications between customers, showing how such information is diffused by relations. Richins (1983) examined some of the antecedents leading to the antecedents of negative world-of-mouth (WOM) by dissatisfied customers with an open end interview and a questionnaire, to find out why or under what conditions is social learning of negative information facilitated. Given these different examples of social learning, there are various perspectives and definitions as to what counts as social learning and how to classify them: In the psychology perspective (Cialdini and Goldstein 2004; Deutsch and Gerard 1955), social learning can be either persuasive by suggesting to the customer what actions are socially acceptable by others (normative social influence), or by suggesting that information from others is evidence of some state of reality (informational social influence). Examples of these concepts have been applied to the area of product evaluation (Burnkrant and Cousineau 1975; Pincus and Waters 1977; Wooten and Reed 1998); in one example, Burnkrant and Cousineau (1975) showed with ground coffee as a product, for example, that a customer’s awareness of positive ratings and recommendations from designated experts (normative influence) positively influenced that customer’s evaluation of the product, as with being able to see positive ratings of other anonymous customers (informational influence).

In the economics perspective, social learning is described as customer-to-customer (C2C) interactions, where the depth of information available and transmitted between customers determines whether it is classified as observational learning or word of mouth (WOM) (Chen, Liu, et al. 2011; Libai et al. 2010). Observational learning is defined as the “inference resulting from rational processing of information gained by observing others” (Bikhchandani et al. 1998, pg. 153); the only pre-condition is for one customer to see what other customers have done, but the customers do not convey the motivation or reason for their action. There is no verbal communication between the one who learns and the one who acted (Libai et al. 2010). In a retail context, inference or “learning” of about a product’s quality comes primarily from the observing customer. A retail example of this are best sellers list on E-commerce sites, where customers are able to see products ranked by their weekly sales, and thus observe - without knowing the reason why - roughly how many people in the past have decided to buy a certain product (Chen, Wang, et al. 2011).
In contrast, word of mouth is usually communicated between two parties: the “sender” is the person who transmits the message – in a retail context, the one who bought and experienced the product - and the “receiver” who receives the message and is the one seeking information about a product (King et al. 2014). Unlike observational learning, the sending reason or motivation for the choice is clearly indicated (Chen, Liu, et al. 2011; Chevalier and Mayzlin 2003). In retailing, an example of this are product reviews on E-commerce sites; a customer clearly signals what he thought of a product he bought and used, which a new customer can learn from with a higher granularity of information than by observing that someone bought the product (Chen and Xie 2008; Z. Zhang et al. 2012). The psychology and economics perspectives are not mutually exclusive and can overlap; recall from psychology that informational social influence occurs when information from others is taken as evidence of some state of reality; in a retail context, both word of mouth and observational learning are forms of this, as both involve learning about a product without prior direct experience.

Although both word of mouth and observational learning are now both commonly used particularly in E-commerce retailing (Chen, Liu, et al. 2011; King et al. 2014), there are several aspects that render observational learning attractive not only on the level of research, but also from a practitioner’s view. On a practitioner level, observational learning is less complex to employ than word of mouth, since all that is needed is a popularity metric (such as the volume of sales of a product), which is typically already gathered by the retailers’ Point of Sales (PoS) systems. Furthermore, effective implementations of observational learning came with recent advances in information systems technology, particularly in the area of online shopping. Online shopping platforms such as Amazon.com or eBay.com already enable customers to view information on product popularity in the form of number of sales and product reviews. Word of mouth however, requires active user generated content (for example, written product reviews), which may not be available for all products – indeed, a body of research in the area of recommender systems focuses on algorithms that can overcome the “cold-start” problem of sparse user generated content (Levi et al. 2012; Schein and Ungar 2002; Victor and Cornelis 2008). Since retailers already have PoS data, the deployment of an observational learning system faces less “cold-start” problems compared to word of mouth.

Additionally, on the level of research, word of mouth has already been fairly well studied both online (Chen, Liu, et al. 2011; Chevalier and Mayzlin 2003; Dellarocas 2003; Godes and Mayzlin 2004) and offline (Bowman and Narayandas 2001; Herr et al. 1991). For an example, researchers have proven the positive effect of word of mouth on the uptake of products and services across
different domains and in forms: via online book reviews on sales (Chevalier and Mayzlin 2003), in forums on TV show uptake (Godes and Mayzlin 2004) and in product reviews on media and movie uptake (Chen, Liu, et al. 2011), among many domains. The effect of negative word of mouth have also been studied, with Chen, Wang, et al. (2011) finding an asymmetrically stronger effect of negative word of mouth compared to positive word of mouth on sales. Research has also investigated the antecedents for giving word of mouth feedback (Cheung et al. 2012; Hennig-Thurau et al. 2004); other research deals with how to analyze the text content of a word of mouth message (Jansen and Zhang 2009; Z. J. Zhang et al. 2012). Research frameworks have also been established to classify and charter the future direction of word of mouth research (Dellarocas 2003; King et al. 2014; Libai et al. 2010). Taken together, all of these indicate word of mouth research to be a mature and well understood field. In contrast, observational learning has only recently been employed for exploratory research in persuasive systems, mostly in E-commerce portals (Chen, Wang, et al. 2011; Duan et al. 2009; Tucker and Zhang 2011) while research gaps remain in investigating alternative observational learning metrics and dimensions, as discussed in the subsequent section. Accordingly, within the bigger picture of social learning, the focus will be on observational learning rather than word of mouth. The focus on observational learning in this thesis and the subsequent research gap that is addressed will be relevant to both the psychology and economics community – namely to informational social learning of the psychology perspective and the observational learning in the economics perspective.

2.2.2 The Sales Velocity Research Gap

Observational learning was defined as the “inference resulting from rational processing of information gained by observing others” (Bikhchandani, Hirshleifer, and Welch 1998, pg. 153); although this definition is broad enough to encompass both learning from observing the actions of others, as well observing as the rate at which the actions occur (this rate change of the actions can be thought of as the “velocity”), the observational learning literature subsequent to Bikhchandani, Hirshleifer, and Welch (1998) dealt primarily in the context of the former. These include studies of observational learning in E-commerce portals which implemented observational learning as a static sales rank (Chen, Wang, et al. 2011), website clicks in a wedding product portal (Tucker and Zhang 2011), or even in the area of organ donation waiting lists (Zhang 2009). It has also been studied in offline settings, such as customer menu choices in a restaurant (Cai et al. 2009). Although the static effect of observational learning is thus well established, in all of these aforementioned cases, the velocity dimension is not considered.
However, research beyond observational learning suggests that observing the velocity component of others’ actions could have an effect on purchase decisions. Research on intertemporal choice, for example, has shown that customers have a preference for a sequence of improving outcomes (Loewenstein and Prelec 1993); that is they prefer things that get better and better, and that this effect gets stronger with the speed with which the improvement occurs over time (Hsee and Abelson 1991; Hsee et al. 1991, 1994). Extending this into the area of retailing, an example would be a product’s sales velocity – that is, how much a product has improved in sales or sales rank over a period of time.

But up until now, it has not been shown that the concept of improving outcomes could be leveraged as a marketing tool, or that there is a link between the sequence of improving outcomes and the velocity component of observational learning. To confirm this, this thesis conducted a forward and backwards search for papers which cited either core observational learning papers (Bikhchandani et al. 1998), the aforementioned studies (Hsee and Abelson 1991) or models of social learning reviewed by (Chamley 2004). Further searches included “rate change”, “sales velocity”, “velocity outcomes” and “velocity performance”. The search was applied to the databases ABI/INFORM, Business Source Complete (EBSCO), the Web of Science and Google Scholar, as well as the top ten marketing journals defined in analysis by Hult et al. (1997) and Steward and Lewis (2009). Articles under consideration were from 1991 to July 2014.

The articles found do not leverage velocity effects as a marketing tool: Improving outcomes was used as a diagnostic metric within firms’ and in the supply chain (Bronnenberg and Sismeiro 2002; Groves et al. 2011; Sandoh and Larke 2002), studied for its effect on goal striving (Chang et al. 2009; Elicker et al. 2009) and evaluated for its effect on a firm’s satisfaction with logistics service providers (Briggs et al. 2010). Although parallels to the retail domain exist (e.g. in the study by Briggs et al. (2010), the logistics provider can be seen as the product which the firms – the customer – have to choose between), there is a visible research gap in evaluating this phenomenon in the retail domain. This motivates an extension of these results in the context of E-commerce portals, where it can be leveraged to actuate behavior change of customers; and furthermore, with a controlled experiment, so that one can gain deeper insight on the effect.

In a retailing context, one might therefore expect that revealing velocity information of product sales would positively influence the likelihood of purchase. From an observational learning point of view, this would be akin to observing not only the actions that others have taken but also the rate at which they are taken. A product’s sales velocity might also embody the idea of a product
being discovered and being better received, leading to a positive product evaluation for the deciding customer. Additionally, as might be expected from the accessibility-diacnosticity model of information (Feldman and Lynch 1988), if a piece of information is accessible and useful for a choice at hand, it would be accepted. Accordingly, depending on the context of the product or situation, the change in a product’s popularity could give a variety of informative signals to the customer, and therefore influence their purchasing choices. For example, for fast moving goods and routine purchases such as groceries, it is known that customers occasionally seek variety in their purchases (Van Trijp et al. 1996), sometimes even eschewing their known, pleasurable favorites for the sake of variety (Ratner et al. 1999). It therefore follows that in this situation, rather than the most popular product, which a customer is likely to have tried already, the customer might rather choose a less popular product – provided that they get some positive signal. The alternative product’s rising popularity might provide this signal.

2.3 Customer Relationship Management in the Pre-Smartphone Age

With the popularization of loyalty cards, retailers have been able to collect large amounts of purchasing data linked to individuals, measuring customer purchases and marketing activity effectiveness (Verhoef and Venkatesan 2010). In this section, the focus is on the research concerning Customer Relationship (CRM) management systems, the primary information system which handles this customer data. In particular, focus is given to metrics and models which are used in CRM systems for evaluating customer value to the retailer.

2.3.1 Research in Customer Relationship Management Profitability

The primary information systems which handle customer data are Customer Relationship Management (CRM) systems. CRM systems are described as information technology enabled relationship marketing that support organizations in identifying, attracting and increasing retention of profitable customers by managing relationships with them, with the inference being built on analysis of customer transaction data (Payne and Frow 2005; Ryals and Payne 2001). Strongly linked to CRMs is the concept of using them to build and sustain customer loyalty. Loyalty has been described as arising from behavior and attitudes (Kumar and Shah 2004): behavioral loyalty is observed from past and present purchases at a given retailer, while attitudinal loyalty refers to the long term, higher order commitment of a customer to an organization – i.e. pointing to the future. To this end, CRM research has been studied and

---

2 It is acknowledged that a product can decrease in sales rank. In the thesis, the focus is on establishing the sales velocity effect in the positive direction – similar to a best sellers list, which lists only positive observational learning information. For a complete picture, the effect of negative sales velocity is also examined in this thesis.
applied in a variety of domains, ranging from financial services (Ryals and Payne 2001), to online retailers (Pan and Lee 2003), to grocery retail loyalty programs (Heerde and Bijmolt 2005; Liu 2007; Meyer-Waarden 2007; Zhang and Wedel 2009); the common thread in application is the idea of cyclically using historic and incoming transaction data to estimate customer value and buying preferences, and, depending on the retailer goals, selecting a subset of these customers for a marketing policy accordingly. However, in spite of their widespread usage, there are practical problems in profitably implementing a CRM system: without a link between loyalty and profitability, companies often waste valuable marketing resources attempting to build loyalty systems that are unprofitable (Reinartz and Kumar 2002). Thus having an information system in of itself in being able to collect and act on customer transaction data is not enough.

There are several approaches in current research for addressing this loyalty-profitability gap. One research stream deals with CRMs at the strategic level; they provide guidelines and CRM process models for management as to what CRM processes to take into consideration in order to be profitable (Goodhue et al. 2002; Kumar and Shah 2004; Kumar et al. 2006; Pan and Lee 2003; Payne and Frow 2005; Zablah et al. 2004). For example, Kumar and Shah (2004) provide a framework where attitudinal and behavioral loyalty are combined with an explicit evaluation of a customer’s future transactional profitability to the retailer; the optimal marketing action then depends on the levels of these three metrics. Attitudinal loyalty can be measured by surveys, focus groups and interviews, while behavioral loyalty and lifetime value can come from customer transaction data. Zablah et al. (2004) present a framework which argues that profit maximization requires an evaluation of the desirability of individual customers, a model of customer defection intentions, knowledge of the needs and preferences of customers and knowledge and emergence of market threats. Goodhue et al. (2002) provide a set of best practices based on case studies of six companies who have successfully implemented CRMs. At a strategic level, owing to the increasingly multi-channel (online, offline and in augmented environments such as with smartphones in stores) aspect of retailers (Verhoef and Venkatesan 2010), current CRM research focuses on frameworks for adapting CRMs to multi-channel retailing (Kumar 2010; Neslin and Shankar 2009; Verhoef and Venkatesan 2010). The problems pertaining to the profitability of single-channel CRMs still apply to the multi-channel case (Verhoef and Venkatesan 2010).

Another stream of research goes one level lower from this strategic view: it addresses the profitability problem by improving on the information systems which deploy loyalty schemes. For example, Jonker et al. (2006) propose an information system that sends a reward to a
customer at an optimal frequency in order to improve the recency of the last purchase, the frequency of purchases and the monetary value of the purchases, all of which are used to predict future profitability. Lee and Park (2005) present an information system which segments customers based on the predicted profitability of individual customers, with a marketing scheme that aims to maximize the overall profitability given a particular marketing budget.

A common enabler and thread between all of the aforementioned literature is the dependency on an underlying metric for evaluating the profitability of a particular customer to the firm; without such a metric it is impossible to lead to any improvement in profitability. As identified by Kumar et al. (2006, pg. 281), one important question in both research and practice regarding CRM profitability is “what is the right metric to manage customer loyalty”? To this question, past research have proposed different performance models and metrics for evaluating customer loyalty to the retailer. For example, as reviewed in Kumar and Shah (2004, pg. 318) loyalty measures in past research have included proportion of purchase, probability of purchase, probability of product repurchase purchase frequency, repeat purchase behavior and purchase sequence. Common metrics used by practitioners include share of purchase (SOP) or share of wallet (SOW), which measures the relative share of a customer’s purchase at a retailer as compared to the total number of purchases; share of visits (SOV), which measure the number of visits to the store as compared to the total number of visits across all stores; Past Customer Value (PCV) – based on the past profit contribution of the customer; and Recency, Frequency and Monetary Value (RFM) – a measure of how recent, how frequent and the how much was spent by a customer (Kumar and Shah 2004). A common thread between them is that they do not reflect future profitability (Kumar and Shah 2004); mere behavioral loyalty, of which the majority of the aforementioned metrics measure, is weakly correlated with future profitability (Dowling and Uncles 1997; Reinartz and Kumar 2002), since many schemes today reward customers for past actions or actions committed today rather than considering future potential of the customer (Reinartz and Kumar 2003; Yi and Jeon 2003).

Accordingly, current research advocates for and applies the Customer Lifetime Value (CLV) or the lifetime value (LTV) metric for identifying valuable customers for use in a CRM (Gupta and Zeithaml 2006; Hwang et al. 2004; Kim et al. 2006; Kumar and Shah 2004; Kumar et al. 2006; Reinartz and Kumar 2000, 2003). It is defined as the present value of the future cash flows associated with the customer (Pfeiferis et al. 2004). The review by Gupta and Zeithaml (2006) showed that marketing decisions based on CLV improve a firm’s financial performance and play a major role in customer retention, while the review by Blattberg et al. (2009) showed that CLV
can be used effectively to diagnose marketing efforts. With a forward looking model for computing CLV, it becomes possible to model their future profitability and thus segment them for marketing action to be delivered by an information system. Indeed, CRM systems and algorithms have already been proposed which segment customers based on CLV (Chen et al. 2008; Hwang et al. 2004; Kim et al. 2006; Ngai et al. 2009; Verhoef et al. 2002). As such, the focus of this thesis will be on addressing CLV research gaps and on its application in evaluating the effectiveness of marketing actions.

2.3.2 Models of Customer Lifetime Value (CLV)

Different models exist for computing CLV (Fader and Hardie 2009; Gupta et al. 2006; Jain and Singh 2002; Kumar et al. 2006), however a common thread is that in order to determine a customer’s overall lifetime value, they factor in past buying behavior of a customer and the cost of future marketing action for retaining him. At a high level, this means:

\[
\text{CLV} = f(\text{cost of acquisition}, \text{expected revenue per transaction}, \text{future number of transactions})
\]  

(Equation 1)

Equation 1 answers what is the present value of future cash flows associated with the customer. The current state of the art in CLV models are the so-called customer base analysis models (Fader and Hardie 2009; Fader, B. G. Hardie, et al. 2005; Jain and Singh 2002), which take into account the past purchase of the entire customer base (i.e. the population) in order to come up with individual probabilities of purchase in the next time period. These models factor in the stochastic behavior of individual customers in making purchases; as such, these estimates of individual buying can be used as inputs to compute the future expected transactions, and in turn, the customer lifetime value. This is in contrast to first generation models (reviewed in Jain & Singh (2002)) where some finite time horizon is assumed, and the stochastic nature of the purchase process and cash flow timing are ignored.

The appropriate mathematical models for each of the components of Equation 1 depend on the type of relationship with the customer (contractual vs. non-contractual) and the type of opportunities for transactions (continuous vs. discrete). In the contractual setting, the firm is informed when a customer wants to end a relationship with them, and in the non-contractual setting the time at which a customer becomes inactive is unknown to the firm (Reinartz and Kumar 2000). Regardless of whether the setting is contractual or not, the opportunity for a transaction can either occur at any time the customer wants (continuous) or at a well-defined
point in time (discrete) (Fader and Hardie 2009). A mapping of common business-to-customer relationships to this contractual/non-contractual vs. continuous-discrete paradigm, modified from Fader and Hardie (2009) is given in Table 1.

Table 1: Classifying customer bases, as modified from Fader and Hardie (2009)

<table>
<thead>
<tr>
<th>Opportunities for Transactions</th>
<th>Type of Relationship with Customers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Contractual</td>
</tr>
<tr>
<td>Continuous</td>
<td>Grocery Purchases</td>
</tr>
<tr>
<td></td>
<td>Doctor Visits</td>
</tr>
<tr>
<td></td>
<td>Hotel Stays</td>
</tr>
<tr>
<td>Discrete</td>
<td>Event Attendance</td>
</tr>
<tr>
<td></td>
<td>Prescription refills</td>
</tr>
<tr>
<td></td>
<td>Charity fund drives</td>
</tr>
<tr>
<td></td>
<td>Contractual</td>
</tr>
<tr>
<td></td>
<td>Credit Card</td>
</tr>
<tr>
<td></td>
<td>Student Mealplan</td>
</tr>
<tr>
<td></td>
<td>Mobile Phone Usage</td>
</tr>
<tr>
<td></td>
<td>Magazine subs</td>
</tr>
<tr>
<td></td>
<td>Insurance policy</td>
</tr>
<tr>
<td></td>
<td>Gym membership</td>
</tr>
</tbody>
</table>

In grocery retailing, since consumers are not bound by a contract to a particular retailer, and they can make a purchase at any time, it is therefore a non-contractual continuous setting. A key challenge in this type of setting is estimating the future number of transactions, without knowing whether a customer has truly left the retailer or not; the standard probability model used to predict this customer “death” is the NBD/Pareto model (Morrison and Schmittlein 1988). This model assumes that while the customer is alive, he is purchasing randomly around his mean transaction rate, which is assumed to vary between customers. The customer’s unobserved “death” is treated as if random, characterized by the exponential distribution, and heterogeneity in death rates is characterized by the gamma distribution. This model has shown empirical validity across various non-contractual continuous domains such as catalog retailers (Reinartz and Kumar 2000, 2003), office retailing (Schmittlein and Peterson 1994), online music retailers (Fader, B. G. Hardie, et al. 2005). Its extensions are documented in Fader & Hardie (2009), the most critical of which for CLV are the extensions from Fader et al. (2005) (which is referred to as the Fader model) which derives a probabilistic expression for the average spending per transaction and an expression for the future number of transactions based on the NBD/Pareto model, thus allowing the estimation of Equation 1. The Fader model evaluates the lifetime value of individual customers at the firm level, given their recency, frequency and average monetary value of purchases (RFM) at the firm - data which are widely available from a firm’s CRM system. The simplicity of the input data and the predictive validity across domains makes it an attractive model.
2.3.3 Spatial Resolution for CLV Models and Their Timely Application

Although the Fader model is promising for use in an in-store high resolution marketing CRM, it is currently applied at the level of the customer’s total purchases in the firm (Fader, B. G. Hardie, et al. 2005; Glady et al. 2009)- without evaluation at the product category level. It is well documented that consumers exhibit heterogeneity in buying preferences for different product categories (Allenby and Rossi 1998; Leszczyc and Bass 1998; Lim et al. 2005), and so it follows from this category-dependent interest that a customer would have different levels of value for these categories, scattered between patronized retailers. One could imagine that a customer who has a high CLV overall could however have certain product categories where he is not valuable, which would not be known without a category level CLV. There may even be categories where he runs the risk of partial defection to another retailer (Buckinx and Van den Poel 2005), but this may not emerge from a person-level CLV until it is too late. Quantifying this consistency or spread of CLV between categories within a given person forms a research gap.

In addition to this granularity gap in CLV for grocery retailing, there is also a temporal gap: the application of CLV in a CRM is typically as a one-time or infrequently applied segmentation tool (Hwang et al. 2004; Kim et al. 2006). Recent work by Wang & Hong (2006) challenges this, by proposing a framework where a customer’s lifetime value is continuously computed and the appropriate marketing action is chosen, depending on the lifetime value’s amount, trend and volatility, plus the accessibility to the customer (the extent that which a customer has historically accepted a marketing offer). The “amount” is defined as the CLV value of the customer computed at a given time, the “trend” is defined as “movement behind past and current customer profitability” (Wang and Hong 2006; pg. 719) and the “volatility” refers the temporal fluctuation of the CLV in a time period. Notably two of these dimensions – the trend and volatility – are time dependent. It is thus one of the first works to consider the dynamics of CLV in a CRM. Combinations of the CLV amount, trend and volatility as well as the customer accessibility lead to different customer types and hence a different corresponding marketing tactic; a summary, modified from Wang & Hong (2006), is found in Figure 3:
<table>
<thead>
<tr>
<th>CLV Dimensions</th>
<th>Levels</th>
<th>Combination s of CLV Dimensions</th>
<th>Customer Classification</th>
<th>Corresponding Tactic</th>
<th>Required Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount (P)</td>
<td>Loss – Low – Medium - High</td>
<td></td>
<td>Defecting Customers</td>
<td>Win back tactics: Win back customer share</td>
<td>Churn analysis</td>
</tr>
<tr>
<td>Trend (T)</td>
<td>Down – Flat – Up</td>
<td></td>
<td>Growing Customers</td>
<td>Upgrade tactics: increase customer share</td>
<td>Upgrade analysis</td>
</tr>
<tr>
<td>Volatility (V)</td>
<td>Significant - Safe</td>
<td></td>
<td>Steady Customers</td>
<td>Loyalty tactics: broaden market spaces</td>
<td>Loyalty analysis</td>
</tr>
<tr>
<td>Accessibility (A)</td>
<td>High – Moderate - Low</td>
<td></td>
<td>Inactive Customers</td>
<td>Reactivation tactics: Increase customer base</td>
<td>Custome r base analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Unprofitable Customers</td>
<td>Cost down tactics: decrease contact cost</td>
<td>Channel analysis</td>
</tr>
</tbody>
</table>

Figure 3: Customer classification and corresponding marketing tactic, depending on CLV dimensions (temporal dimension included), as modified from Wang & Hong (2006)

In Figure 3, the two temporal dimensions of CLV – the volatility (abbreviated as V) and the trend (abbreviated as T) feature prominently in detecting whether a customer is defecting, growing or staying steady in profitability. Although the work by Wang & Hong (2006) acknowledges the temporal dimension of CLV, these are operationalized at a person-level – a gap exists in evaluating the business value that arises from having a product category-level CLV recomputed repeatedly as a diagnostic metric. Furthermore, the consistency of CLV between categories within a given person can augment the customer development framework by Wang & Hong (2006); in addition to the lifetime value’s amount, trend and volatility, CLV consistency becomes another managerial decision making metric. That is, while the framework by Wang & Hong (2006) can inform managers “what” overall marketing tactic to employ depending on the temporal trend of a customer’s overall CLV, the CLV at the product category level could identify the exact category where the marketing tactic should be applied, thus helping managers decide “where” to apply the marketing tactic.
The contribution in this thesis will thus address these gaps by expanding on the Fader model to a product-category level, with an emphasis on how this extension could capture more economic value from PoS data. In the spirit of the work by Wang & Hong (2006), the thesis will also apply the product-category level model at multiple time points and also characterize the consistency of CLV between categories within a given person. The aim is to enable high resolution evaluation of customer profitability, and by extension, high resolution marketing for physical stores via ubiquitous technology such as smartphones and tablets.
3. Towards High Resolution Identification of Variety-Seeking Behavior

This section concerns with how a multi-channel retailer could segment its customers – a core identified research gap in multi-channel retailing (Dholakia et al. 2010; Konus et al. 2008; Neslin and Shankar 2009; Payne and Frow 2005). It presents a data-driven model for measuring variety seeking behavior at a high level of granularity, based on a consumer’s purchases in individual product categories. The model thus moves towards the goal of one-to-one marketing, having a segment of one, thereby addressing RQ1 – providing an analytical method for retailers that accounts for the habitual aspects of customers and can be readily operationalized in a customer relationship management system.

3.1 Introduction

Segmentation schemes help retailers identify customers with heterogeneous preferences, and thus form the basis of finer grained marketing (Dickson and Ginter 1987). A particularly interesting and high value target group is the one that is variety seeking, since this group is most likely to respond positively to new offers. To identify customers’ psychological purchase preferences, the costly method of deploying questionnaires is often used (Dodds et al. 1991; Fisher et al. 2002; Garretson et al. 1998; Kerin et al. 1992; Van Trijp et al. 1996; Voss et al. 2003; Wakefield and Inman 2003). However, this approach is not scalable; retailers cannot sustainably survey all of their consumers, and thus the standard marketing action is not personalization, but rather having all consumers receive the same promotion, typically a price cut. Furthermore, current questionnaire approaches neglect that the actual disposition of variety seeking of a person might in fact differ from product category to product category. Indeed, it is well know that there is heterogeneity in customer buying preferences for different product categories (Allenby and Rossi 1998; Leszczyc and Bass 1998; Lim et al. 2005).

This section therefore proposes a data-driven model for measuring individual variety seeking behavior at the level of product level categories – i.e. segmentation of variety seeking with a segment size of one. This is in line with the widely held view that such one-to-one marketing is more effective than mass marketing (Changchien et al. 2004; Jiang and Tuzhilin 2006; Sheth et al. 2000; Thomas and Sullivan 2005). The model is estimated on Point of Sale (PoS) loyalty card data widely available for the retailer. The model is in contrast to previous variety seeking models made before the smartphone age, which segment customer behavior at a store level (Bawa 1990; Kim et al. 2006; Konus et al. 2008; Trivedi 1999; Verhoef et al. 2002). A data-driven
segmentation approach is preferable to a questionnaire approach, since it can then be operationalized into CRMs which retailers already possess. It can then be automated on widely available CRM data, and marketing actions such as recommendations can then be automatically delivered through smartphones.

The chapter has three main contributions. First, it contributes to the customer segmentation research stream by providing a novel way for identifying customers’ overall extent of variety seeking as well as their specific variety seeking at a category level. Unlike the Trivedi model (1999), a customer’s stated preference of the different brands are not required. Second, for the most important retail categories, this section will characterize the extent of variety seeking and provide a data-driven approach that is easy to operationalize by practitioners – especially for deploying large-scale personalized marketing measures in social or mobile commerce in physical stores. Finally, this chapter will provide a method to reconcile the highly granular category-level results with existing per person typologies found in questionnaires.

3.2 Research Framework

The first research question this chapter asks is how a high resolution segmentation model can be developed for multichannel marketing for physical grocery stores. Extending from the previous theoretical discussion, this section aims to develop a method for estimating an individual’s extent of variety seeking in a given product category as the basis for a high resolution scheme. In order to develop the variety seeking model, it is thus necessary to answer the question: “How many unique products must a customer buy in a product category in a given period of time to be considered a variety seeker?” To address this, this thesis employs two assumptions:

1. Consumers’ extent of variety seeking in a given product category can be described with a zero-order model, that is, independent of state (i.e. previous purchases).
2. It is also assumed that a consumer’s pattern of variety seeking in a given product category can be described by counting the number of distinct products that they have bought in a period of time, conditioned against the extent of variety exhibited by other consumers.

In the first assumption, it was noted by Bawa (1990) that in practice, the descriptive power of the models did not differ much by assuming a higher order (i.e. state dependence); the main advantage of a high order model was added research insight in the underlying mechanisms of variety seeking. Hence, this thesis assumes a zero order model. The second assumption proposes conditioning the variety seeking on the product category, the time period and the behavior of
others in the population because in the domain of grocery retailing, it is known that the inter-purchase time (and hence, frequency and variety) of product purchases varies by product category (Leszczyc et al. 2004; Rhee and Bell 2002). For example, a product like milk may have only one to two substitutes and on average a consumer may stick to one variant; as such, a person buying three varieties of milk would be considered variety seeking, whereas for fruit yoghurt, buying three varieties per year might not be considered to be variety seeking, if on average people buy eight distinct flavors per year out of twenty available. Thus, by evaluating the variety of an individual’s purchases against the variety shown by the overall population in that category, it is possible to control for category-level differences.

3.2.1 Probabilistic Model Fit for Variety of Goods Purchased

Consequently, the second assumption leads naturally into stochastic counting-based models (Andersen et al. 1993), employed previously in consumer base analysis to successfully predict customer lifetime (Fader and Hardie 2009; Fader, B. G. S. Hardie, et al. 2005). In the models described by Fader et al. (2005), an individual consumer’s frequency of purchases at a retailer are counted and used as an input to estimate how long they are likely to continue buying from that retailer. A similar idea is used by counting the number of distinct products bought in a time period to describe variety seeking. Previously, such models have not been applied to the area of variety seeking.

Since count data is discrete, non-negative and has no upper limit, it follows that the probabilistic distribution for modeling the population’s variety seeking should be a discrete and positive skewed distribution. For parsimony, the model aims for a univariate probability mass functions (PMFs), since this lends itself easily towards an interpretable, one dimensional variety index. PMFs of this type include Poisson, geometric and negative binomial distribution (NBD).

Conceptually, the Poisson distribution is the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event; modeled on variety seeking, this implies individuals have differing rates of variety seeking and satiation, subject to some stochastic process, and some probability of consuming a certain amount of variety – in-line with the evidence on satiation (Chintagunta 1999; McAlister and Pessemier 1982). This conceptualization is also in-line with Wood & Neal (2009), which noted that individual consumers tend to be habitual in their purchases (i.e. implying the population’s variety seeking is positively skewed). The Poisson distribution is given by:
Where \( k = 0,1,2\ldots \) and \( \lambda \) is the Poisson shape parameter, with \( \lambda > 0 \). The distribution gives the probability a consumer has bought \( k \) products, given the behavior of the overall population in that product category. Note also that mathematically, the Poisson shape parameter \( \lambda \) also describes the average number of distinct products bought by customers in that category. Thus, with \( \lambda \), it becomes possible to also compare the expected variety of goods consumed between categories.

Other distributions are possible, with similarly plausible storylines: The geometric distribution is the probability distribution of the number \( X \) of Bernoulli trials needed to get one “success” event, supported on the set of natural numbers. A Bernoulli trial is a random experiment in which there are exactly two outcomes, “success” or “failure”, with the probability of success defined the same way each trial (by a fixed parameter). Applied to variety seeking, the Bernoulli trial would be a check at a given time of whether someone is satiated (the “success” event) and will seek no more variety, and thus the geometric distribution would tell us, in the whole population, how many “failure” events this would take before satiation. The geometric distribution is given by:

\[
Pr(X = k) = (1 - p)^k p
\]  
(Equation 3)

Where \( k = 0,1,2\ldots \) and \( p \) is the probability of “failure” (i.e. non-satiation/ desire for continued variety).

The NBD proposes a similar story. It models the number of successes in a sequence of Bernoulli trials before a specified (non-random) number of failures (denoted \( r \)) occur. For NBD, if one considers “success” to be non-satiation, and “failure” to be satiation, then by letting \( r = 1 \), then NBD provides the \( k \) number of events before satiation for the population. The negative binomial distribution is given by:

\[
Pr(x = k) = \binom{k+r-1}{k}(1-p)^r p^k
\]  
(Equation 4)

This study will test the fit of these models for each product category. Accordingly, in answering RQ1 – namely, how a high resolution variety seeking model can be developed - the following research sub-question is proposed:
RQ1.1: Which discrete probability distribution best captures the population’s variety seeking?

3.2.2 Customer Variety Seeking Heterogeneity Across Product Categories

It was reviewed in Section 2 that variety-seeking is a personal characteristic – as such one would expect the degree of variety seeking for a given person to be the same across product categories (i.e. the null hypothesis would hold true for the variety indices of a person). However, this thesis proposes that consumer variety seeking will differ between product categories. It is known that the inter-purchase time of product purchases varies by product category (Leszczyc et al. 2004; Rhee and Bell 2002), and hence so would variety seeking. Furthermore, the attribute-based theory of variety seeking (reviewed in McAlister & Pessemier (1982)) suggest that since different product categories have different attributes which are “consumed”, it follows that some categories are more prone to satiation (and hence variety seeking) than others.

The study would confirm comprehensively whether the variety seeking varies from category to category. Accordingly the next research subquestion is:

RQ1.2: How does a customer’s variety seeking differ from category to category?

3.2.3 Customer Typologies

As discussed in Section 2, questionnaires are used by marketers to develop customer typologies for better targeting consumers with offers and products. In the case of questionnaires, the variety seeking typology applies at a general level, irrespective of product category. It was argued that the variety seeking indices of a given person at the category level will be heterogeneous across categories. However, in spite of this heterogeneity, it is possible that there is a high-level similarity between consumers, leading to typologies or groups of similar consumers which marketers can address at a coarser granularity. Thus:

RQ1.3: What are higher level similarities between customers in spite of their heterogeneous variety seeking between categories?

In answering this question, by finding a higher level typology of variety seekers, one would achieve the same result as the traditional questionnaire. Hence this thesis would be able to offer information systems such as mobile RS and marketers different granularities of analyzing their customers.
3.2.4 Time Effects on Consumer Typologies and Overall Variety Seeking

The psychological questionnaire methods of identifying overall variety seeking within a customer posits that variety seeking is an internal customer trait and therefore overall consistent over time (Laroche et al. 2003; Van Trijp et al. 1996), outside of circumstantial external stimuli known to trigger variety seeking such as coupon deals, stock-outs or manipulation of the perceived assortment of goods (Burke et al. 1992; Diehl and Pynor 2010; Kahn and Wansink 2004; Read et al. 1995; Simonson 1990). Thus, by applying the variety seeking model presented in this thesis at different time points, one could obtain a snapshot of a person’s variety seeking with time, and thus numerically confirm whether variety seeking is in fact a time-invariant characteristic trait as suggested by the psychology literature. This leads to the following question:

*RQ1.4: How does the overall variety seeking of a customer change with time?*

At the same time, although literature suggests that the overall variety seeking of a customer stays stable over time, whether category level variety seeking remains stable and consistent over time is unknown, since previous research, with the exception of Van Trijp et al. (1996), did not consider category-level differences in variety seeking. This motivates the next research question:

*RQ1.5: How does category level variety seeking change with time?*

3.3 Method and Results

3.3.1 Dataset for Estimation

To address the research questions, a year of point of sale data from a physical grocery retailer partner was examined. The receipt data comes from a complete set of all transactions from a single store, and consists of over 150,000 unique receipts covering a total of two million transactions, with a total of 19,374 unique products sold that year. A transaction refers to the purchase of an item as it appears on the receipt; each transaction event records the name of the item, a timestamp of when it was bought, the European article number (EAN), a receipt ID, the number of units bought, the price per unit, the loyalty card ID of the household who bought it and the product’s category as defined by the retailer. Although the customer’s age distribution and gender are not known, it is the full dataset of one particular store and was deemed “typical” by the retail partner. Therefore, from this data it is possible to construct the purchasing histories of the households, and identify how many items, unique or in total, they bought per category and at what time.
It is acknowledged that loyalty card holders might be different from the larger population of all customers and that a given loyalty card might represent a household with different people with different tastes; furthermore it known that consumers often split their purchases between multiple retailers (Leszczyc et al. 2000; Rhee and Bell 2002); these external purchases are unobservable to the retailer. However, since retailers only have measurable data about their own loyalty card holders, this study is relevant and in-line with a retailer’s practical limitations. Given the unobservable purchases, in order to determine the “natural” (from the point of view of the consumer) amount of variety and frequency of purchases for different product categories, the model is estimated on consumers who have shopped more than 51 times in the studied store (N=848). This is based on the observation that across all stores, consumers shop weekly (Leszczyc et al. 2000; Rhee and Bell 2002). Furthermore, the focus of the analysis is on products which were available all year round (defined as having been sold at least once per week in the entire store). Since many products are only available few days a year, if one treated these products as "available all year around", this would inflate the apparent number of products available for a consumer (and hence variety) to choose from. The analysis thus covers 638,000 transaction events. The minimum purchase and product availability filtering are in line with past models estimated on consumer scanner panel data (Bawa 1990; Bucklin et al. 1998; Guadagni and Little 1983; Gupta 1991).

3.3.2 Product Categories Definitions

The product categories of the retailer are defined at four hierarchical levels. At the highest (first) level, there are ten categories: “meat and sausages”, “fruits and vegetables”, “fresh goods”, “ingredients”, “preserved food”, “drinks”, “baked goods and sweets”, “washing material”, “perfume”, and “non-food items”. These are then broken down into subcategories at the second level – for example, “drinks” would be broken down into the subcategories of beer, coffee, tea, etc., which in turn can be broken down into a third level category (“bottled beer”, “canned beer”) to finally the fourth and lowest level (different variants of beer such as “draught beer”, “wheat beer”, etc.). The number of categories at each level is as follows (in brackets are the number of categories which are sold all year around): there are 66 (50) second level categories, 277 (146) third level categories and 900 (308) fourth level categories.

Although each retailer could have their own arbitrary classification scheme, these categories are mappable to reference categories defined by the widely adopted and open Global Product
Classification (GPC) international standard\textsuperscript{3}. This is because any product with a standardized barcode would automatically have a standardized Global Trade Item Number (GTIN), which has a predefined mapping to a specific GPC. As such, the results presented in this thesis are applicable and can be “translated” across all grocery retailers, by simply taking their GTINs and mapping to the associated GPC.

Admittedly, even the GPC standard categories can seem arbitrary to consumers and might not necessarily match the mental model of the consumer in terms of defining substitutes; hence, the amount of apparent variety seeking can be inflated due to the way the categories themselves are defined, (i.e. a category is too broad). Note however that the model presented accounts for this because the very definition of “variety” is evaluated with respect to the rest of the population-wide behavior in that category. Accordingly, the model fit is tested at the different four levels of product categories to assure its generalizability.

3.3.3 Probabilistic Model Fit for Variety of Goods Purchased

In order to answer RQ1.1, the fit of the three proposed models (Poisson, geometric and NBD) was tested on the POS data. Given a product category that is sold all year round, the total number of unique products bought by each consumer in the one-year data set was counted –this vector of frequencies is denoted as $X$. A Poisson, geometric and negative binomial distribution was then each fitted to $X$ using maximum likelihood estimation (MLE).

For each distribution fit the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) was computed in order to compare the goodness of fit of the models. It was found that both AIC and BIC gave nearly identical conclusions, so for parsimony the following results show conclusions derived from the AIC. The AIC is also well established for comparing goodness of fit models in various disciplines (Andrews and Currim 2003; Bozdogan 1987; Kamakura and Russell 1989). It measures of how well the model fits (relative to other models) – but there is a logarithmic penalty on the complexity of the model. In other words, AIC balances parsimony with goodness of fit, which is appropriate to the goal of having a parsimonious model.

The models were fitted using first the lowest level product categories, and then repeated the process for the higher level. For each level, there were also categories for which all consumers only bought one item; these categories were not considered in the model fit, since no distribution could be meaningfully fitted. The results are summarized in Table 2, which show that

\textsuperscript{3} Defined in http://www.gs1.org/gdsn/gpc/what
best fitting distribution, irrespective of product category level used, was the Poisson distribution. Therefore, the Poisson distribution was adopted for all subsequent analysis. Notably, the geometric function did not emerge once as a best fitting function. RQ1.1 is addressed.

Table 2: Summary statistics of best fitting distributions using different category levels.

<table>
<thead>
<tr>
<th>Category level used</th>
<th>% Categories fitted best to:</th>
<th># of Categories where no fit was possible (all consumers bought 1 variety of item)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poisson</td>
<td>NBD</td>
</tr>
<tr>
<td>4th (223 categories fitted)</td>
<td>97%</td>
<td>3%</td>
</tr>
<tr>
<td>3rd (128 categories fitted)</td>
<td>86%</td>
<td>14%</td>
</tr>
<tr>
<td>2nd (47 categories fitted)</td>
<td>68%</td>
<td>32%</td>
</tr>
</tbody>
</table>

The next step was to determine which category level to choose for subsequent analysis. Notably from Table 2, the lowest level (4th level) category had a high (28%) number of categories for which consumers bought only one variant of a product; a subsequent check was conducted to see if each of these categories had only one product available. They did. This suggests the 4th level category is too fine and overfits a category 1:1 to a single product. At the same time, it was shown that the 2nd level is too coarse; for example, the “milk products” category encompasses yoghurt, pudding and butter, which are arguably not substitutes for each other. Accordingly, the 3rd level category was chosen (128 categories). Figure 4 shows for selected product categories a frequency distribution of the variety of products bought, and the fitted Poisson distribution for that category:
Figure 4: Frequency distribution of customers who bought n-number of products

In line with the empirical findings of the AIC computations, Figure 4 qualitatively corroborates the good fit between the measured data and the Poisson model.

In order to give a more descriptive snapshot and reference point of what the data looks like at a category level, Table 3 presents a sample of the 128 categories, showing the top 5 product categories in terms of sales volume and the top 5 categories in terms of absolute variety. Table 3 also shows the Poisson parameter, $\lambda_{\text{Poisson}}$, which describes the population mean variety seeking. Notably, these categories with high variety seeking also have a large assortment of available products.
### Table 3: Sample of variety seeking categories. Note that $\lambda_{\text{Poisson}} = \mu_{\text{variety}}$

<table>
<thead>
<tr>
<th>Product Category</th>
<th># Unique Products</th>
<th>$\mu_{\text{variety}}$ (σ)</th>
<th>$\mu_{\text{Units Bought}}$ (σ)</th>
<th>Top 5 by variety</th>
<th>Top 5 by sales volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft Drinks</td>
<td>48</td>
<td>3.8 (2.9)</td>
<td>12.0 (16.6)</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Yoghurt</td>
<td>78</td>
<td>7.2 (5.2)</td>
<td>28.8 (35.1)</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Cheese</td>
<td>54</td>
<td>6.1 (4.3)</td>
<td>16.5 (17.6)</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Baked Goods</td>
<td>35</td>
<td>4.7 (3.1)</td>
<td>17.7 (21.7)</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Milk</td>
<td>11</td>
<td>2.6 (1.4)</td>
<td>33.0 (30.9)</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Cream</td>
<td>14</td>
<td>3.6 (2.0)</td>
<td>22.3 (22.3)</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

### 3.3.4 Customer Variety Seeking Heterogeneity Across Product Categories

#### Computing the Customer-Product Variety Index

In order to answer RQ1.2 and check whether consumer variety seeking is heterogeneous across product categories, it is first necessary to compute the variety index for each person for each product category. With $\lambda_{\text{Poisson}}$ computed for each product category, the Poisson distribution (Eq. 2, Section 3.2.1) would give the probability a consumer has bought $k$ unique products in that category. The higher the $k$, the more improbable it is. However, this probability would not be a direct surrogate of variety seeking: ideally, an individual variety seeking index should be monotonically increasing and would describe the person’s relative propensity to seek variety compared to the population. The Poisson cumulative distribution function (CDF) fits this and describes the probability someone bought less than or equal to $k$ products. It is given by:

$$
Pr(k \mid \lambda) = e^{-\lambda} \sum_{i=0}^{\text{floor} \lfloor k \lambda \rfloor} \frac{k \lambda^i}{i!}
$$

(Equation 5)

Thus someone who has bought a large variety of products relative to the population will also have a high probability, given by Equation 5, that the rest of the population has bought a lesser number of products. As such, the CDF captures the relative degree of variety seeking at the individual level. Using Equation 5, the variety index for each consumer was computed, for each product category where the customer shops.

By using this variety index instead of a raw value of variety of goods purchased, it is therefore possible to compare a consumer’s variety seeking across product categories, and compare consumers within the same product category. This is illustrated in Figure 5, for the example category of Yoghurt.
In the category of Yoghurt, the Poisson parameter is 7.2, meaning that on average in a year a customer buys 7.2 varieties of Yoghurt. The resultant CDF allows the computation of a variety index; referring to Figure 5, a person who buys on average 7 varieties of Yoghurt would have a variety index of 0.6; i.e. there is a 60% probability given the population examined that one bought less than 7 varieties of Yoghurt per year. A person who buys 15 varieties or more would have a variety index of 1.0. Since the relationship between the number of unique products bought and the variety index is non-linear, the CDF allows an exact mathematical definition of variety seeking of an individual relative to the overall population.

**Confirming and Characterizing the Heterogeneity of Variety Seeking Across Product Categories**

Having computed the variety indices of a consumer, it follows that if variety seeking was a trait that was category independent, then a given person should have the same extent of variety relative to the rest of the population for each product category (i.e. their variety index should not be significantly different across categories). Furthermore, this would also imply that the population distribution of variety indices for each category would be similar between categories.
The classical test for this statement would be a repeated measures ANOVA, since a consumer is subjected to multiple “treatments” (i.e. shopping in product categories) as stimuli, and the resultant outcomes (i.e. the variety indices) are compared between the treatment groups (the product categories). However, as shown in Section 3.3.3, since variety indices within groups (the product categories) are not normally distributed, the assumptions of ANOVA are violated. As discussed in Field et al. (2012), a non-parametric test which can be used in this instance is the Friedman test (Friedman 1937); however a requirement of the Friedman test is completeness of data. In the analyzed data set, since not every consumer purchased from every product category, there will be consumer-product category pairs for which no variety index exists. The analysis was therefore completed with the Skillings-Mack test (Skillings and Mack 1981), which generalizes the Friedman test to allow for missing data. The Skillings-Mack statistic was significant, \( T = 10205, p < 0.05 \); the product category did significantly change a person’s degree of variety seeking. RQ1.2 is addressed. This therefore justifies the need of computing a variety index at the granularity of person’s interaction with a specific category.

In addition to this empirical confirmation of the existence of heterogeneity, a frequency plot showing which bins the variety indexes between product categories within a given customer also affirms the non-uniformity and gives some insight about the characteristic of the heterogeneity. A random sampling of four anonymized customers are shown in Figure 6:
If there was no variability in variety seeking across product categories within a given customer, then the variety index should be the same across all categories. However, Figure 6 clearly shows that for a given customer, there are some categories which the variety seeking is high (>0.75) where there are others where little variety is sought (<0.25).

To characterize the heterogeneity of variety seeking within a customer, a metric is needed which describes the spread of heterogeneity. In order for comparisons to be made between customers who may have purchased in a differing number of categories, the metric should therefore be dimensionless and captures unequal variety within a person. For this, the Gini coefficient was computed for the category level variety indices. The Gini is a common measure of distributional inequality; it has been applied to many problems, the most well-known being income inequality (Sen 1976). Drawing a parallel to the variety seeking case, the coefficient quantifies to what extent a person is consistent in their category level variety indices. A low Gini coefficient indicates that a person is either consistently having high variety or low variety. It is thus possible to segment and characterize customers by their overall variety and their Gini coefficient, to
identify those who are consistently seeking variety, or for which categories does that particular customer seek variety.

In order to define the Gini coefficient $G$, let $L(u)$ be the Lorenz curve denoting the percentage of the total variety generated by the lowest $u\%$ of product categories during a fixed time period. The Gini coefficient is defined as:

$$G := \frac{A}{A+B}, \quad A = \int_0^1 \left( u - L(u) \right) du, \quad B = \frac{1}{2} - A \quad \text{(Equation 6)}$$

Thus $G \in [0,1]$ and a value of $G = 0$ reflects diversity (all variety indices are equal) whereas values near one represent concentration (a small number of product categories have most of the variety seeking). The coefficient’s natural boundary between $[0,1]$ and it’s dimensionless nature is a very useful property that makes it ideal for comparing variety seeking across people and categories. In contrast, other inequality measures such as the coefficient of variance (CV) - which in this context would be the variety indices’ standard deviation divided by its mean – have no upper bound and thus make comparing and interpretation across customers difficult (De Maio 2007).

A distribution of the resultant Gini coefficients of the 848 customers is depicted in Figure 7:

![Histogram of Gini Coefficients of Variety Indices](image)

**Figure 7: Distribution of Gini Coefficients which show the extent of diversity of variety indices**

Figure 7 shows that most consumers do not have equal variety indices across categories: only a minority has a Gini coefficient lower than 0.2, indicating low inequality. From Figure 7’s overall distribution of heterogeneity, the majority of customers (N=808, or 95%) have a Gini greater than 0.2, with the median at 0.245, indicating customers indeed have different variety indices across categories. With this method, it is possible to identify at the category level which
consumers are valuable, and at a person-aggregated level, how consistent (via the Gini) that customer is across categories.

### 3.3.5 Customer Typologies

While the findings of RQ1.2 prove that an overall consumer typology is not sufficient to explain the heterogeneity of variety indices on a category level, this section now look at the other way around. In order to address RQ1.3, a combination of a consumer’s variety indices per category is used to establish an overall aggregated consumer typology.

Since the variety index of a single category cannot be used to predict the index of another category, the mean variety index of each customer is calculated, based on their category level variety indices.

For ease of illustration, a subset of the POS data was investigated: In particular, the top five categories in terms of purchase frequency were selected as well as those customers (N = 476) that have purchased products in all five categories. The selected five categories account for 22% of all purchases and 11% of the revenue – a large and managerially relevant set of categories. The aggregated mean variety index of the selected five categories per person were calculated and visualized; the result is shown in Figure 8.

![Figure 8: (a) A histogram of the customer’s average variety index. (b) A plot of the customer’s mean variety index and the degree of consistency, indicated by the Gini coefficient of the variety indices](image)

In the \( \mu_{\text{variety}} \) distribution of Figure 8 (a), the mean of \( \mu_{\text{variety}} \) (the grand mean) was 0.61, and the standard deviation 0.17. Those one standard deviation or more below the grand mean were classified to be overall low variety seekers, those one standard deviation or more above the grand mean as overall high variety seekers, and those in between as medium variety seekers. As a result, there are 301 (63.2%) medium variety seekers, 92 (19.3%) low variety seekers and 83
(17.4%) high variety seekers. These proportions are consistent with questionnaire typology data (Van Trijp et al. 1996), and show that there are some higher level similarities between consumers that can be leveraged for traditional, coarse-granular marketing methods. An additional test probed whether the distribution of $\mu_{\text{variety}}$ can be generalized with a normal distribution; the Kolmogorov-Smirnov test of normality found that the distribution of $\mu_{\text{variety}}$, $D(476) = 0.5944$, $p < 0.05$ was significantly different from a normal distribution, so the result is not generalizable with a parametric normal distribution. The distribution was also platykurtic (-0.804) and wider than the normal distribution.

To explain the effect of the heterogeneity, the population was visualized as a scatter plot in Figure 8 (b). Each dot corresponds to an individual customer. The x-axis shows the computed mean of the aggregated variety index for each customer, and the y-axis shows the corresponding standard deviation. The result is once again consistent with the findings of RQ1.2 and shows especially that there are some significant differences between customers and also within customers, especially in the group of “medium” variety seekers. This means that some average variety seekers are actually high variety seekers in some categories and low variety seekers in others, which has significant implication on the success of marketing measures. Meanwhile, one can observe that customers who exhibit high variety seeking are also more consistent in their behaviour, shown by their lower Gini coefficients. To further illustrate the heterogeneity, four different customers (shown as coloured dots in Figure 8 (b)) are shown in Figure 9 and elaborated on regarding their differences in terms of their typology.

![Figure 9: Four examples of customers with their overall variety seeking typology and their variety indices in the top five categories.](image)

As Figure 9 shows, customer A and D are fairly consistent in their behavior. In all five categories they exhibit the same level of variety seeking. However, customer B and C are completely different. While their overall typology both amount to medium variety seeking, they each exhibit different behavior across the categories. Customer C is a medium variety seeker in three of the five categories whereas customer B is a high variety seeker in two categories and a low variety
seeker in three categories. Thus, while customers can be described by an overall typology of variety seeking, there are also category level differences that have to be accounted for, which is provided for in the model introduced in this chapter.

3.3.6 Time Effects on Consumer Typologies and Overall Variety Seeking

The previous sections estimated the customer variety seeking and overall typologies by using the full one-year set of customer data. This section tests for the validity and consistency of the previous results when considering instead multiple, smaller time windows. Namely, the one-year set of customer data is split into four quarters – each quarter consisting of four months - and for each quarter, Equations 2 and 5 are estimated to characterize the overall extent of variety-seeking in a product category (Equation 2) and the individual-category variety index (Equation 5). In doing so, the consistency of the previous results can be affirmed (RQ1.3), the volatility of a customer’s overall variety seeking can be studied (RQ1.4), and a customer’s category-level variety seeking with time can be examined (RQ1.5). In order to determine the consistency within customers over a period of time, it is necessary to study customers who have shopped in every quarter. Accordingly, out of an original sample of 848 customers, this section studies the 838 customers who have made a purchase in all four quarters.

Probabilistic Model Fit for Variety of Goods Purchased

The first step was to check whether the Poisson model remains the best fitting model in the face of a reduced (i.e. quarterly) data set. As before, the total number of unique products bought by each consumer in the one-year data set was counted, and a Poisson, geometric and negative binomial distribution were fitted to this vector of frequencies for each product category, using maximum likelihood estimation (MLE). For each distribution fit, the Akaike Information Criterion (AIC) was computed in order to compare the goodness of fit of the models. The models were fitted using first the third level product categories. The results are summarized in Table 4.

<table>
<thead>
<tr>
<th></th>
<th>1st Quarter</th>
<th>2nd Quarter</th>
<th>3rd Quarter</th>
<th>4th Quarter</th>
<th>Whole Year Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>74%</td>
<td>73%</td>
<td>66%</td>
<td>75%</td>
<td>86%</td>
</tr>
<tr>
<td>NBD</td>
<td>30%</td>
<td>25%</td>
<td>23%</td>
<td>22%</td>
<td>14%</td>
</tr>
<tr>
<td>Geometric</td>
<td>2%</td>
<td>2%</td>
<td>4%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>No Fit (all customers bought only 1 variety of item)</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
</tr>
</tbody>
</table>
Table 4 show that the Poisson remains the best fitting distribution across all quarters, based on the AIC. Notably, with the reduced data set, compared to the whole year data set, more categories see NBD as a good fit as well. There were also categories for which all consumers only bought one item during that quarter, in which case no distribution could be meaningfully fitted in that quarter. This occurred 1% of the time, or 1 category, in the 4th quarter.

Accordingly, the Poisson distribution was chosen for all subsequent analysis.

Confirming and Characterizing the Heterogeneity of Variety Seeking Across Product Categories

As per Section 3.3.5, this section tests whether variety seeking was a trait that was category independent – in which case a given person should have the same extent of variety relative to the rest of the population for each product category (i.e. their variety index should not be significantly different across categories). The analysis was completed with the Skillings-Mack test (Skillings and Mack 1981). The Skillings-Mack statistic was significant for all four quarters (Q1: $T = 2438$, $p < 0.05$; Q2: $T = 2137$, $p < 0.05$; Q3: $T = 2758$, $p < 0.05$; Q4: $T = 2210$, $p < 0.05$) indicating that the product category did significantly change a person’s degree of variety seeking at a given period of time. RQ1.2 is thus further confirmed. This therefore justifies the need of computing a variety index at the granularity of person’s interaction with a specific category.

Consistency of Customer Typologies over Time

With within-customer heterogeneity established between the four quarters, this section now looks at how consistent are customers overall (i.e. in their overall typologies) across each quarter – their extent of temporal consistency. This was evaluated in two ways.

In the first test, for the sample of customers ($n=838$), the Cronbach’s $\alpha$ was computed between the variety indices computed at each of the four quarters. Typically in psychological and social science research, the Cronbach’s $\alpha$ is used to determine the extent of reliability or internal consistency between measures which are supposed to measure the same underlying construct (Field et al. 2012). In this application, the underlying construct is variety seeking, and thus if the repeated measures of the variety index have strong internal consistency, the Cronbach’s $\alpha$ should be high. The threshold of which Cronbach’s $\alpha$ should exceed in order to indicate sufficient internal consistency differs depending on the literature. Nunnally and Berstein (1994) propose a threshold of 0.7 while, as with Bagozzi and Yi (2011). Malhotra and Birks (2007) propose a threshold of 0.6 For this study, the higher threshold of 0.7 is chosen, which is most commonly applied in research practice. For $n = 838$, the results show strong reliability between the four
quarters for mean variety seeking, $\alpha = 0.88$. The results show that globally, there is strong temporal consistency in variety seeking between customers across the four different time points.

In a second test, in order to examine the consistency within customers at an more individual level, the Gini coefficient was calculated for each individual, based on his four overall variety indices in the four quarters. The Gini coefficient would characterize to what extent these four variety indices are equal to one another, with 0 indicating perfect equality and 1 indicating high inequality. The results are plotted as a histogram in Figure 10:

![Figure 10: Consistency over time of mean variety seeking](image)

Figure 10 shows that the distribution of inequality for mean variety seeking is heavily weighted towards equality – across the four quarters, 98% of customers have a Gini coefficient for mean variety seeking less than 0.1, indicating near equality. This confirms that at the individual level as well, there is strong consistency of mean variety seeking over time.

Recall also that in the one-year estimation of variety seeking, for each customer his consistency of variety seeking between product categories was computed. Accordingly, this consistency,
represented by a Gini coefficient, was also computed in each of the four quarters. From these four measurements of a person’s consistency, computation of the Cronbach’s $\alpha$ showed strong reliability between the four quarters, $\alpha = 0.78$. The Gini coefficient for these consistency measures were also computed and the results are plotted in a histogram in Figure 11.

![Distribution of Customers (n=838) by Gini Coefficients of Variety Seeking Consistency Over Time](image)

Figure 11: Consistency over time of variety seeking consistency

The results show that 94% of all customers have a Gini coefficient for variety seeking consistency less than 0.1, indicating near equality. This confirms that at the individual level as well, that customers are not only consistent across time for their extent of overall variety seeking, but also consistent across time in terms of how consistent they are in seeking variety between product categories.

RQ1.4 is thus addressed – in line with the psychology literature, variety seeking is indeed empirically stable over time- and the overall typologies under investigation in RQ1.3 are confirmed across multiple time points.

**Consistency of Category-level Variety Seeking**
The previous section found that overall variety seeking within a customer is consistent over time. RQ1.5 asks, within a given customer, how consistent over time is that person in variety seeking within individual product categories.

To answer this question, for a given product category, the variety index for each customer in that product category was computed at each quarter. From these four variety indices, the Gini coefficient was calculated to quantify the extent of temporal consistency exhibited by that person at the category level. For each product category, this results in a distribution of Gini coefficients describing the extent in which customers are consistent over time. The mean of these coefficients were then computed to obtain the average consistency in that product category. This process was then repeated for all 128 product categories.

![Distribution of Gini Coefficients](image)

Figure 12: Distribution of mean consistency of variety seeking exhibited within product categories

Figure 12 shows that 74 out of 128 categories (57%) have a mean Gini coefficient for variety seeking consistency less than 0.1, and 107 out of 128 categories (84%) have a mean Gini coefficient for variety seeking consistency less than 0.15. Thus the majority of customers exhibit highly consistent variety seeking over time, in the majority of categories. RQ1.5 is thus answered – customers are not only consistent over time in terms of their overall variety seeking, but are also consistent over time in the degree of variety they seek at the level of individual categories. The results also reinforce the overall trend of temporal consistency in RQ1.4.
3.4 Discussion

The findings confirm the need for a higher resolution perspective on variety seeking behavior. The study shows that the currently dominant approach of using a per-person typology is not sufficient to explain the differences in variety seeking behavior per category. These differences are, however, critical for successful marketing measures in mobile or social commerce. The presented approach allows for measuring variety seeking behavior on a category level and comparisons between consumers and across categories. The study showed that a variety index can be modeled nicely with a Poisson distribution and provided both confirmation and some characteristics of reference categories based on real point-of-sale data. Because often a high-level typology of a consumer is already sufficient for decision making, this study therefore also provided a method for reconciling the highly granular, consumer-product category perspective with the overall typology used in research and practice today. These overall typologies were also found to be consistent with time, confirming the position of the psychological literature. The consistency in behavior over time was also found at the category level: that is, although customers may differ from each other in terms of which categories they choose to seek variety in, within a customer at the category level there is strong consistency over time.

3.4.1 Applicability at Different Granularity Levels of Categories

The study confirmed that the variety of goods purchased in a product category for a customer fits a Poisson distribution. This relationship holds true even when the analysis was conducted across different granularity levels of product categories and also when the model was estimated at different time points. As such, the relationship is robust against the arbitrariness of product categories as set by the retailer and also the sampling window, and therefore can be generally applied in different grocery retail settings. The Poisson “storyline” is also intuitive; individuals have differing rates of variety seeking and satiation, subject to some stochastic process, and some probability of consuming a certain amount of variety - in-line with the evidence on satiation (Chintagunta 1999; McAlister and Pessemier 1982). It was also seen that the categories where variety seeking was high also had a large assortment of goods available, which confirms past psychological studies that merely having a wide assortment of goods can drive variety seeking (Diehl and Poynor 2010; Kahn and Wansink 2004; Read et al. 1995; Simonson 1990). The product categories which showed high variety of goods sold (ex. soft drinks and fruit yoghurt) also matched the categories selected for study in the literature (Trivedi 1999). The observed temporal consistency of an individual customer’s degree of variety seeking within a product
category also meant that once identified as a variety seeker a particular category level, this customer is likely to be receptive to try new offerings in that category.

3.4.2 From Category Level Variety Seeking to Overall Variety Seeking

Finally, it was shown that the category variety indices can also be generalized to determine a person’s overall inclination to seek variety, thus replicating the insights of a psychometric questionnaire. The result was also robust to timing; even when the variety seeking model was estimated in smaller time windows of quarters, the individual level of overall variety seeking remained consistent. Where the model differs from the questionnaire is that in addition to the overall picture, it was also possible to quantify individual variety seeking at the category level, thus characterizing the heterogeneity seen in the spread of standard deviations of $\mu_{\text{variety}}$.

Accordingly, it is expected that the model presented will be able to better explain consumer receptiveness to try a variety of goods, and also for designing category-level recommendations that cater to a consumers’ extent of variety seeking in a particular category.

3.4.3 Contribution to Research

With respect to the existing literature, the variety seeking model which is proposed in this thesis can be seen as a data-driven model of the psychological research on variety seeking (Baumgartner and Steenkamp 1996; Raju 1980; Steenkamp and Baumgartner 1992; Van Trijp et al. 1996). The consistency of the results over time confirms the position of the psychological research that variety seeking is a personal attribute. At the same time, the heterogeneity between categories within a customer adds an extra dimension – that of variety seeking category differences and consistency – previously not considered in the variety seeking literature, but confirming the literature which noted that customers have heterogeneous shopping preferences between categories (Allenby and Rossi 1998; Leszczyc and Bass 1998; Lim et al. 2005). The Poisson model and storyline also matches the satiation view on variety seeking (Chintagunta 1999; McAlister and Pessemier 1982) and the finding that high variety seeking occurred in categories with large assortments confirm past studies (Diehl and Poynor 2010; Kahn and Wansink 2004; Read et al. 1995; Simonson 1990). The temporal consistency of variety seeking also confirms the view that variety seeking is a personal and consistent trait.

In the area of information systems, the results also overlap at a high level with the core idea of a collaborative recommender system, since it uses the behavior of others – the population distribution of variety sought in a product category - to make an inference about relative variety seeking in a customer. Furthermore, because the variety seeking model estimates variety
seeking at the category level, there is also an overlap with a content-based recommender system because in both cases, products are assumed to consist of attributes, and a product category is thus seen as a collection of exchangeable products with similar attributes – and therefore, the unit of analysis chosen for estimating variety seeking. As such, the estimation of variety seeking presented in this thesis also complements the recommender system research.

3.4.4 Contribution to Retailers: An Appropriate High Resolution Segmentation Scheme for Physical Grocery Stores

Past research indicated that variety seekers tend to increase overall consumption quantity (Kahn and Wansink 2004; Read et al. 1995; Simonson 1990) and are open to promotions (Ailawadi et al. 2001), and thus form a potentially valuable customer base for individualized segmentation in multi-channel marketing. This chapter’s results have found a data-driven method of identifying such variety seeking customers and their category-level preferences – it has shown that although customers can be described by an overall typology of variety seeking, category level differences in variety seeking exist within a customer, and also the extent of differences vary from customer to customer. Consequently, the combination of overall variety seeking and the consistency can lead to different marketing strategies for each individual customer. An example of how this insight can be put into practice is illustrated in Figure 13.
Figure 13: Mapping of customer overall variety seeking and consistency towards retailer strategy

Figure 13 shows the same subset of customers as discussed and selected in Section 3.3.5 for ease of illustration; they are the customers (N=476) who have made purchases in all of the top five product categories in terms of purchase frequency, where the selected five categories account for 22% of all purchases and 11% of the revenue. The mean variety index of the selected five categories per person and their associated Gini coefficient of the variety index were calculated and visualized onto a scatterplot. The corresponding recommended retailer action can be seen in Figure 13. For customers who exhibit high overall variety seeking and a low Gini, they are very consistent in their variety seeking behavior and therefore the retailer can recommend a product from any category to them – in this case, the retailer could narrow down which category according to another parameter, such as customer lifetime value or predicted profitability in a particular category. For customers who exhibit a high Gini, it means there are select categories where variety is sought, and therefore from a recommendation success and relevancy point of view, the retailer should only focus on categories where that individual seeks variety for a recommendation. Finally for those who exhibit a low Gini and a low overall variety seeking, it means those customers consistently have strong habits in most categories, and therefore it may be a waste of retailer resources to spend marketing effort in convincing these consumers to try new goods. In any case, regardless of which category is exhibiting high variety seeking, the retailer is free to act on this knowledge with a recommendation, a direct offer with a discount, or in the order of how results are presented when customers explore or browse a retailer’s product line. RQ1 is thus addressed: by a presenting an individual level assessment of variety seeking and its category-level consistency, a high resolution variety-seeking segmentation scheme appropriate for multichannel grocery retailing was devised, which enables
a retailer to target specific customer interests with offers. This scheme is data driven, can be implemented in a retailer’s customer relationship management system, and addresses the highly habitual aspects of customer shopping particular to grocery retailing.

The final implication of this chapter’s results for retailers is that the temporal stability of customer variety seeking at the overall level and at the level of product categories means that the estimation of variety seeking can already be performed with a quarter of purchase data; there does not need to be a long period of time to detect such customer inclination. The difference between the infrequently deployed questionnaires is that the estimation method presented here is data-driven and can be scaled up easily to a large base of customers; applying a variety seeking questionnaire for every category in which a customer shops, and for every customer in a retailer is, in contrast, a difficult endeavor. One limitation to note for this result is that it was estimated on a year of data – although the customer extent of variety seeking was consistent in this timeframe, it is quite possible that customers may change in their variety seeking over a longer time horizon – such as over several years. Future research could therefore test the variety seeking estimation on a data set with a longer time window, to see if variety seeking changes over a timeframe of years. If it does, then the data-driven estimation method presented would be even more valuable as compared to the questionnaires, since the change in variety seeking can be detected by regularly applying the estimation method as new transactions are recorded.

### 3.4.5 Accuracy of results

Although the results are already very promising and the effects are clearly visible, the results can be further improved. In addition to studying the effects in different stores and industries, four factors are relevant for future research. First, the study based customer identification on the loyalty card number. This card might be shared within a whole family and thus increase the noise level of the findings. A future study might control for this noise by attempting to identify the demographics behind the loyalty card holder, for example by combining a PoS data set with a retailer’s smartphone app usage statistics; since smartphones are personal, they can offer a finer resolution of individual identification when combined with PoS data. Secondly, seasonally offered products were excluded for the analysis. It might also be of interest to consider seasonal products and all product categories when studying variety seeking behavior, and that extending the automated measuring of variety-seeking to the whole population would be for future work. Finally, in the overall variety seeking aggregation, a simple (unweighted) aggregation of variety indices was used; since consumers have a finite budget, a consumer can only explore a finite
number of categories at depth, and as such, the overall variety index could appear much lower due to the presence of many sparsely explored categories. A future study could propose and evaluate a weighted aggregation scheme, where for example categories with low relative spending are weighted less.

3.5 Conclusion and Future Research
This study was one of the first to develop a model for measuring variety seeking behavior on a high level of granularity. In particular, it showed how to obtain a high resolution classification based on individual consumers’ purchases in individual product categories can be used instead of the established psychological per person typologies typically operationalized by questionnaires, which is difficult to deploy to every customer. The study has the following main contributions. Firstly, it contributed to the customer segmentation research stream by providing a novel way for identifying variety-seeking customers and their specific variety-seeking behavior on a category level. Second, for the most important retail categories, the study characterized the extent of variety seeking and provided a data-driven approach that is easy to operationalize by practitioners – especially for deploying large-scale personalized marketing measures in social or mobile commerce in physical stores. For practitioners, the model is parsimonious and can be estimated on observable consumer purchasing data logged at the PoS; it does not depend on demographics, which are not always available to every retailer due to privacy regulations, nor does it depend on unobservable motivations. Thirdly, the study proposed a method to reconcile the granular category-level results with per person typologies used today, and thus provide also a starting point for further research. The study found that the per person typologies are stable over time, thus confirming the position of psychological research which held that variety seeking was a personal trait. Unlike the psychological questionnaires, however, the data-driven approach can easily identify variety seeking for all customers and categories, which would have been prohibitively expensive using questionnaires. Finally, the study addresses a gap in recommender systems research, which until now was largely based on the content of what consumers have bought or what others have bought; this model of characterizing consumers’ extent of variety seeking adds another dimension, that of personal inclination to try new goods. This opens up new research opportunities in combining the classical approaches in recommender systems with the insights of this chapter. For example, the method could help recommender systems identify which consumers seek more variety than others for specific product categories; in turn, once both have been identified, a classical content-based recommender system can be applied to decide what product to recommend.
Future work can integrate the model into a content-based approach for recommender systems in real-time decision making for a smart phone application. Finally, for computing overall variety seeking, a study with a weighted aggregation scheme should be investigated.

In order to answer RQ2 and determine how social learning marketing can be developed further to benefit physical grocery stores, this section presents and tests the idea of “sales velocity”. This idea is operationalized and key hypotheses are tested in three empirical studies. Recommendations are given on how such a form of marketing can be deployed with a multi-channel physical grocery retailer.

4.1 Introduction

In an increasingly crowded marketplace, retailers need innovative ways of promoting products to their consumers. To this end, E-commerce retailers have utilized observational learning (OL) – a form of social learning marketing – to great effect. In the context of marketing, OL occurs when a customer observes that others have bought or tried a product and thus makes an inference about the quality of a product, but has no further information otherwise. In E-commerce, this is typically operationalized as a list of top ranked products to promote product sales; the higher the sales rank, the more likely consumers buy that product.

Prior OL research assumed OL arises from observing a static outcome, such as the current sales rank of a product; however, prior research on intertemporal choice showed that people prefer outcomes with increasing trends over stable or decreasing trends. This suggests that observing an increasing sales rank, denoted as sales velocity, would have a positive effect on purchase likelihood.

Because unprecedented, in this chapter three studies are presented which test the sales velocity effect. Results show that sales velocity has a significant effect on the likelihood of purchases, reversing even participant preferences for a product with a higher sales rank. This effect is consistent across four broad products tested. For researchers, by joining the two previously disparate branches of research in OL and intertemporal choice, a gap in OL research is addressed, where the velocity dimension of OL had been ignored.

This chapter is laid out as follows. Hypotheses concerning the sales velocity effect are developed. Three studies that explore the effect of sales velocity on consumers’ likelihood of purchase are presented and discussed, and the boundaries of this effect are explored. By exploring and understanding sales velocity in the different studies, practical implications are derived and the
limited of this research are presented. Finally, the chapter is concluded and opportunities for future research are presented.

4.2 Research Framework

4.2.1 The Sales Velocity Effect

As discussed in Section 2.2, the literature on intertemporal choice suggests that observing the velocity component of others’ actions could have an effect on purchase decisions, since it represents an improving outcome. Accordingly, this thesis proposes that the higher the sales velocity of a product, the more likely a consumer will choose it, all else being equal. However, in any choice set, realistically it is unlikely that the choices are equal in rank. Therefore, any study of the sales velocity would have to account for differences in the current rank between choices. Thus the main hypothesis is formulated as:

**H1**: Given a choice of similar product alternatives, when the sales velocity is high (low), the likelihood of purchase of that product is higher (lower).

While the first hypothesis is aimed at establishing the effect, the following hypotheses further explore some boundary conditions that might boost or reduce the effect of sales velocity on likelihood of purchases.

4.2.2 The Role of Numeric Framing

The sales velocity cue can play a major role in influencing consumer choice. It can be presented in many different ways including as different units (ex. as percentage or in raw numbers), icons (ex. an up arrow), colors (ex. a green number for the velocity), and a combination thereof. In this chapter, the focus is on unit framing, which was widely established in research and practice and used in the past for optimizing the effect of price discounts; thus one can expect a similar result on the framing of sales velocity information. The idea of how choice problems are framed can affect cognitive judgment and preferences dates back to the work by Kahneman & Tversky (1984), which presented an example on how the way a number is represented influences perception of the underlying quantity. In the marketing literature, much focus has been on price discount framing (Chen et al. 1998; DelVecchio et al. 2009; Krishna et al. 2002) – i.e. how the numbers displayed in a discount influences its evaluation by customers. In the study by Chen et al. (1998) for example, they found that consumers perceived a price reduction in absolute dollar terms to be larger than the same price reduction framed in percentage terms, when the number displayed for the price reduction was larger than the number displayed of the percentage
reduction. In other words, the unit and framing had a significant influence on people’s perceptions of the quantity, and taking this logic into the area of rank change, it could be expected that the same to be true here as well. Namely, given a sales velocity metric, then one could expect the representation that leads to the largest number to have the strongest effect. The phenomenon exhibited in Chen et al. (1998) can be described by the numerosity effect (Bagchi and Li 2011; Chen and Rao 2007; Kruger and Vargas 2008; Monga and Bagchi 2012; Pandelaere et al. 2011; Zhang and Schwarz 2012); namely, when something is expressed in alternative units, the perception of magnitude increases if the unit is on a finer grained scale. In Pandelaere et al. (2011), it was shown that consumers see a bigger difference between two products when the warranty information was represented in months than years. The effect is attributed to people focusing on the numbers rather than the units. This effect has been largely ignored in E-commerce and since sales velocity can be represented in different forms as well, one could expect that the effect of sales velocity as hypothesized in H1 to be higher when the sales velocity is expressed in a unit that leads to larger numerical values. Therefore, the following hypotheses are proposed:

\[\text{H2a) Consumers perceive an objective rank change of a product to be higher when it is expressed in a unit that leads to larger numeric values (and vice versa).}\]

\[\text{H2b) A product evaluated to have a high (low) rank change has a stronger (lower) sales velocity effect on the likelihood of purchase.}\]

4.2.3 The Negative Sales Velocity Effect

Up until now the focus was primarily with boosting sales due to positive sales velocity in OL. Given a period of time, a product can increase in sales rank (positive sales velocity) but also decrease in sales rank (negative sales velocity). This provides an important boundary of the sales velocity effect: since sales velocity is not yet widely implemented, retailers should know the effects of negative sales velocity, before making a decision to implement sales velocity in general. The literature on loss aversion suggests that losses are perceived to be stronger than gains of equivalent size and that the perception of the loss is relative to the position from where the loss occurred (Hardie et al. 1993; Kahneman and Tversky 1984; Tversky and Kahneman 1991); accordingly one would expect a negative sales velocity to have a larger negative impact on the likelihood of purchase, than the reverse case with positive sales velocity. Accordingly, the following hypothesis is proposed:
H3a: Given a choice of similar product alternatives, a product with a decreasing (increasing) sales velocity has a lower (higher) likelihood of purchase.

H3b: Given a choice of similar product alternatives and the same relative sales velocity, the effect of the decreasing sales velocity on the likelihood of purchase is stronger than the effect of increasing sales velocity.

4.3 Overview of Studies

Three studies were conducted to test the hypotheses. Study 1 consists of two parts: the first part tests the basic sales velocity effect, namely, whether consumers have a higher likelihood in purchasing a product when the sales velocity is high (H1). The second part of Study 1 tests for the numerosity effect to see whether the framing of the sales velocity information can lead to changes in the perception of the sales velocity (H2a) and subsequently to changes in the likelihood of purchasing a product (H2b). In Study 2, the generalizability of the sales velocity effect in the case of different focal products is tested. In Study 3, the effect of negative sales velocity on the likelihood of purchases (H3a/b) is tested.

4.4 Study 1: Establishing the Sales Velocity Effect

To investigate H1 and H2a/b, an online consumer choice experiment was conducted as follows.

4.4.1 Task

Consumers often have well-established preferences for products (Hoyer 1984), making it difficult in a consumer choice experiment to select a “preference neutral” stimuli with minimal confounds.

Although in a real-world scenario, such preferences would inevitably be present, in a first experiment it is desirable to control against them. For this reason, the following consumer choice experiment was based on the agent/principal task common in the marketing literature (for examples, please see Ariely 2000; Diehl & Poynor 2010; West 1996), where the participant (the agent) has to make the best choice on the behalf of another person (the principal) based on the principal’s preferences which are transparently given to the participant. In this manner, the aim was to neutralize the effect of individual attitudes or individual involvement to a particular product or product domain.

Accordingly, participants read a scenario that asked them to imagine that they wanted to treat a good friend who likes chocolate. They were told this friend has tried out a wide variety of chocolate before, including the most popular ones, so therefore the friend wants to try
something new and refreshing. The scenario then told the participants to imagine that they subsequently checked an online E-commerce store serving their market, scrolled through the options, and now have to consider two offerings.

Subsequently, participants were given a tutorial where they were taught and tested on how to interpret the rank information on the E-commerce store. It should be noted that most E-commerce sites do not force onto or give users a detailed explanation of their popularity metrics and how they are computed; however for internal validity and to ensure the users would understand the information in the task, this tutorial was provided. After the tutorial, participants were presented with the stimuli.

4.4.2 Stimuli Development
Selecting the Focal Product

The domain of retail, online and offline, covers a wide variety of domains and products, ranging from clothes to electronics to food. Without loss of generality, since a principal-agent task was implemented where personal preferences are taken out of the equation, for the purpose of this experiment, chocolate was chosen as a focal product. Because chocolate is sold at both online and offline physical retailers (ex. Walmart sells chocolate both in their online store and also in their bricks and mortar chains), the results of this chapter can be relevant to both channels. Furthermore, chocolate is also an example of a repeatedly-bought and hedonic good (Voss et al. 2003), which facilitate variety seeking behavior (Van Trijp et al. 1996); as discussed in the hypothesis formation, popularity change might aid choices in a variety seeking situation.

Since the stimuli will consist of sales rank data, a year of receipt data from a European physical grocery retailer was analyzed in order to (a) have a basis for realistic sales rank numbers, (b) validate how realistic it is to use sales velocity to promote products like chocolate, given its actual sales behavior, and (c) to also determine the potential impact of such a promotion. The receipt data comes from a single store, and consists of over 150,000 unique receipts covering a total of two million transactions, which map to 19,374 unique products sold that year. First, how often these products were sold were examined, since the stimuli deals with sales rank data. These 19,374 products sold followed an extreme long-tail distribution in terms of aggregate sales; that is, only a few products have a large frequency of sales (the “hits”), while the majority sell infrequently (the “niches”) (Anderson 2006). The distribution of the first 500 products are depicted Figure 14. In Figure 14, the circles depict a product’s average sales per week and the associated sales rank. For that same product at the given sales rank, the triangles then indicates
the average rank change of the product. What becomes immediately apparent from this
distribution is that the high ranking products exhibit a very low rank change – i.e. they sell well
each week and tend to stay that way throughout the year.

Figure 14: A product’s average sales per week and the associated sales rank. For that same product at the given sales rank, the triangles then indicates the average rank change of the product, and the dashed box the mid tail.

Note that Figure 14 contains the retailer’s products, irrespective of which product category the product belonged to; to check the generalizability of the sales rank and sales rank change distributions, the sales distributions within each of the retailer’s 900 product categories were also examined; it was verified that the average sales and sales rank change trend also holds true at the category level. Namely, that the average sales follow a long tail distribution and the maximum sales rank change grows the further from the head of the distribution. An excerpt is shown in Figure 15.

Figure 15: Average sales and maximum rank change distributions for select product categories.
Since there are no universally accepted cut-off points between the niches and the hits for the sections a long-tail distribution\(^4\), for the purpose of further discussion and analysis, the following definitions are used: a product is a “hit” if it is in the top 1% in terms of total number of products sold in a year, in the long tail if it is in the bottom 90% of products, and those in between are called the mid tail. Some characteristics from the retailer are shown in Table 5.

### Table 5: Product characteristics for “hit”, “mid-tail” and “long-tail” products for a physical retailer

<table>
<thead>
<tr>
<th></th>
<th>Average Week-to-Week Rank Change</th>
<th># Products in this category</th>
<th>Average Units Sold per Week</th>
<th>Combined Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>“Hit”</strong></td>
<td>171.5</td>
<td>192</td>
<td>&gt;30.3</td>
<td>19%</td>
</tr>
<tr>
<td><strong>“Mid-Tail”</strong></td>
<td>726.4</td>
<td>1742</td>
<td>30.3&gt;= units sold &gt;=5.61</td>
<td>41%</td>
</tr>
<tr>
<td><strong>“Long-Tail”</strong></td>
<td>844.7</td>
<td>17440</td>
<td>&lt;5.61</td>
<td>40%</td>
</tr>
</tbody>
</table>

As Table 5 shows, the long-tail products make up a large segment of the retailer’s revenue and have a large sales velocity – accordingly they could be well promoted with sales velocity, but promoting a long-tail product is not helpful to the physical retailer, who has limited inventory. Therefore, a long-tail product would be inappropriate as a focal product in this study. Similarly, the hits in the distribution are also inappropriate as a focal product. Although the hits in the long-tail head sell well (over 30 units per week), they exhibit low sales velocity, form only 19% of the sales revenue, and furthermore are often items which consumers already buy regularly - for the data of this retailer partner, 75% of the receipts contain an item from the top selling 1% of all products. This is not a problem in the E-commerce context for products which consumers infrequently buy (ex. cameras), since consumers are unlikely to have bought the recommended popular product. However, for the grocery retail domain at least, a top seller list could suffer from irrelevancy since - as noted in the dataset in this study - most of the top selling products are already bought by the majority of consumers, so from the consumer point of view, there is nothing new. For the retailer, this leads to no additional new purchases. Furthermore, the insights from the field of recommendation systems (Adomavicius and Tuzhilin 2005; Zhou et al. 2010) caution against continuously recommending products that are too close to what a user already buys, as it reduces the relevancy of the recommendation. Thus, a classical OL metric which would expose hit products is neither appropriate for physical grocery retailers or their customers and thus a hit product should not be used as a focal product in this study either.

\(^4\) for differing models, please see Brynjolfsson, Hu, and Simester (2011) and Elberse (2008)
In contrast, mid-tail products exhibit high sales velocity and additionally make up a large segment of the retailer’s revenue; the differing sales velocity and revenue characteristics between hits and mid-tail products suggests that sales velocity can be an effective form of marketing persuasion in promoting specific mid-tail products, thereby increasing revenue. Sales velocity avoids the aforementioned problem of promoting the most popular products which consumers already buy – a problem that did not exist for infrequently bought E-commerce goods, where OL has historically been deployed. In this data set, it was found that chocolate is in the mid-tail, exhibiting high sales velocity. Taken together, these results suggest that using a mid-tail product (like chocolate) in the experiment has revenue potential for retailers, is valid with respect to actual sales data, and therefore an appropriate choice.

**Development of the Choice Profiles**

Each chocolate offering was presented as a profile consisting of the product, the price, the product description, the product rank and - depending on the experimental group - a different representation of the product sales rank change. The rank change information varied depending on which of the experimental groups the participant was in. Participants were randomized into either the control group (shown product rank only), the rank change group (product rank + rank change), or the numerosity group (product rank + rank change in percentage). The profile pairs within each experimental condition were designed to be as similar as possible except for their rank change metric and its representation. This would allow the study of the effect of rank change and its framing: comparing the rank change group with the control group addresses H1, and comparing the rank change group with the numerosity group addresses H2a/b. A decomposition of the product profiles and experimental groups are depicted in Figure 16.
Figure 16: Components in a product description; the descriptions were made as similar as possible, and parameters such as the current rank, past rank and choice set size were controlled for in a pre-test.

The product profiles components, listed below and in were designed and pre-tested (n=184) for similarity:

- **Choice Set Size**: Here the choice set size refers to the size of the market for that category when browsing an E-commerce site. For sites such as Amazon.com, for categories like cameras, there are thousands of cameras on offer, while for memory cards there are only hundreds. The assumption is made that the current rank and the past rank (see below) are only meaningful relative to the choice set size (i.e. a product with a sales rank of 9 in a market where there are 10 products is probably perceived worse than when the market has 1000 products). A pre-test that was conducted found that having a larger choice set size minimized the perceived difference between consecutively ranked products, but beyond 100 products the perceived difference did not change. Therefore, a choice set size of 100 was chosen.

- **Current Rank**: The two products in each comparison received a rank of 50 and 51 respectively; the pre-test showed that participants (correctly) identified 50 as the better ranked product and perceived the difference between 50 and 51 to be small. This will be confirmed in the main study with a manipulation check.

- **Product Descriptions**: To minimize differences between the profiles except for the ranking information, the product names were generically chosen as Chocolate M or N, their respective prices were set equal to each other, and the product descriptions were
written with only subtle variations between them. The pre-test confirmed that the product description text was indeed perceived as similar.

Finally, for the manipulation itself, within each product pair the aim was to polarize rank change between the offerings. For the rank change component, the year of purchasing data from the grocery retailer was examined and it was found that chocolate in the mid-tail could fluctuate up to 100% in rank change, while more popular chocolate does not change in rank at all week to week. Accordingly, one product received a rank change of 0% while another received a rank change of 96%. The % rank change was calculated as \((\text{Past Rank} - \text{Current Rank})/\text{Current Rank} \times 100\). To cue the consumer that the rank change was positive, an up arrow was placed next to the change; for rank change, since most popular E-commerce websites use arrows, this de-facto standard was used, which consumers already know. Note that other definitions of calculating this % rank change could exist; however since this study examines numerosity from the point of view of designing a proper stimuli, only the end numerical magnitude of the rank change measure is sufficient. To create a conservative experimental design, the less popular product was assigned the higher rank change and the more popular product zero rank change. Therefore, if people choose the product with the higher rank change, it would be in spite of it being the less popular product.

4.4.3 Participants

120 participants from an online panel (Amazon Mechanical Turk, https://www.mturk.com) (\(\mu_{\text{age}} = 32, \sigma_{\text{age}} = 11, 45\% \text{ female}\)) were recruited for the study.

The participants were paid a token amount for the study and were mostly from the United States (95% from USA; the rest from the United Kingdom, Ireland and Canada). In line with Mason & Suri (2012), responses from 6 participants were recruited, whose responses did not pass attention-check questions in the survey, resulting in 114 samples (a 95% pass rate). The reliability of participants and the demographics are in-line with past studies conducted on the Mechanical Turk (Downs et al. 2010; Goodman et al. 2012; Mason and Suri 2012) which showed that Mechanical Turk participants were as reliable as a “traditional” panel and could replicate classical behavioral experiments. There were no significant differences in results between genders.

4.4.4 Measures

After being presented with the stimuli, the participants then had to first indicate which of the two products they were more likely to buy, i.e. they were asked to indicate the likelihood of
purchase. In line with the study by Kruger & Vargas (2008), this was measured on an unnumbered seven point scale, anchored by product M on one end and product N on the other end of the scale.

Subsequently, on another screen participants had to compare the two profiles in terms of sales rank change; they had to indicate which product they thought had the higher sales rank change. This was measured in the same manner as the likelihood of purchase. This measure served to determine whether the numerosity effect changed the perception of the sales rank change.

Attention and Confounding Checks

On the same screen as the sales rank change comparison, participants had to compare the two profiles in terms of price and sales rank; for each profile pair, they had to indicate which product they thought had the higher price and higher sales rank. These too were measured on the same scale and anchors as the main measures. Price served as an attention-check question, since the two offers were equal in price in all groups; therefore the correct answer would be a “4”. Those who failed this question were not considered for analysis in the study. Sales rank was a confound check; since the sales rank information remained the same across all groups, there should be no difference between them. The wording of the questions is given in the Appendix. The presented order of the profiles was counterbalanced between participants; the product with the higher rank was randomly assigned to product M or N. In the analysis, the product with the higher rank was labeled “M” (represented by a score of “1” on the seven-point Likert scale) and the score was reverse coded when the higher ranking product was assigned to “N”.

As another check, the participants were also asked questions concerning their involvement when it comes to chocolate purchases for themselves and also for others; the intention was to check whether the scenario selected had an involving product (chocolate) and situation (purchases for another). The scales are from Chandrashekaran (2004). Finally, the participants were asked how large they thought the differences were between the product profiles on a seven-point Likert-scale ranging from very small (1) to very large (7). Since group B and group C share the same content as A, with only the addition of the rank change information, in order to isolate the effect of this manipulation, it would be desirable for the perceived difference in group A to be significantly small.
4.4.5 Results

Prior to the main analysis, several checks were conducted. First, since most of the scale items were measured with the same method (a seven point Likert scale), the common method bias was tested using Harman’s one factor test (Podsakoff and Organ 1986; Podsakoff et al. 2003). Accordingly, a principal component analysis was conducted for the variables in the study and multiple factors with eigenvalues greater than one were found; accordingly, no single factor explained the majority of variance and the common method bias was not significant.

Second, prior to the main analysis, a test was conducted to check for homogeneity of variances and normality for responses to the main measure “likelihood of purchase”, and also for the checks for sales rank change and sales rank. The Levene statistic for testing homogeneity of variances was significant ($p < 0.05$) so therefore, the variances between groups differs significantly. Additionally, the Shapiro-Wilk test was significant ($p < 0.05$) so the data is not normally distributed. Accordingly, the assumptions of ANOVA and the t-test are violated and thus the non-parametric Kruskal-Wallis and Mann-Whitney U tests were conducted on the likelihood of purchase measure and for the sales rank change and sales rank checks when looking for significant differences between groups.

Attention and Confounding Checks

For the price attention checking question, a sample of $n = 114$ (95% of participants) passed and were considered for further analysis.

For the sales rank, since the current rank was the same across all groups, there should be no differences across experimental groups. The Kruskal-Wallis test confirmed this and did not show a significant effect of the experimental group on the reported sales rank ($H(2) = 3.326, p = 0.190$). Pairwise Mann-Whitney U also did not show any significant effect either ($p > 0.0175$ for all pairwise comparisons between groups). Therefore, the sales rank was perceived (correctly) to have no significant differences across experimental groups.

---

5 A Bonferroni correction of $1/n_{\text{comparisons}}$ was applied to control for the Type 1 error that arises from multiple comparisons; with three comparisons, all Mann-Whitney U effects were thus reported at a 0.0167 level of significance to obtain an even more strict requirement for significance.
For the involvement check in chocolate purchases, involvement in purchases for one self was high \((\mu = 5.12, \sigma = 1.65)\) as well as for others \((\mu = 4.92, \sigma = 1.48)\), suggesting that the scenario depicted, as intended, an involving product (chocolate) and situation (purchases for another). Finally, for the last check, participants were asked how large they thought the differences were between the product profiles on a seven-point Likert scale from very small (1) to very large (7). For group A, which formed the basis profile pairs for all other groups \((\mu = 1.81, \sigma = 1.25)\), a one-sample t-test showed the perceived difference lies significantly below the neutral scale value of four, \(t(41) = -11.32, p < 0.001^*\). This assured that the base profile pairs generated were, as intended, very similar, allowing one to infer that the differences in the other experimental groups came entirely from the manipulation.

The Sales Velocity Effect

In order to test H1, a Kruskal-Wallis test was conducted between groups for the likelihood of purchase. There was a significant effect of the experimental group on the reported sales rank \((H(2) = 20.22, p < 0.001^*)\). Subsequently, two Mann-Whitney U tests were conducted to compare the control group A with the two groups where sales velocity was present (groups B and C) to find where this effect was. A summary of the descriptive statistics and Mann-Whitney U tests for the main dependent measure is in Table 6. A total of 114 subjects participated after filtering from the attention check questions. For interpreting the descriptive statistics, recall that all outcomes were measured on an unnumbered seven point scale, anchored by “Definitely product M” on one end and “Definitely product N” on the other end of the scale, with “No Difference” in the middle. “Definitely product M” was coded as a “1” and “Definitely product N” was coded as a “7”.
Referring to Table 6, the paired Mann-Whitney U tests also show a significant difference between the control group A and the numerosity group C, and between control group A and the rank change group B. In other words, both groups where the rank change manipulation was present (groups B and C) were significantly different than group A. Descriptive statistics (see Table 6) for the three groups show that in group A, people prefer the higher ranked product, are less sure in group B, and show a preference for the lower ranked product under a numerosity framing in group C.

From the empirical data, H1 is therefore supported; rank change has a significant effect on the likelihood of purchases and that it weakened the likelihood of choosing the higher ranked product.

The Numerosity Effect

In order to test the second hypothesis, a Kruskal-Wallis test showed that there were overall differences between the three groups for the perceived rank change (H(2) = 66.680, p<0.001*). Pair-wise Mann-Whitney U tests were conducted to see where the differences lie. The results are given in Table 7.
Table 7: Summary of paired Mann-Whitney U statistics (given by U) for the sales rank change. The * indicates significant p-values after a Bonferroni correction was applied

<table>
<thead>
<tr>
<th>Sales Rank Change</th>
<th>Groups Involved</th>
<th>Result</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptives µ (σ)</td>
<td>Group A (n=42)</td>
<td>3.31 (0.52)</td>
<td>The perceived rank change for numerosity group C is larger than B.</td>
</tr>
<tr>
<td></td>
<td>Group B (n=32)</td>
<td>5.72 (1.55)</td>
<td>Furthermore, p&lt; 0.0167* for “B vs. C” so, H2a is supported.</td>
</tr>
<tr>
<td></td>
<td>Group C (n=40)</td>
<td>6.50 (1.28)</td>
<td></td>
</tr>
<tr>
<td>Mann-Whitney U</td>
<td>A vs. B</td>
<td>U =145, r = 0.71, p&lt;0.001*</td>
<td></td>
</tr>
<tr>
<td>Tests U, effect size r, p</td>
<td>A vs. C</td>
<td>U= 83.5, r= 0.82, p&lt;0.001*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B vs. C</td>
<td>U=457, r=0.30, p&lt;.0167*</td>
<td></td>
</tr>
</tbody>
</table>

In Table 7, descriptive statistics show that in group A, when no rank change information was presented, participants simply chose the higher ranked product (µ= 3.31, σ = 0.52), but correctly identified the higher rank change product in groups B (µ= 5.72, σ = 1.55) and C (µ= 6.50, σ = 1.28). Notably, there was a significant difference between groups B and C, which showed that the numerosity effect was present in C and did lead to a difference in perceived rank change.

These results provide evidence that the manipulation of the rank change was successful – and also that the numerosity effect significantly changed the perception of the rank change. Thus, H2a is supported: the numeric framing influences perception of sales velocity too.

To address H2b – i.e. whether this change in perception of sales velocity leads to a stronger sales velocity effect on the likelihood of purchase – a Mann-Whitney U test was conducted between the control group and the numerosity group (U=83.5, r=0.82, p<0.001); the test showed that the numerosity group C is significantly different from the control group in the likelihood to purchase, indicating that with numerosity present, the sales velocity effect is still significant. Furthermore, as per Borenstein et al. (2011) the effect sizes for the two treatment groups were computed. It was found that the effect size (0.50) is also larger than the case without the numerosity framing (0.26). Thus, H2b is supported. Mean values of the likelihood of purchase for group C (µ= 4.53, σ = 1.88) is also larger than for group B (µ= 3.69, σ = 1.99). These results are also summarized in Table 7.
4.4.6 Discussion

For the first hypothesis, the sales velocity, operationalized by the rank change, had a positive effect on the likelihood of purchase, as evident by the significant differences of the two rank change groups (group B and group C) with respect to the control - this confirms the main hypothesis (H1). This result bolsters the results in intertemporal choice that, indeed, the velocity of an improving outcome positively influences consumer evaluation of choices, and that this outcome can be operationalized in the rank change metric.

The result also reveals and addresses a large gap in OL research, which previously only considered current outcomes arising from others’ choices, without factoring in the velocity dimension of these outcomes. Counterintuitively, the positive signaling from an improving product was even able to reverse the preference for a higher ranked one: in the control group, nobody chose to buy the worse ranked product, but a significant number of those in the rank change groups did. This preference reversal occurred even though the sales rank heuristic is much more commonplace in practice, and therefore more familiar to consumers. The implication is that this metric can be leveraged by retailers to promote mid-tail products, which are not necessarily the most popular, but typically exhibit good gains in sales rank. The results also extend the findings of Briggs, Landry, and Daugherty (2010) who showed that velocity metrics of performance is positively correlated with a firms’ perception of the providers; in particular, the study showed with a controlled experiment that there is also an influence on choices, in a context where retailers can use the result for their marketing initiatives. The results also suggest that the general findings of studies in velocity outcomes from Hsee, Abelson, and Salovey (1991) also hold true in the context of marketing.

For the second hypothesis H2a, there was a significant difference in the perceived sales rank change between the rank change group B and the numerosity group C, which showed that the numerosity effect was present and did lead to a significant difference in perceived rank change (H2a is thus supported); also, group C’s higher effect size and a greater likelihood of purchase, as evident by the increased mean, supports H2b.

4.5 Study 2: The Sales Velocity Effect with Different Products

Study 1 showed that effect of sales velocity is quite strong for a particular type of product – in this case it was the highly hedonic and relatively fast moving class of chocolate. However, the effect of sales velocity might be less or more pronounced depending on the characteristics of the product marketed.
It should be noted that from an economic perspective of observational learning, the role of the product characteristics should play no role in the observation and interpretation of other shopper’s actions; the positive signal that is inferred from observing others’ choices should only be a function of how many others have chosen that same product (Chen, Wang, et al. 2011; Duan et al. 2009), and as per Study 1, the velocity at which was chosen. However, since each of these economic studies had focused on a single product category (cameras and software respectively), they might have ignored the differential effects of observational learning that may exist across product categories. There is a research gap in systematically controlling for different product characteristics in doing a comparative study of the effect of observational learning.

Study 2 aims to address this. Since there are an infinite number of products available, it is therefore useful to apply a product classification for the purpose of studying the boundaries of the sales velocity effect. This study extends the results of Study 1 to include four additional products sampled from two common product classification paradigms; the hedonic paradigm and the non-durable goods paradigm. In the hedonic paradigm, a hedonic good is one that “relates to the multisensory, fantasy and emotive aspects of product usage experience” (Hirschman and Holbrook 1982; Holbrook and Hirschman 1982) – that is, there is an affective component from consuming the product. Examples of such products in the area of grocery retail for example are chocolate, mixed nuts and bubble bath (Chandon et al. 2000). It has been shown in the benefit-congruency framework by Chandon et al. (2000) that the extent of hedonism in the product affects how consumers respond to the type of sales promotions (which they classified as utilitarian – ex. price cuts - or hedonistic like free gifts); the promotion is most effective when the promotion and product type are matched in hedonism. For a product that is highly hedonic, such as chocolate for example, a promotion is more effective and satisfying when they provide intrinsic stimulation, fun and self-esteem, such as giving a free sample of the product, which encourages discovering and exploring a previously untried product. On the other hand, price cuts are seen to have low hedonic benefit and therefore are not as effective for these hedonic products. Therefore, a product’s extent of hedonism influences the type of promotion that is most effective for it. Accordingly, this study investigates whether the sales velocity effect still holds across products of differing degrees of hedonism.

The second paradigm which is investigated is the paradigm of the durable good vs. the non-durable good, and concerns the frequency of which the associated product category is

---

6 Due to the many similarities between Study 1 and 2, for parsimony the manuscript will note where the procedure and manipulations of Study 2 are the same in Study 1, and refer to the pertinent sections of Study 1.
purchased. From a consumer point of view, a durable good is defined as one that lasts multiple time periods; a consumer who buys a product can continue to use it over time, hence a purchase in the present serves as a substitute for a purchase in the future (Mantena et al. 2012; Waldman 2003). As such durable goods are often not purchased as frequently as non-durable goods and they tend to be more expensive. A non-durable good is the opposite and is one that is purchased frequently and consumed almost immediately upon purchase, and therefore tends to be non-durable. Often cited examples of durable goods include computers and cars, while in contrast, food items are often cited as non-durable goods. It had been found that whether a product is a non-durable or a durable good influenced consumer perceptions of various relevant marketing heuristics. Völckner & Hofmann (2007) found that for durable product, compared to a non-durable good, consumers had a weaker association of a high price being linked to high quality, but this difference decreased with product familiarity; this is in line with the finding that suggested consumers resort to heuristics often for non-durable goods (Bearden 1982; Burke et al. 1992; Hoyer 1984). Estelami & De Maeyer (2004) found that consumers had a much lower knowledge of price of goods for durables than non-durable goods. Derbaix (1983) found that consumers’ perception of risk and uncertainty for purchases differed depending on whether it was a durable or non-durable good. As such, the innate frequency of purchases of a product category matters and influences consumer interaction with these products. Thus the study will also examine whether the sales velocity effect holds across products of differing purchase frequency.

4.5.1 Task
The task was exactly the same as Study 1, except the product which the participant was shopping for was randomly drawn from a list of four products. In other words, participants were assigned randomly to one of eight groups; for each of the four products compared, there was a control group where only rank information was given and a sales velocity treatment group where additionally rank change information was given. As per Study 1, the participant proceeded with an agent/principal task and read that they were buying a particular product on behalf of a relative on an E-commerce site. They were then taken to a page that presented two offers of the product in question and subsequently asked to indicate which offer they would buy. After indicating which product they would buy and answering subsequent questions, the task was over.

4.5.2 Stimuli Development
Four product profiles were developed, whose selection and design are explained as follows.
Development of the Choice Profiles

As per Study 1, each product offering was presented as a profile consisting of the product, the price, the product description and in the control groups, the product rank. For the treatment groups, the sales velocity information of the product was also shown.

The profile pairs within each experimental condition were designed to be as similar as possible, except for the rank change metric and its representation. Unlike Study 1, only one representation of the sales velocity across all groups was used – the numerosity representation used was group C in Study 1, since in Study 1 this manipulation was found to have the strongest effect on the likelihood of purchase, and as such, this would be the recommended representation in a practical setting. The presented order of the profiles was counterbalanced between participants.

Selecting the Focal Product

The four product categories which represent the extreme points of the hedonic and durable axes were selected as follows. For the non-durable, fast moving goods, consistent with Laurent & Kapferer (1985), two products were selected from the hedonic scale: chocolate (high degree of hedonism) and laundry detergent (low degree of hedonism). Both of these product categories are considered to be non-durable goods in past studies that have employed a durable/non-durable goods paradigm (Derbaix 1983; Harlam et al. 1995; Holbrook and Hirschman 1982; Yeo and Park 2006). For the durable goods, consistent with Grewal et al. (2004), a vacuum cleaner was selected as the non-hedonic product (Laurent and Kapferer 1985) and televisions were selected as the hedonic product (Voss et al. 2003; Zheng and Kivetz 2009) which, as per Laurent & Kapferer (1985), scored similarly in hedonism to chocolate.

4.5.3 Participants

366 American participants were recruited from an online panel (Amazon Mechanical Turk, https://www.mturk.com) ($\mu_{\text{age}} = 31$ years, $\sigma_{\text{age}} = 10$, 40% female) for the study. The groups and number of participants in each are as follows.
### Table 8: Groups and number of participants in Study 2

<table>
<thead>
<tr>
<th>Products</th>
<th>Vacuum Cleaner</th>
<th>Detergent</th>
<th>TVs</th>
<th>Chocolate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td>41</td>
<td>41</td>
<td>47</td>
<td>45</td>
</tr>
<tr>
<td>Sales Velocity Group</td>
<td>45</td>
<td>46</td>
<td>46</td>
<td>39</td>
</tr>
</tbody>
</table>

#### 4.5.4 Measures

The likelihood of purchase was the main measure; after being presented with the stimuli, the participants then had to first indicate which of the two products they were more likely to buy. The scales were as per Study 1.

**Manipulation, Attention and Confounding Checks**

Participants also had to compare the two profiles in terms of price, sales rank change and sales rank. Price served as an attention-check question and sales rank and sales rank change were included as a manipulation check. Participants were also asked how large they thought the differences were between the product profiles Q and R as a further manipulation check.

In addition to the measures of Study 1, user perceptions of the product categories’ extent of enjoyability and its purchase frequency were both measured on a semantic differential scale, to verify that the selected products that were distinctly hedonic/non-hedonic and durable/non-durable.

#### 4.5.5 Results

**Manipulation, Attention and Confounding Checks**

For the price attention checking question, a sample of 353 passed (96%) and was considered for further analysis. As per Study 1, Sales Rank Change, Sales Rank and Intention to Buy were not normally distributed as per the Shapiro-Wilk test (p<0.05), so all comparisons for these measures between the control and treatment groups for each of the products were thus conducted with the Mann-Whitney U test instead of the t-test. The perceived differences between product profiles were compared in a one-sample t-test against the neutral value of “4”.

A summary of these checks are given in Table 9; all manipulation checks were successful.
Table 9: Summary of manipulation checks; the * indicates significant p-values (p<0.05) for the corresponding test

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Test of Manipulation</th>
<th>Vacuum Cleaners</th>
<th>Detergent</th>
<th>TVs</th>
<th>Chocolate</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales Rank Change</td>
<td>Mann-Whitney U test applied to control vs. treatment group</td>
<td>U = 85,  ( r = 0.83 ),  p&lt;0.001*</td>
<td>U = 164,  ( r = 0.74 ),  p&lt;0.001*</td>
<td>U = 110,  ( r = 0.81 ),  p&lt;0.001*</td>
<td>U = 170,  ( r = 0.74 ),  p&lt;0.001*</td>
<td>p &lt; 0.05 in all cases; rank change correctly perceived as significantly different from control for all products</td>
</tr>
<tr>
<td>Sales Rank</td>
<td></td>
<td>U = 897,  ( r = 0.10 ),  p = 0.328</td>
<td>U = 905,  ( r = 0.04 ),  p = 0.681</td>
<td>U = 893,  ( r = 0.21 ),  p = 0.044</td>
<td>U = 731.5,  ( r = 0.18 ),  p = 0.100</td>
<td>p &gt; 0.05 except for TVs; sales rank correctly perceived to be the same across most product profile pairs</td>
</tr>
<tr>
<td>Perceived Differences in Product Profiles</td>
<td>One-sample t-test between ( \mu ) in the control groups vs. &quot;4&quot;</td>
<td>( \mu = 1.52 ),  ( \sigma = 1.05 ),  ( t(43) = -15.72 ),  p&lt;0.001*</td>
<td>( \mu = 1.80 ),  ( \sigma = 1.17 ),  ( t(40) = -12.05 ),  p&lt;0.001*</td>
<td>( \mu = 1.55 ),  ( \sigma = 0.97 ),  ( t(46) = -17.23 ),  p&lt;0.001*</td>
<td>( \mu = 1.51 ),  ( \sigma = 1.06 ),  ( t(44) = -15.78 ),  p&lt;0.001*</td>
<td>p &lt; 0.05 in all cases and ( \mu &lt; 2 ) in all cases; thus perceived differences between product profiles were small, across all products</td>
</tr>
<tr>
<td>Perceived Enjoyment</td>
<td>2-Way ANOVA with whether the product was hedonic or not as one factor, and whether its durable as another factor</td>
<td>( \mu = 3.45 ),  ( \sigma = 1.47 )</td>
<td>( \mu = 3.76 ),  ( \sigma = 1.28 )</td>
<td>( \mu = 6.06 ),  ( \sigma = 1.21 )</td>
<td>( \mu = 6.20 ),  ( \sigma = 1.33 )</td>
<td>A product being hedonic significantly influenced perceived enjoyability, ( F(1, 349) = 322.002 ),  p&lt;0.001*. A product being durable significantly influenced purchase frequency ( F(1, 349) = 79.19 ),  p&lt;0.001*.</td>
</tr>
</tbody>
</table>

**Hypothesis Testing: The Sales Velocity Effect**

For each of the four products, the Mann-Whitney U test was applied between the control and treatment groups for the Likelihood of Purchase measure. The results are given in Table 10.
Table 10: Results for the Mann-Whitney U tests and descriptive statistics for the likelihood of purchase. The * indicates significant p-values of the Mann-Whitney U test.

<table>
<thead>
<tr>
<th>Likelihood of Purchase</th>
<th>Mann-Whitney U Test</th>
<th>Descriptives</th>
<th>Control Group vs. Treatment Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U, effect size r, p</td>
<td></td>
<td>Control Group</td>
</tr>
<tr>
<td>Control Group vs. Treatment Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vacuum Cleaners</td>
<td>U = 495.5, r = -0.44, p &lt;0.001*</td>
<td>2.41 (1.207)</td>
<td>n= 44</td>
</tr>
<tr>
<td>Detergent</td>
<td>U = 506.5, r= -0.40, p&lt;0.001*</td>
<td>2.71 (1.553)</td>
<td>n= 41</td>
</tr>
<tr>
<td>TVs</td>
<td>U= 609.0, r = -0.38, p &lt;0.001*</td>
<td>2.60 (1.378)</td>
<td>n= 47</td>
</tr>
<tr>
<td>Chocolate</td>
<td>U= 379.5, r = -0.49, p &lt;0.001*</td>
<td>2.67 (1.297)</td>
<td>n= 45</td>
</tr>
</tbody>
</table>

The Mann-Whitney U tests between the control and the treatment groups were significant for all product groups; thus H1 is further supported and is robust in spite of the differing properties of the products. Furthermore, the means of the treatment groups indicate that the presence of sales velocity leads to a higher likelihood of purchase for the frequently purchased products ($\mu_{\text{detergents}} = 4.13$, $\mu_{\text{chocolate}} = 4.51$) compared to the less frequently bought products ($\mu_{\text{vacuum cleaners}} = 3.93$, $\mu_{\text{TV}} = 3.98$). Likewise, sales velocity seems to exert a higher influence on hedonic products ($\mu_{\text{TV}} = 3.98$, $\mu_{\text{chocolate}} = 4.51$) as compared to the less frequently bought products ($\mu_{\text{vacuum cleaners}} = 3.93$, $\mu_{\text{detergents}} = 4.13$).

4.5.6 Discussion

Study 2 reaffirmed H1 and added generality to the results: the sales velocity is influential even for products that differ in their characteristics. The descriptive statistics suggest that additionally, the effect is stronger for hedonic and frequently purchased goods – indeed, the effect on the likelihood of purchases is the strongest for the originally chosen product in Study 1, chocolate ($\mu = 4.51$), and weakest for its diametric opposite (the non-hedonic, infrequently purchased vacuum cleaner, $\mu = 3.93$).

4.6 Study 3: The Negative Sales Velocity Effect

Study 1 and 2 showed that effect of sales velocity holds even for a variety of products. However, both studies focused on the effect of positive sales velocity on influencing likelihood of purchases. Study 3 extends these results by examining negative sales velocity. The task was exactly the same as Study 2, with the same scenario text, sequence of steps and measures.
4.6.1 Stimuli Development

The profiles followed the same format as Study 1 and Study 2; the control and sales velocity group were as per Study 2. Compared to Study 2, there was additionally a third group added, the negative sales velocity group (group C). Group C was manipulated to have the same magnitude in % sales velocity decrease as the increase in group B. Since Study 1 and Study 2 showed that in the absence of sales velocity (i.e. the control group A), participants are more likely to choose the higher ranked product, to create a conservative design, the decrease in sales velocity was given to the higher ranked product. Thus, if negative velocity would have an effect in influencing sales towards the other (lower) ranked product, it would be in spite of participant preference for the better ranked one. A representation of the product profile as shown to participants is depicted in Figure 17.

![Figure 17: Components in a product description; two generic profiles Q and R are presented.](image)

The presented order of the profiles was counterbalanced between participants. Two focal products from Study 2 were sampled: the vacuum cleaner (low hedonic, high durability) and chocolate (high hedonic, low durability). These were found in Study 2 to respectively have the least and strongest effect from sales velocity in likelihood of purchases. The product descriptions remained the same as Study 2.

4.6.2 Participants

271 American participant were recruited from an online panel (Amazon Mechanical Turk, https://www.mturk.com) (\(\mu_{\text{age}} = 31\) years, \(\sigma_{\text{age}} = 10\), 39% female) for the study.
4.6.3 Measures
The main measure, likelihood of purchase, and the manipulation and attention checks were identical as per Study 1 and Study 2.

4.6.4 Results
Manipulation, Attention and Confounding Checks

For the price attention checking question, a sample of 257 passed (95%) and was considered for further analysis. As per Study 1, Sales Rank Change, Sales Rank and Likelihood of Purchase were not normally distributed as per the Shapiro-Wilk test (p<0.05). Consistent with Study 1 and 2, the Kruskal-Wallis and Mann-Whitney U tests were used with Bonferroni corrections.

For the sales rank manipulation check and for both products, the Kruskal-Wallis test was not significant (for vacuum cleaners, $H(2) = 1.102$, $p > 0.05$, and for chocolate, $H(2) = 5.9$, $p > 0.05$), and all there pairwise Mann-Whitney U tests were not significant ($p>0.0167$ with the Bonferroni correction). The sales rank was perceived (correctly) to have no significant differences across experimental groups.

For the sales rank change manipulation check, there was a significant effect between groups on perceived rank change (vacuum cleaners, $H(2)=65.789$, $p<0.001^*$ and chocolate $H(2) = 60.192$, $p<0.001^*$). Pairwise Mann-Whitney U tests for the sales velocity groups B and C compared to the control were significant (for both products and both comparisons, $p<0.001^*$), thus as intended, the sales velocity groups were perceived different from the control. Furthermore, the difference between groups B and C were not significant (for both products and comparisons, $p>0.0167$), as intended. Thus the sales rank change manipulation was successful. Note that in order to compare the perceived rank change between the positive and negative velocity groups (where the rank change was applied given to opposite products), the negative velocity results were reverse coded.

As per Study 2, for both products, the perceived differences between product profiles was significantly different from the neutral value of “4” (for the vacuum cleaner, $\mu=1.52$, $\sigma=1.05$, $t(43) = -15.72$, $p<0.001^*$, and for the chocolate $\mu=1.51$, $\sigma=1.06$, $t(44)=-15.78$, $p<0.001^*$). Both means are also low (both nearly 1.5). Thus, as intended, the profiles were perceived similar other than their rank change information.
The Negative Sales Velocity Effect

Kruskal-Wallis tests showed there was a significant effect of the sales velocity information on likelihood of purchase (vacuum cleaners H(2) = 29.643, p < 0.001* and chocolate H(2) = 36.52, p < 0.001*). Pairwise Mann-Whitney U tests were applied to determine between which groups the differences were. The results are given in Table 11.

Table 11: Summary of paired Mann-Whitney U statistics for the likelihood of purchase. The * indicates significant p-values of the Mann-Whitney U test

<table>
<thead>
<tr>
<th>Likelihood of Purchase</th>
<th>Groups Involved</th>
<th>Result Vacuum Cleaners</th>
<th>Result Chocolate</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptives ( \mu (\sigma) ) ( n )</td>
<td>Group A</td>
<td>2.41 (1.207) ( n=44 )</td>
<td>2.67 (1.297) ( n=45 )</td>
<td>Sales velocity groups B and C have higher likelihood of purchase than the control. Furthermore, ( p &lt; 0.017^* ) for both “A vs. B” and “A vs. C”, so, H1 and H3a are further supported.</td>
</tr>
<tr>
<td></td>
<td>Group B</td>
<td>3.93 (1.75) ( n=45 )</td>
<td>4.51 (1.805) ( n=39 )</td>
<td>Effect size in “A vs. C” &gt; “A vs. B” so H3b is supported.</td>
</tr>
<tr>
<td></td>
<td>Group C</td>
<td>4.56 (1.933) ( n=40 )</td>
<td>5.02 (1.823) ( n=44 )</td>
<td></td>
</tr>
</tbody>
</table>

| Mann-Whitney U Tests | A vs. B | U=495.000, r=0.44 \( p<0.001^* \) | U=379.500, r=0.49 \( p<0.001^* \) |
| | A vs. C | U=333.000, r=0.54 \( p<0.001^* \) | U=317.000, r=0.59 \( p<0.001^* \) |
| | B vs. C | U=704.500, r=0.19 \( p=0.081 \) | U=703.500, r=0.16 \( p=0.150 \) |

Referring to Table 11, the paired Mann-Whitney U tests show that the likelihood of purchase is increased towards the lower ranking product when it has positive sales velocity (the group A vs. B comparison, \( p < 0.001^* \)), and also when the higher ranking product has a negative sales velocity (the group A vs. C comparison, \( p < 0.001^* \)). Thus H1 is further supported, bi-directionally.

Likewise, H3a is supported. It was found that the effect size in the A vs. C negative velocity comparison is larger (0.54 for vacuum cleaners, and 0.59 for chocolate) than the A vs. B positive velocity comparison (0.44 for vacuum cleaners, and 0.49 for chocolate); this supports H3b – given the same perceived relative sales velocity change (as proven in the manipulation check), the sales velocity effect on likelihood of purchase is larger in the negative sales velocity case. It is also seen from the descriptive statistics that the negative sales velocity has a stronger influence on the likelihood of purchasing product R (for chocolates, \( \mu_{\text{group C}} = 5.02 > \mu_{\text{group B}} = 4.51 \); likewise for vacuum cleaners \( \mu_{\text{group C}} = 4.56 > \mu_{\text{group B}} = 3.93 \)).
4.6.5 Discussion
Study 3 further supported H1 by showing that both negative and positive sales velocity influences the likelihood of purchase. Furthermore, as per H3a, it showed that a decrease in sales velocity led to a lower likelihood of purchase for that product, and as per H3b it showed that this decrease in sales velocity has a stronger positive effect on the likelihood of purchase of the non-decreasing product than a corresponding increase in sales velocity. This likelihood of purchases in the negative velocity case is also higher than in the positive case.

4.7 Discussion

4.7.1 Theoretical Implications
The results show that increasing sales velocity positively influences the likelihood of purchases (Study 1), and that this result is robust across different products (Study 2). The positive signaling from an improving product was even able to reverse the preference seen in the control group for a product whose current rank is high – that is, even in the face of the more common sales rank heuristic. These results reveal and address a large gap in OL research, which previously only considered current outcomes arising from others’ choices, without factoring in the velocity dimension of these outcomes. This sections has also extended the general studies in velocity outcomes from Hsee et al. (1991) into the context of marketing, and furthermore builds on the work by Briggs et al. (2010), which had shown that velocity metrics of performance are positively correlated with a firms’ perception of the providers; the thesis extended these findings by conducting a controlled experiment which proved that there is also an influence on choices, in a context where retailers can use the result for their marketing initiatives.

It was further shown that the numerical framing of the sales velocity can strengthen the sales velocity effect on the likelihood of purchases (Study 1), by changing the perception of the rank change. This extends the work on numerosity (Cheema and Bagchi 2011; Chen and Rao 2007; Kruger and Vargas 2008; Monga and Bagchi 2012; Pandelaere et al. 2011; Zhang and Schwarz 2012), and that furthermore, it could be leveraged on non-price metrics like sales velocity to even boost sales. Finally, study 3 showed that both negative and positive sales velocity influences the likelihood of purchase, and that a product’s decrease in sales velocity has a stronger positive effect on the likelihood of purchase of the non-decreasing product than a corresponding increase in sales velocity. Taken together, these studies give researchers a firm ground for future work in using velocity metrics like sales velocity for marketing purposes.
4.7.2 Implications for Customers

Sales velocity can act as a useful heuristic for consumers in making choices; it has been well established that consumers are often presented with too many choices, suffer from choice overload, and subsequently are less satisfied with their choices (Diehl and Poynor 2010; Scheibehenne et al. 2010). In these situations, consumers are known to resort to heuristics (Oliver 1993) to help them focus their attention on product attributes of importance to them, thus increasing post-choice satisfaction. Since sales velocity presents the less-often bought products, it may be a useful and satisfying heuristic for variety seekers who want to try something different from the usual popular and already bought products.

4.7.3 Implications for Retailers

The results demonstrated the impact of the sales velocity metric, which can be easily deployed for both online and offline physical retailers as an additional marketing tool. For example, currently many E-commerce portals use a list of top selling products; similarly, these retailers can also have a list of high sales velocity products (sorted by rank change) to promote mid tail products.

Furthermore, this chapter found that alternate numerical representations of sales velocity (i.e. as a % change) leads to a stronger sales velocity effect, when the alternative representation leads to a larger number; this representation can be leveraged to boost the perception of rank change. However, one cautionary note should be considered before deciding on a % change representation; additional analysis after the study found that a % change representation exposes different products in the long tail. Whereas the maximum expected rank change increased when the sales rank of the product got worse (and thus is appropriate for promoting mid-tail to long tail products), it was found that except for the #1 product, every product had a chance of exhibiting a large % rank change (in the order of magnitude near 100%) from weekly fluctuations in sales. Thus, which representation is ideal depends on which products the retailer wants to promote.

Retailers can also combine sales velocity with other marketing tools; for example, they could feature targeted products prominently on a website’s front page, like a normal advertisement, but justifying the product’s presence by stating its sales velocity attribute. Retailers can also present sales velocity information in recommendations, to boost its effect. A schematic of how this could be operationalized is presented in Figure 18.
In Figure 18 (a), the sales velocity is used to promote a product in place of a typical price discount, embedded in the customer’s search experience of products, while in Figure 18 (b), sales velocity is paired with a recommendation that targets a product category where the customer normally seeks variety. Optimization of how to display the sales velocity and validation of its effect on consumer sales and the resultant real-world profitability can also be easily implemented with A/B testing by the retailers; retailers can simply have a group who do not see sales velocity in their apps, and different treatment groups are tested with different sales velocity representations.

The results are also particularly relevant to physical retailers; as shown in the analysis of the one year of receipt data from the European physical grocery retailer partner, sales velocity can be used to promote mid-tail products - which form a large portion of the retailer’s revenue. Furthermore, since sales velocity is more suitable to be used for the mid-tail products, it allows retailers to have an alternative and cost-effective promotional vessel other than a series of increasing price cuts for promoting lesser sold goods, which would continually incur an inventory
cost otherwise. Having sales velocity as an alternative to price discounts can thus increase revenue. Having sales velocity as another method for promoting products also gives the retailer flexibility; since many price promotions are negotiated with manufacturers and are planned in advanced (Ailawadi et al. 2009; Moreau et al. 2001; Murray and Heide 1998), the ability to feature a product without resorting to price negotiations helps the retailer. The following sections discuss here some practical recommendations and considerations of the findings.

Sales Velocity Lists can complement Sales Rank Lists

While mid-tail products benefit strongly from sales velocity, products in the head of a long tail tend to be stable and therefore rarely change or increase in rank (and therefore have a low sales velocity). One might therefore ask whether selectively showing sales velocity could be detrimental to the promotion of the popular products.

Since the existing marketing instrument of a “top ten list of products by sales rank” already addresses the products in the head of the long tail, having another instrument that promotes the top ten products by the sales velocity gives the retailer another method to address a previously neglected and different segment of his products. By the very definition of the long tail, since the vast majority of products in fact cannot be the most popular ones (as measured by the current sales rank), it means a much larger pool of products (the majority of products, in fact) can benefit from a sales velocity promotion. An example of how this could be operationalized is presented in Figure 19, which show how the top ten list and the sales velocity lists can be operationalized as separate features, schematically showing the two lists in separate tabs.
Since the two product lists do not overlap in products and since the two lists draw comparisons between products on two different criteria (one based on OL and the other on sales velocity), arguably consumers would not perceive the lists as competing. That is, sales velocity does not replace existing marketing measures, but complements them. It provides an alternative marketing instrument to price discounts (which are also selectively applied to specific products). In the case of the physical grocery retailer partner, mid-tail products consist of 41% of the retailer’s revenue, so this is a managerial relevant segment of goods. Operationalized as a list of high velocity mid-tail products (which by definition, consumers do not often buy), sales velocity can encourage variety seeking. Operationalized as a single promoted product, sales velocity can be the attribute highlighted in place of a price discount.

Positive versus Negative Sales Velocity

One practical matter concerns the positive and negative velocity. Given a period of time, a product can increase in sales rank (positive sales velocity), but also decrease in sales rank (negative sales velocity). Study 3 showed that a product with negative sales velocity would have a low likelihood of purchase. Accordingly, the position of this thesis advocates for implementations of sales velocity (such as a top ten list of products) which avoids showing
negative information. This thesis recommends against showing sales velocity in every product
description, embedded as an attribute since this ensures negative velocity is seen, and
furthermore, this is not done at all by practitioners such as Amazon.com; they have always a
separate screen devoted only to the top products sorted by either current rank or sales velocity.
Accordingly, it is recommended to use sales velocity as a promotional tool to selectively promote
products, in the place of price discounts, thus saving retailer margin.

Sales Velocity per Category

It is also the position of this thesis to recommend that the sales rank and velocity to be
computed within the retailer’s product categories. For an example, if a consumer browses
chocolates, he would get the top ten (or whichever number the retailer chooses) chocolate
products sorted by sales velocity, and if he clicks on drinks, he would get the top ten drink
products only. The sales rank and velocity of the chocolates would be independent of the sales
rank and velocity of drinks. By presenting the velocity as an attribute within a category, one can
avoid the confound that arises from different product categories having different interpurchase
times (Leszczyc and Bass 1998; Leszczyc et al. 2004; Rhee and Bell 2002) (and thus, differing
influences on the velocity). Furthermore, grouping products by category is in line with consumer
mental models of search and choice making (Dellaert and Haebul 2012; Häubl and Trifts 2000);
attributes are compared in the context of similar products. This was in fact how it was
implemented in this study and how Amazon.com does it with their sales velocity feature.

A schematic of how this could be operationalized is presented in Figure 20.
In Figure 20, sales velocity is presented as a category-grouped list of products, sorted by the change in sales rank within that category. By limiting the number of products shown to the top “n” products for each category, the retailer can also avoid showing negative sales velocity information. Indeed, the guidelines for displaying the results of a recommendation system advocate for not more than ten products shown, preferably sorted by a relevant attribute of interest (Pu et al. 2011).

**In-Store Feasibility**

One question might be whether the ideas in this chapter can be implemented in-store, both technologically and also with the buy-in of retailers. Technologically, physical retailers can already operationalize and present sales velocity to consumers via information system artifacts such as mobile phone applications; this has been proven in both research and in practice. In research, in-store recommendation system research (van der Heijden 2006; Kowatsch and Maass 2010; Lee and Benbasat 2010) have already proven the feasibility of such systems; in practice, physical retailers (ex. Walmart) already have apps for their stores where consumers can browse
and obtain information about the retailer’s products. It would be therefore a simple matter to allow consumers to browse and sort products by the sales velocity.

In terms of actual buy-in from retailers, the author of this thesis has worked with a start-up in developing and deploying a smartphone app for the retail partner, available for download that contains a top ten list of products sorted by sales velocity and a separate list that contains the products with a high sales rank. The sales rank change is computed from the retailer’s PoS data, which is acquired through an interface developed in cooperation with the retail partner’s PoS vendor. Therefore, the sales velocity ideas described in this chapter is feasible technically and also agreeable with retailers.

4.8 Overall Conclusion and Future Research

This chapter has proven and validated the effect of sales velocity on influencing the likelihood of a future purchase. The studies conducted showed the asymmetrical effect of negative sales velocity in reducing the likelihood of purchasing a product compared to the effect of positive sales velocity in raising the likelihood of a purchase. The positive effect of sales velocity was validated across diverse products, and it was also shown that in the area of fast moving consumer goods, showing a list of products sorted by sales velocity would reveal and promote mid-tail products. Taken together, these results answer RQ2; sales velocity represents a viable form of social learning marketing that is appropriate for physical grocery retailers.

Although the chapter has shown the robustness of the sales velocity effect across different products and explored its boundaries, future work should test whether the sales velocity effect is diminished in the presence of price and actual monetary loss, similar to the incentive-aligned studies (Miller et al. 2011). This can be achieved by having choices which have real monetary consequences.

This chapter also studied sales velocity’s effect with regard to promoting physical products; however it may also be highly effective for services or intangible goods – for example, for vacation packages and destinations or music downloads which are rising in popularity. This should be investigated in future studies.

During this study, the effect of sales velocity was established, operationalizing it as a change over two time periods. The sales velocity can also be measured over multiple time periods, and of interest to consumers might also be the longevity of a positive rank change; for how long has a product been rising? This however, would expand the scope of the research beyond sales
velocity, as there would then be an implied acceleration or trending component, which was first studied by Hsee et al. (1994) in a cognitive psychology experiment. Since intention of this chapter was to establish the sales velocity effect in a retail context and to evaluate which type of products a retailer could promote with this method, it was beyond scope to deal with acceleration measures. The chapter focused on a simple, easy to implement and understandable metric that addresses the mid-tail of sold products. An extension of the work by Hsee et al. (1994) into the area of sales acceleration would yield further studies and the understanding on the dynamics of sales velocity.

The velocity’s interaction effect with the positional rank is also an area of future work – the impact of sales velocity may depend on which rank it rose from, and the positions of the final ranks of the products compared. In the studies presented, there was always an improving brand (rank 100) having a final rank (rank 51) very close to a superior brand (which stayed at rank 50). Future work should investigate further combinations.

Finally, for practitioners, future work may calibrate the effectiveness of sales velocity in a field deployment. With a field deployment, several practical considerations can be evaluated with respect to their real-world effect on purchases and profitability; for example, the study suggested that the sales velocity heuristic can be a valuable cue, particularly for space-scarce smartphones. Practitioners can thus examine the tradeoffs between having this cue on display versus having less information, in its different visual representations: for an example, would a table of "rapid risers" be overall more effective than a table showing both up and down movements? In addition to these visual calibrations, future work of interest to practitioners would be the interaction of sales velocity with aspects such as the persistency of the velocity (how many "weeks on the chart"), which in of itself could be a metric explicitly revealed to consumers to influence their behavior in select product categories (ex. a rapid riser in a stodgy market might be especially valuable and worthy of note). A real-world deployment would also help identify under what conditions would sales velocity complement vs. cannibalize sales of more popular products. These deployments can be easily implemented both online and offline: for a field experiment deployed with an online retailer, similar to the OL experiment by Tucker & Zhang (2011), sales velocity can be a product attribute exogenously revealed on the retailer’s E-commerce portal, and as such, it becomes possible to see whether sales velocity influences real-world product choices, and for which products is the influence stronger. For a field experiment with a physical retailer, sales velocity can be potentially operationalized via information systems artifacts such as mobile phones, as an extension to the work on mobile recommendation agents
for consumers' in-store choice making (van der Heijden 2006; Kowatsch and Maass 2010; Lee and Benbasat 2010).
5. Towards High Resolution Evaluation of Customer Profitability

In order to answer RQ3- i.e., determining a method by which high resolution marketing can be evaluated - this section presents, estimates and tests a Customer Lifetime Value (CLV) model at the granularity of individuals and their purchasing behavior within categories. Thus, the model is appropriately matched to evaluate the one-to-one marketing insights developed, for example, in the variety seeking chapter.

5.1 Introduction

The ubiquity of smartphones enables personalized marketing for physical grocery retailing. This opportunity requires a high resolution understanding of customers’ space of preferred products and the timing of purchases, and correspondingly, it also requires an evaluative method that helps understand what marketing methods are effective.

An appropriate evaluative metric is the customer lifetime value (CLV) or the lifetime value (LTV) metric which is a widely accepted concept from marketing science. It is defined as the present value of the future cash flows associated with the customer, factoring the cost of retention (Pfeiferis et al. 2004). As a forward facing metric that is computed from - but distinct from - historic profitability, CLV can therefore be used to evaluate whether a marketing action actually improved the lifetime value and profitability of a customer; indeed it is already studied in various information systems for contractual settings (Jonker et al. 2006; Kim et al. 2006). However, there are gaps in the existing paradigm of CLV which inhibit its use as a high resolution, rapid response metric for CRMs in grocery retail. As noted earlier, real-time marketing in grocery retail necessitates both a high resolution understanding of the customer’s space of preferred products and the timing of purchases. Previous studies have shown that customer’s buying preferences exhibit heterogeneity according to product category (Allenby and Rossi 1998; Leszczyc and Bass 1998; Lim et al. 2005), and thus it follows that the lifetime value of customers, and the corresponding effectiveness of a marketing action, will vary by product category. Current CLV models only consider a customer’s total purchases in the firm without evaluation at the product category level, and are often deployed infrequently as a one-time segmentation tool. Taken together, CLV as applied today lacks the necessary product resolution for the aforementioned CRM.

This chapter is thus motivated to address these this gaps by giving researchers and in-store grocery retailers a model for conceptualizing the customer lifetime value at the product category level and to evaluate the application of such a product category level CLV model at higher
temporal resolution. In contrast to other models, the approach in this thesis will show there are indeed differences in lifetime value across categories for the same customer, an important factor for any retailer’s marketing strategy. The approach can be easily implemented by physical retailers with already available in-store point-of-sale (PoS) data and would allow for measuring changes in CLV in almost in real-time. The chapter will also show that the time-variant nature of CLV at the product-category level can be used to evaluate marketing strategies.

The key contributions of this chapter are two-fold. First, it will show that the current conceptualizations of CLV neglect critical properties required for personalized marketing and thus have to be extended for higher spatial resolution. Second, the extended CLV is evaluated with real PoS data of over 638,000 transactions to show the economic impact of such a model. This contribution also complements current CRM strategies, which do not factor in consumer lifetime value at the level of individual product categories for their recommendation strategy.

5.2 Research Framework

5.2.1 Drivers and Value of Category Level CLV

It was noted earlier that current applications of CLV are computed at the firm or store level; that is, the customer’s frequency, recency and average monetary value of purchases irrespective of product category are used to compute CLV.

While the lifetime value of a customer to a retailer is valuable, at the level of individual product categories and for personalized recommendations, a retailer level CLV is insufficient to evaluate the effectiveness of marketing actions at the product category level.

First, in the domain of grocery retail, which is characterized by highly differentiated product categories, there is evidence customers behave differently across categories. According to the attribute view of products, different product categories could be seen as having different attributes which are “consumed” (McAlister and Pessemier 1982); it thus follows that given customer’s individual taste towards these attributes, they would have different purchasing patterns from category to category, and hence lifetime value. Furthermore, since consumers often split their purchases between multiple retailers (Rhee & Bell 2002; Leszczyc et al. 2000), they may have greater lifetime value in certain categories at one retailer while lower lifetime value in others. Additionally, the natural interpurchase time of different product categories are different (Neslin et al. 1985) – perishables like bread are bought more frequently than canned goods, for example, and thus the effect of recency in a CLV calculation would differ by category,
depending on a consumer’s deviation from the “natural” purchase frequency of that category.

Finally, having a customer-store level CLV may obfuscate the actual risk that a customer would remain valuable to a retailer; although a customer may appear overall to run a low risk of defecting to another retailer and overall is spending sufficiently well to be profitable to the retailer, due to the category-differential tastes of a customer, it may be that there are categories where he runs the risk of defecting, even if overall he is not.

To illustrate this, consider the following example. Assume there are two customers, X and Y. Customers X and Y are able to make purchases at discrete times $t_1, t_2, ..., t_n$, and can choose to purchase products from categories A, B and C, which are defined to be mutually exclusive. For ease of illustration, it is further assumed that there is exactly one product offering per category, with a time-invariant price. Therefore, at time $t$ and at the customer-category level, one can define the average monetary value of a purchase in the $i^{th}$ category, $M_i$, to be equal the price of the purchased product $P_i$. The average monetary value of a purchase of a customer at the customer-store level at time $t$ would be $\sum_{i=0}^{k} \frac{P_i}{n}$, for $k$ products purchased.

With these assumptions, two customers, X and Y, are presented in Figure 21. Both have shopped at times $t_1, t_2, t_3$, and within this time range [$t_1 ... t_3$], for each of the product categories A, B and C, both have purchased the same total amount of goods: both purchased one product from category A, one product from category B and two products from category C.

Recall that the CLV models under study in this thesis are functions of Recency, Frequency and Monetary Value (RFM). That is, the more recent, more frequent, and greater the amount a customer has bought at a store, the less likely he will defect and the greater the lifetime value. In this example, at the traditional customer-store level, for both customers X and Y, their RFMs are identical and therefore their corresponding customer lifetime values would also be the same. Indeed, because (1) they have made the same number of purchases, (2) their total and average spending are the same and (3) their most recent purchase occurred equally recently, from a retailer and mathematical point of view, the two have the same level of risk of defecting from the retailer and both provide the same amount of value.

However, the picture changes when one looks into the category level patterns and RFM values; although overall both customers have the same CLV and risk of defection overall from the store, Customer X notably has a much worse recency time for Category A and B compared to Customer Y. Thus, after estimating a CLV model for each product category, it may emerge that Customer X...
has a real risk of defecting in categories A and B, in spite of his overall apparent low risk of
defecting from the store overall at $t_3$ – an insight that would not have been apparent with the
traditional customer-store view of CLV. Furthermore, as this example suggests, the current
customer-store view of CLV may also be overly optimistic about the lifetime value of a customer
– taking Customer X as an example, if in fact it emerges from a category-level CLV computation
of CLV that the large time gap ($t_3 - t_1$) in Category A and B truly indicated a defection in those
categories, then he is only one category away (Category C) from defecting outright from the
retailer.

Figure 21: Two customers who have the same RFM at the store-level, and therefore, same CLV, but different category
level RFM, and therefore, different CLV at the category level

Taken together, it can be seen that a CLV model that is at the grocery store level thus ignores the
differences between product categories and consumers’ differentiated interest in them. Without
acknowledging these differences in CLV between categories, targeted promotions run the risk of
losing their maximal value to the retailer, and retailers run the risk of overlooking categories
where the lifetime value of a customer is low and hence could be potentially developed. This
study will empirically investigate these proposed drivers for differences in category level CLV, leading to the first set of research questions:

**RQ3.1a:** What is the extent of heterogeneity in category-level CLV for consumers?

**RQ3.1b:** To what extent do interpurchase times and individual level spending in product categories affect heterogeneity in category-level CLV?

Furthermore, as an extension of the problem presented in Figure 21, it is not known what is the relationship between the traditional person-level CLV, computed at the customer level using all purchases, versus the category level CLV. That is, is the CLV computed at the customer level simply the sum of category level CLV? If the “over-optimism” of the overall CLV as suggested by the example of Figure 21 turns out to be true, then would that mean that the CLV computed at the customer level would be larger than the sum of the category level CLV? To investigate this aspect, the following research question is proposed:

**RQ3.1c:** What is the relationship between a person-level CLV and their category-level CLVs?

In contrast to a person-level CLV today, by determining this extent of heterogeneity within customers, one could potentially identify product categories for individual customers where a promotion could help revive them, and categories where their CLV is high and thus no further promotions are necessary. One could also potentially synchronize CLV and marketing efforts with the different natural interpurchase times between categories. This study will evaluate these possibilities, thus leading to the next research question:

**RQ3.2:** What is the value of category-level CLV for managing customer profitability?

### 5.2.2 The Timely Application of Category Level CLV for Marketing Evaluation

With smartphones it becomes increasingly important to enable real-time marketing, and as discussed earlier, CLV could provide one such measure of the success of marketing actions. On the matter of time, CLV measures have been typically used for a one-time segmentation of customers (Gupta et al. 2006; Kim et al. 2006; Malthouse and Blattberg 2005; Verhoef et al. 2002); since customers are acknowledged to have different degrees of lifetime value, and that a small minority of consumers drive most of the profits of a firm (Gupta and Zeithaml 2006), the course of action recommended in the literature is to compute the CLV of all customers once and decide on whom to continue with promotions accordingly. However, this misses one big
potential of CLV as an evaluative metric: since CLV is computed on past purchases, it can always be updated as new purchasing events are observed, and hence beside segmentation, it can also be used to evaluate the impact of individualized recommendations and promotions. This means that the CLV for a customer can be immediately re-calculated after a purchase and thereby obtain clear feedback on the success of a promotion (CLV goes up). Indeed, the empirical evaluation of promotions on CLV was identified by Blattberg, Malthouse, & Neslin (2009) as being an open research issue, and the timely application of CLV at a person level was first proposed by Wang & Hong (2006), who proposed strategies to address customers depending on the temporal dynamics of their CLV - i.e. whether their CLV was rising/falling, consistent across time, or high vs. low – however this was evaluated at the person-level. This idea will be complemented with an extension to the category level, which leads to the next research question:

**RQ3.3:** How can marketing effectiveness be clarified ex-post with a high-resolution, category-level CLV model?

### 5.3 Method and Results

#### 5.3.1 A Model for Category Level CLV

A CLV model at the product category level would thus enable quantifying and characterizing the product category differences in lifetime value within and between customers. Conceptually, this is illustrated in Figure 22:

![Figure 22: Concept of the category level CLV](image)

This model is based on the Fader model discussed in the related work. The Fader model was chosen as it has been shown to be empirically valid across domains, and furthermore, it is
computable with readily available data: only RFM is needed, which in turn can be easily
computed from the retailer’s point-of-sale data. Note that one can interchange the Fader model
with another since the underlying concept of CLV at a product category level does not depend
on a particular model of CLV; the requirements are that (1) the CLV is computationally tractable
based on available point of sale (PoS) data, and thus, automatable for an information system and
(2) the model’s input reflects counting phenomenon.

The model is as follows. For a given product category $i$, the CLV of a customer is given by:

$$CLV_i = \text{margin}_i \times \frac{\text{Expected average revenue}}{\text{transaction}_i} \times \text{Discounted Estimated Residual Transactions}_i$$

(Equation 7)

Where:

- $\text{margin}_i$ is the expected gross profit margin that arises from each transaction in a given
  product category
- $\text{Expected average revenue per transaction}_i (M_i)$ is the Bayesian expected future average
  revenue per transaction, described by a submodel
- $\text{Discounted Estimated Residual Transactions}_i (DERT_i)$ is the future projected number of
  transactions, based on a submodel of consumer buying behavior until “death”

The difference from the Fader model is that the model presented here evaluates CLV at the
product category level, and thus allow the margin, average transaction value and residual
transactions to vary per category. Since the margin for different product categories are not the
same, a CLV computed at a person level would introduce noise when they assume the same
margin applies overall. This customer level CLV model will be referred to as the CLV-Category
model.

As per Fader et al. (2005), the evaluation of the Discounted Estimated Residual Transactions is
conducted by using the Pareto/NBD model as follows, however, at the category level $i$:

$$DERT(\delta_i | r_i, \alpha_i, \beta_i, X=x_i, t_{x_i}, T_i) = \frac{\alpha_i^r \delta_i^{\alpha_i-1} \Gamma(r_i+x_i+1) \Psi[x_i,t_{x_i},\delta_i | r_i+T_i]}{\Gamma(r_i)(\alpha_i+T_i)^{r_i+x_i+1}L(r_i,\alpha_i,\beta_i | X=x_i,t_{x_i},t_{x_i},T_i)}$$

(Equation 8)

Where $\delta_i$ is the compounded interest rate and $r_i$, $\alpha_i$, $s_i$, $\beta_i$ are the Pareto/NBD parameters to be
estimated. These parameters are allowed to vary at the category level – heterogeneity in
transaction rates across customers follows a gamma distribution with shape parameter $r_i$ and
scale parameter $\alpha_i$, and heterogeneity in dropout rates across customers follows a gamma
distribution with shape parameter $s_i$ and scale parameter $\beta_i$. $\Psi(\bullet)$ is the confluent hypergeometric function of the second kind and $L(\bullet)$ is the Pareto/NBD likelihood function. The inputs to $DERT_i$ are a transaction stream $(X_i = x, t_{x,i}, T_i)$, where $x_i$ is the number of transactions observed in the time interval $(0, T_i]$ (the frequency) and $t_{x,i}$ ($0 \leq t_{x,i} \leq T_i$) is the time of the last transaction (the recency) for a given product category $i$. Thus having frequency and recency within a product category are sufficient in computing $DERT$.

For computing the $M_i$ for each individual at the category level $i$, the following model from Fader et al. (2005) is applied. The $M_i$ model assumes the average transaction values vary across customers but do not vary over time for any given individual, and the dollar value of a customer’s transaction varies randomly around his average transaction value $m_{x,i}$. Thus:

$$E(M_i | p_i, q_i, \gamma_i, m_{x,i}, x_i) = \left( \frac{q_i-1}{q_i+q_i-1} \right) \frac{y_i p_i}{q_i-1} + \left( \frac{p_i x_i}{p_i x_i + q_i-1} \right) m_x \quad (Equation \ 9)$$

Where $p_i, q_i$ and $\gamma_i$ are shape and scale factors of the model. The logic of having the submodel for $M_i$ is to factor in the biasing effect of a small number of observations on $m_{x,i}$; that is, for small number of observations, the weighing of $m_{x,i}$ towards $M_i$ is diminished and the population mean is instead given more weight. Conceptually, the combined model of Equations (7), (8) and (9) assumes customers go through two stages in their “lifetime” with a product category of a firm; they are active for some period of time and then permanently become inactive (Equation 8). A customers can purchase whenever he wants and the number of transactions in a given interval of time varies randomly around his average rate (Equation 8). The point at which a customer becomes inactive is unknown and random to the retailer, so this is estimated by the Pareto/NBD model captured in DERT (Equation 8). The inclination to dropout is heterogeneous and varies per person depending on the category, hence Equations (7),(8) and (9) are evaluated at an individual and category level.

With this category-level CLV model at hand, it is possible to confirm comprehensively to what extent the CLV differs between and within customers for different categories, and also evaluate the consequences on profitability, thus addressing the research questions.

5.3.2 Dataset for Estimation

To estimate the category level model and address the research questions, a year of point of sale (PoS) data from a European physical grocery retailer partner was examined. The receipt data comes from a complete set of all transactions from a single store, and consists of over 150,000 unique receipts covering a total of two million transactions, with a total of 19,374 unique
products sold that year. A transaction in this context refers to the purchase of an item as it appears on the receipt; each transaction event records the name of the item, a timestamp of when it was bought, the European article number (EAN), a receipt ID, the number of units bought, the price per unit, the loyalty card ID of the household who bought it and the product’s category as defined by the retailer. Although the customer’s age distribution and gender are not known, it is the full dataset of one particular store and was deemed “typical” by the retail partner. Therefore, from this data it is possible to construct the purchasing histories of the households, and identify how many items, unique or in total, they bought per category and at what time.

Since RQ3.2 intends to estimate the “lost” CLV potential at the category level that arises from individual customers purchasing “infrequently” for some categories, for each category it is therefore necessary to determine what is a “normal” frequency (from the point of view of the customer) of purchase given a natural shopping cycle (i.e. weekly). Thus the models are estimated on consumers who have shopped more than 51 time at the focus store (N=848). This is based on the observation that across all stores, consumers shop weekly (Leszczyc et al. 2000; Rhee and Bell 2002). Furthermore, the focus of the analysis on products which were available all year around (defined as having been sold at least once per week in the entire store) in order to determine a conservative (since non-year round shoppers were excluded from the potential profits to be made) but stable potential in recovering “lost” CLV. Thus, the analysis looks at 638,000 transaction events. The unit of analysis for time in this study is weekly; thus all purchase events in a given product category are summed up on a weekly basis, and each weekly sum are denoted as a “transaction”.

Note that such a customer sampling would likely have an overall CLV higher than the general population, which would have included those who do not shop weekly or more at the focal store. However as it will be seen in the data, a high purchase frequency overall at that one store does not translate to a high CLV uniformly across all product categories, and thus the heterogeneity examined by RQ3.1 and the motivation for obtaining an interpretable result for RQ3.2 is still preserved. Nonetheless, the minimum purchase and product availability filtering are in line with past models estimated on consumer scanner panel data (Bawa 1990; Bucklin et al. 1998; Guadagni and Little 1983; Gupta 1991).
5.3.3 Product Categories Definition

The product categories of the retailer are defined at four hierarchical levels. At the highest (first) level, there are ten categories: “meat and sausages”, “fruits and vegetables”, “fresh goods”, “ingredients”, “preserved food”, “drinks”, “baked goods and sweets”, “washing material”, “perfume”, and “non-food items”. These are then broken down into subcategories at the second level – for example, “drinks” would be broken down into the subcategories of beer, coffee, tea, etc., which in turn can be broken down into a third level category (“bottled beer”, “canned beer”) to finally the fourth and lowest level (different variants of beer such as “draught beer”, “wheat beer”, etc.). The study in Chapter 3, which used this same data set, found that 3rd level product category had the best overall granularity; the 4th level category is too fine and overfits a category 1:1 to a single product, while the 2nd level is too coarse. Therefore, as per Chapter 3, the 3rd level category was chosen.

Although each retailer could have their own arbitrary classification scheme, note that these categories are mappable to reference categories defined by the widely adopted and open Global Product Classification (GPC) international standard. As such, the results presented are applicable and can be “translated” and compared across all grocery retailers, as long as the products in a given grocery retailer are mapped back to the standard GPC reference categories.

5.3.4 Model Validation

Given the transaction events, it is possible to compute every customer’s RFM. Given the RFM, the parameters of Equations (7), (8) and (9) can be estimated, and hence CLV can be computed. As per Fader et al. (2005), since Equation 7 depends on Equations 8 and 9, first the transaction model (Equation 8) and the average transaction value model (Equation 9) are validated using a holdout period (13 weeks – the first quarter of a year) for model calibration and the remaining number of weeks (43 weeks – the remaining three quarters) for model validation. For each product category, the maximum likelihood estimates of the model parameters are calculated, which are then used to compute the expected number of transactions (Equation 8) and the average transaction (Equation 9) from the population for each of the 52 weeks. An example for the category milk is shown in Figure 23.

---

7 Defined in http://www.gs1.org/gdsn/gpc/what
It can be seen from Figure 23a that the transaction model (Equation 8) tracks the cumulative number of transactions accurately in both the calibration and holdout periods. It can also be seen from Figure 23b that the distribution of average spending per transaction, as modeled by Equation 9, closely follows the actual distribution of spending observed during the calibration period – for example, in both the model and the actual data, on average, most people spend 1.5€ per transaction on milk. This assures that the model for expected number of transactions and spending is appropriate for the category of milk; for each of the remaining 127 product categories, this check was also repeated and a similarly good model fit was found.

5.3.5 Drivers and Value of Category Level CLV

Having validated the appropriateness of the Fader model with the data set, to address RQ3.1a it is necessary to characterize the differences in CLV across categories within individuals, and show the mechanisms why this is the case (RQ3.1b).

As a first step the category level CLV were computed for each individual using the first quarter of the data. Although the model allows for a different margin per product category, for simplicity of the CLV computation a constant margin of 5% per transaction is assumed, which is conservative and in line with the grocery retailer partner. In order to make CLV comparisons across categories for a given individual, a cumulative distribution function was estimated for each of the product categories’ CLV values, and used this function to generate an individual “CLV index” from 0 to 1 – i.e. to what extent a person in a given product category has a higher CLV than all others. The empirical cumulative distribution function (ecdf) is $F_n(t)$, a step function that jumps $j/n$ at observation values, where $j$ is the number of tied observations at that value, and $n$ is the number
of observations. Missing values are ignored. Thus for observations $X = (x_1, x_2, x_3, x_4)$, $F_n$ is the function of observations less than or equal to $t$, given by:

$$F_n(t) = \frac{\text{number of elements in the sample} \leq t}{n} = \frac{1}{n} \sum_{j=1}^{n} 1\{x_j \leq t\} \quad (\text{Equation } 10)$$

Where $1\{A\}$ is the indicator function of event $A$. The useful property of the CLV index is that it comparable between categories because it is a dimensionless indicator of population-relative lifetime value, in a mathematically correct way.

Then, in order to characterize the CLV heterogeneity within a user, the Gini coefficient was computed for the category level CLV indices. The Gini is a common measure of distributional inequality; it has been applied to many problems, the most well-known being income inequality (Sen 1976). Drawing a parallel to the CLV case, the coefficient quantifies to what extent a person is consistent in their category level CLV indices. A low Gini coefficient indicates that a person is either consistently having high CLV or low CLV. It is thus possible to segment and characterize customers by their overall CLV and their Gini coefficient, to identify those who are consistently valuable, or for which categories are a given consumer valuable.

In order to define the Gini coefficient $G$, let $L(u)$ be the Lorenz curve denoting the percentage of the total CLV generated by the lowest $u\%$ of product categories during a fixed time period. The Gini coefficient is defined as:

$$G := \frac{A}{A+B}, \quad A = \int_{0}^{1} (u - L(u)) du, \quad B = \frac{1}{2} - A \quad \text{(Equation } 11)$$

Thus $G \in [0,1]$ and a value of $G = 0$ reflects diversity (all CLV indices are equal) whereas values near one represent concentration (a small number of product categories have most of the CLV). A distribution of the resultant Gini coefficients is depicted in Figure 24: which shows that most consumers do not have equal CLV indices across categories: only a minority has a Gini coefficient lower than 0.2, indicating low inequality. RQ3.1a is thus addressed: from Figure 24’s overall distribution of heterogeneity, the majority of customers ($N=813$, or 96%) have a Gini greater than 0.2, with the median at 0.31, indicating customers indeed have different CLVs across categories. With this method, it is possible to identify at the category level which consumers are valuable, and at a person-aggregated level, how consistent (via the Gini) that customer is across categories.
To investigate the mechanisms that lead to the different CLVs, this study examines the three aspects which are ignored by traditional CLV at the person level: that (1) the margins can be different across all categories (2) the interpurchase times, which affect the natural recency of each category, and hence CLV, are different across categories and (3) the average value of a transaction, which is used in the CLV calculations, are different by category. Since (1), (2) and (3) are inputs into the CLV computation, when combined with a consumer’s differentiated interest (and hence number of purchases) between categories, there would therefore be differentiated CLV. In the computations, since the same margin was used across categories this would not contribute to the CLV difference, however it is possible to examine (2) and (3) with summary statistics, given in Table 12, which show the five product categories which have the lowest variation (low standard deviation) in interpurchase times.

**Table 12: Interpurchase times for five selected product categories**

<table>
<thead>
<tr>
<th>Five Products with the Lowest Variation in Interpurchase Time</th>
<th>Milk</th>
<th>Non-Leaf Vegetables</th>
<th>Take-Away Buffet Items</th>
<th>Root Vegetables</th>
<th>Yoghurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Interpurchase time (days)</td>
<td>9.64</td>
<td>10.48</td>
<td>9.87</td>
<td>15.82</td>
<td>14.06</td>
</tr>
<tr>
<td>Std. Dev Interpurchase time (days)</td>
<td>15.01</td>
<td>15.54</td>
<td>16.14</td>
<td>20.89</td>
<td>21.47</td>
</tr>
<tr>
<td>Mean Transaction Value (€)</td>
<td>1.48€</td>
<td>2.73€</td>
<td>5.64€</td>
<td>2.09€</td>
<td>1.90€</td>
</tr>
<tr>
<td>Std. Dev Transaction Value (€)</td>
<td>0.86€</td>
<td>1.27€</td>
<td>2.93€</td>
<td>0.66€</td>
<td>1.00€</td>
</tr>
</tbody>
</table>
It can be seen from Table 12 that even for these most stable categories, the average interpurchase time of milk is different than root vegetables, and likewise the average transaction values differ. For the entire set of product categories, the Gini coefficient for the average interpurchase times is 0.25 and for average transaction value 0.33, indicating non-uniformity.

RQ3.1b is thus addressed: These results show that computing the CLV from all purchases at a person level ignore the heterogeneity of product category level interpurchase times and average transaction values. The heterogeneous timing of purchases for products also give evidence that a customer-store level CLV model obfuscates the category level differences, which may mask possible low CLV in specific categories for a specific person – a person may overall be profitable to the retailer, but there may be categories for which there is the danger of defecting. With this idea in mind, to probe RQ3.1c, the relationship between category-level and overall CLV, Table 13 is presented, which shows the CLV scores computed by summing up the category level CLVs (CLV_{summed}) and from the traditional method at the person level, using all purchases from all categories (CLV_{person-level}). This was computed for three selected customers, as well as for the population average. It can be seen that the sum of all category level CLVs for a given person (CLV_{summed}) is not the same as the case where all of the purchases at the store across all categories are lumped together for a person-level CLV calculation (CLV_{person-level}).

Table 13: CLV and CLV indices computed by aggregation on vs. direct purchases

<table>
<thead>
<tr>
<th></th>
<th>Sum of all category CLVs for that consumer (CLV_{summed})</th>
<th>CLV computed from all purchases (CLV_{person-level})</th>
<th>CLV Index computed from CLV_{summed}</th>
<th>CLV Index computed from CLV_{person-level}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer A</td>
<td>1680.03€</td>
<td>1786.83€</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>Customer B</td>
<td>468.30€</td>
<td>658.67€</td>
<td>0.51</td>
<td>0.51</td>
</tr>
<tr>
<td>Customer C</td>
<td>303.05€</td>
<td>467.35€</td>
<td>0.19</td>
<td>0.24</td>
</tr>
<tr>
<td>μ all customers (n=848)</td>
<td>516.26€</td>
<td>714.90€</td>
<td>0.50</td>
<td>0.50</td>
</tr>
</tbody>
</table>

To quantify the difference in person-level CLV for the whole population, the root-mean-square deviation (RMSD) are computed for CLV_{person-level} vs. CLV_{summed}, which is given by average deviation between these two models across all 848 of the customers as:

\[
RMSD (CLV) = \sqrt{\frac{\sum_{j=1}^{n}(CLV_{summed} - CLV_{person-level})^2}{n}} \quad (Equation 12)
\]
Where CLV\textsubscript{summed} is the sum of all category CLVs for the jth customer, CLV\textsubscript{person-level} is the CLV computed at a person-level using all purchases at once, and n = 848 customers. The RMSD was 222.21€, which meant that on average the CLV estimates differ by 222.21€ - and from the full data set of customers, the CLV\textsubscript{person-level} was larger than CLV\textsubscript{summed} for 838 out of 848 customers. Thus RQ3.3c is addressed: the traditional CLV\textsubscript{person-level} overestimates by 222.21€, which is 43% higher than that of the average CLV\textsubscript{summed} (516.26€).

One could view this overestimation as “noise” from the “true” CLV that arises from not taking into account that a person may have category level differences in CLV and that in some categories he may be much less valuable at than others. To examine this and to examine RQ3.2 - the implications of this heterogeneity - three customers are picked out and elaborated on in terms of their differences in CLV values across four key product categories. These product categories have the largest total CLV – a managerially relevant segment of product categories. This is also compared with the traditional CLV at the person level, which used all purchases from all categories (CLV\textsubscript{person-level}). The results are shown in Table 14:

Table 14: Three examples of customers with their CLV indices computed from their purchases in four categories, as well as their CLV indices computed from all 128 categories

<table>
<thead>
<tr>
<th>Customer</th>
<th>CLV\textsubscript{person-level}</th>
<th>Gini of all CLV Indices</th>
<th>Cheese</th>
<th>Take-away Buffet Items</th>
<th>Non-leaf Vegetables</th>
<th>Beer</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.99</td>
<td>0.21</td>
<td>0.97</td>
<td>No Purchase</td>
<td>0.98</td>
<td>0.85</td>
</tr>
<tr>
<td>B</td>
<td>0.51</td>
<td>0.56</td>
<td>0.06</td>
<td>No Purchase</td>
<td>No Purchase</td>
<td>0.98</td>
</tr>
<tr>
<td>C</td>
<td>0.24</td>
<td>0.14</td>
<td>No Purchase</td>
<td>0.73</td>
<td>No Purchase</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Table 14 clearly shows the category differences in CLV within users: with the model presented in this chapter, it is now possible to identify product categories for individual customers which a promotion could help revive them and categories where CLV is high and thus no promotions are needed. An example of the latter is with Customer A, who has a large CLV based on his total purchases across all categories. The Gini is small and one sees that across all product categories, his CLV indices are also large. Thus, Customer A is a consistently high CLV customer – as per the typology from Wang and Hong (2006), such a customer has a low chance of defecting and so marketing efforts should not be in the area of retention, but rather in getting the customer to try new products or to further upsell. The categories with a high Gini would tell us relative to other categories, which ones have potential. Customer B has an overall mid-range CLV with a large Gini; it suggests that most of his CLV comes from high volume buying in certain categories, and indeed this is the case for beer, where he is top, whereas for cheese he is near the bottom. With
only the information from Wang and Hong (2006), a manager would conclude either to upgrade the customer to a higher overall CLV; with the Gini, the retailer can go deeper in determining “where” to apply targeted marketing and might choose to target beer because he is valuable there, or choose to “revived” him for cheese. Finally for Customer C, one sees that his CLV index based on all categories is quite low, and that the Gini is low – this means he is an overall low value customer, consistent across categories. Interestingly, he is a good customer in take-away items and beer - suggesting that his purchasing behavior is quite low in the categories not shown.

Taken together, the Gini and the person-level CLV index can thus be used together by the firm’s information system to determine and deliver product-level marketing strategies – a typology of a possible development strategy is shown in Figure 25, which will be referred to as the CLV-Consistency framework. The naming convention of Wang and Hong (2006) was adopted for labeling those segments with low, medium and high CLV as “less profitable”, “profitable” and “most profitable” respectively, while augmenting them with the dimension of “consistency” (“inconsistent”, “mixed-consistent” and “highly consistent” for high, medium and low Gini respectively) – thus Figure 25 is a combination of the CLV “amount” dimension from Wang and Hong (2006) with the product category consistency dimension, represented by the Gini, as developed in this chapter:

<table>
<thead>
<tr>
<th>Overall CLV Index</th>
<th>Gini Coefficient of CLV Indices</th>
<th>Marketing Strategy if customer has...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Low</td>
<td>Consistently Less Profitable Customer: Consistently low value across categories</td>
</tr>
<tr>
<td>Medium</td>
<td>Mixed-Consistent, Profitable customer: Maintain &amp; preserve, guard against down-trends in CLV</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Inconsistent, Less Profitable Customer: Has some high CLV categories, but is overall not valuable. Identify select CLV categories for development.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 25: Typology of marketing strategies given a person’s Gini coefficient of CLV indices and overall CLV index
Figure 25 illustrates a map of how a customer can be developed depending on where he stands in the Gini-CLV typology. For example, if a customer has overall low CLV but a high Gini, this suggests he could be developed towards a medium value customer by selecting key product categories and upgrading the customer towards a high CLV in those categories. Those who exhibit medium to high overall CLV and medium to high Gini is where this model can show the greatest improvement in insight compared to the Wang and Hong (2006) framework, since it becomes possible to identify exactly “where” and which categories to develop for winning back or retaining potentially defecting customers (i.e. categories where the CLV is low), and it also identifies categories where CLV is already high and therefore a candidate category for cross and up-selling.

In order to explore how many customers of the 848 customers fall in each of these typologies and the implications on profit, the following thresholds are defined: for the Gini, those greater than or equal to 0.5 are defined as high, those less than or equal to 0.25 as low, and those in between as medium. For CLV, the top quartile of CLV\textsubscript{summed} are defined as high, the lower quartile as low, and all in between as medium. Additionally, for each customer, the number of product categories he has made purchases are counted and the average CLV\textsubscript{summed} are computed within each segment.

<table>
<thead>
<tr>
<th># Customers, (μ\textsubscript{avg. # categories visited/customer}), μ CLV/customer</th>
<th>Gini</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>CLV\textsubscript{summed}</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>N= 34 (6.7) 158.45€</td>
</tr>
<tr>
<td>Medium</td>
<td>N=33 (29.73) 509.99€</td>
</tr>
<tr>
<td>High</td>
<td>N=62 (53.92) 1081.37€</td>
</tr>
</tbody>
</table>

The results show that those who have high CLV tend to have a higher number of categories visited and are consistent about it (i.e. the high CLV low Gini group). The interesting groups here are the medium value / maturing customer groups with a medium Gini (the light shaded cells): at the given CLV level, it can be seen that they visit nearly the same number of categories as their low Gini counterparts (the dark shaded cells); but since they have a higher Gini, it means within these same number of categories shopped, they also have more low CLV categories and thus a greater potential for further improvements in category level CLV. Corresponding to this, it can
also be seen that within the medium and high CLV levels, those with a lower Gini exhibit a higher CLV than those with a medium Gini: in the medium CLV segments, its 509.99€ vs. 466.86€ (a difference of 43.13€, and with N=386, a potential of 16648.18€ if all of these customers were developed to be more consistently valuable across categories) and in the high CLV segments 1081.37€ vs. 854.83€ (a difference of 226.54€, and with N=151, a potential of 34207.54€ if these customers were developed). Thus, the high resolution CLV allows identifying those customers who exhibit heterogeneous value between categories, the categories for development, and helps quantify the potential improvement in CLV.

This amount-consistency framework can naturally be extended and fully incorporated with the Wang and Hong (2006) framework to also include the trend dimension and the volatility dimension; this temporal dimension is discussed in the following section.

5.3.6 The Timely Application of Category Level CLV for Marketing Evaluation

In the Wang and Hong (2006) framework, where the CLV dimensions of trends and volatility are included, they recommend addressing customers where the CLV trend is decreasing, because the cost of retaining a customer is typically less than that to recruit one from scratch. This section investigates how the repeated application of the category level CLV may be used to determine the effectiveness of such a tactic of addressing these customers; if today’s mass-promotion are thought of as a form of investment, then the investment can be recovered if purchase frequency or total spending is stimulated. But, there will be those who buy only on discount, potentially decreasing the CLV, in which case, this investment might have been better invested in a more profitable category for that customer – this however, would not be apparent in an overall CLV model that does not look at product category level CLV.

Thus, to answer RQ3.3 and address how category-level CLV could be used for high-resolution marketing evaluation, the following were computed: (1) the category CLV, (2) the number of purchases made in total and (3) the number of purchases made on discount at three time intervals (using all the data up to and including the first quarter, the second quarter and finally the third quarter). As an example, focus is given specifically to those whose new purchases in Q3 arise solely from a discount; Figure 26 shows some of these results for three chosen product
categories, which were found to have the highest percentage of discount transactions having “saved” a customer from inactivity.

![Diagram](image)

**Figure 26: Customers’ reactions to promotions and the effect on CLV**

From Figure 26, there is a clearly differentiated effect of discounts on being able to increase CLV: for pickled vegetables and cream cheese, the discount actually lowered CLV for the majority of those whose new purchases in Q3 were driven solely by the discount. Dishwasher fluid however exhibited a majority of customers having an increase in CLV. The results suggest that for certain customer-category pairs, the discount was a poor investment in maintaining their profitability.

To investigate this further, for the cream cheese category, the trajectories of several customers’ purchase frequencies and average spending in Q1, Q2 and Q3 were examined. Referring to Figure 27, for Customer 103914, one sees that he made no transactions in Q1. In Q2 he receives a discount (a), which prompts his first purchase in that category. His CLV becomes 6.70€. In Q3 he continues buying, but it can be seen that it is entirely due to it being another discount, and his average spending drops (b) as with his CLV (c). Here, one sees a consumer who was recruited by a discount, but ultimately, loyal only as long as it remained. For Customer 109027, there are no new purchase activities between Q1 and Q2. Reflecting this increasing time gap (i.e. recency downgrade), his CLV drops from 5.19€ in Q1 to 3.22€. However in Q3 he is stimulated with a new discount and his average spending rises from 1.09€ to 2.39€; his CLV becomes 4.45€. The
A discount “revived” a possibly defecting customer and even increased the CLV. Finally, for Customer 111797, one sees that he continuously is buying on discount and his CLV and average spending is decreasing with every quarter. The discount is eroding his average spending and his CLV.

Figure 27: Individual level reaction to promotions (a) and the effect on (b) average spending per transaction and (c) CLV

From this one sees how a category level CLV, applied periodically, can be a tool for diagnosing marketing actions. It is possible to distinguish who was “saved” via a promotion, and more importantly, whether it was profitable to save them. It can be seen that repeated discounts can sometimes increase the CLV of a customer, by reviving him after a period of inactivity (customer 109027); however it can also lead to erosion of profits (customer 103914 and 111797). Since it was seen in RQ3.1 that the overall CLV can be very different than the category level CLV, without this granularity it would not be able to determine the effectiveness of promotions at the category level. The result already shows that a quarterly interval for time resolution can already reveal the effectiveness of promotions, although a real-world implementation can of course use a finer cycle. Furthermore, it emerges from this result that current approaches of giving everyone the same discount (i.e. via a weekly promotion) leads to “wastage” of CLV, since for some the discount will lower their lifetime value, and only for certain people will it be effective.

To investigate the cumulative effect of “saving” consumers as well as the “wastage”, across all product categories, three groups from Figure 26 were examined in closer detail: (1) those who did not buy anymore between Q2 and Q3, indicating either “death” in that category (recall that the sample consist of those who shop all year) or a missed opportunity (or a failed promotion) to “revive”, (2) those who were “fully saved” and whose CLV went up, indicating an effective discount, and (3) the “fully saved” whose CLV went down, indicating an ineffective discount. The first scenario had 42240 instances where a customer in a given product category lost CLV from Q2 and Q3 when they showed no further buying in that time frame; the total CLV lost was 11866.38€. The second scenario had 572 cases and the total CLV boosted was 1533.12€. The
final scenario had 606 cases, and the total decrease in CLV was 1125.12€. Notably, among those who bought on discount, the net CLV is positive (1533.12€ - 1125.12€ = 408€). Among those who did not buy between Q2 and Q3, the loss is an order of magnitude higher than the gain of a discount. Since all consumers got the same promotions, it implies these consumers did not buy because of lack of interest or awareness in the discount or product category, thus implying an even greater need for targeted marketing.

5.4 Discussion

The results showed that within consumers, interpurchase times, average amount spent and consequently, CLV, differed significantly by product category. The contributions of this chapter are four-fold. First, the higher granular approach allows a detailed product-level evaluation of a customer’s CLV, avoiding an overly optimistic valuation that arises from a person-level CLV. Conceptually, when the CLV is estimated at a person level over all purchases at once, there is noise due to the heterogeneity between categories – and the collective number of purchases and store visits seen from a person-store level lacks the granularity to see whether a customer is at risk of defecting in an individual product category. It was found that the person-level CLV overestimation compared to the new method to be 43% on average. It may even be able to further improve on this result by having CLV computed at a product-level, thus accounting for heterogeneity between products within a category. Secondly, if someone has a high CLV in a few categories but not others, for this person to have a high CLV, relative to the population he must shop at a sufficiently high frequency, and thus this presents the opportunity for him to be converted or saved from defecting in his laggard categories – an insight that was not visible in previous frameworks such as the framework by Wang and Hong (2006). With the studied sample of shoppers, this aspect is even stronger: since this sample consists of shoppers who shop the majority of their time (weekly) at the studied retailer, the fact that they are not uniformly valuable across categories indicate that in each shopping trip at the retailer, a given user has potential to be converted into valuable customers for those untapped categories, thus increasing overall CLV. With the presented finely granular approach, it is possible to identify these categories, as well as categories where no promotion is necessary. Thirdly, the differences in natural interpurchase times by category as shown by the results suggest that in an information system, the frequency of marketing actions and the corresponding measurement of CLV ex-post should change depending on the product category involved. That is, to save someone from defecting in a fast-moving product category would require the monitoring of his category level CLV at a much faster cycle than a slow moving one, as with the corresponding promotion. This is
in contrast to the existing paradigm which treats all product categories the same, which also only looks at the customer’s visit to the stores – a customer can thus appear to be not defecting overall, even though they may have defected already in specific categories. Finally, this chapter showed how a category level CLV, applied periodically, can be a tool for diagnosing marketing actions. The results show that the current approach of giving everyone the same discount (i.e. a weekly promotion) leads to “wastage”; this means that the sacrificed margin outweighs expected future profits from a particular customer. Like past work on the long term effect of discounts (Mela et al. 1997; Zeelenberg and Putten 2005), this study found that discounts can actually lower the long term value of customers at the population level for a given product category – this study builds on this in that it becomes possible to distinguish who was “saved” via a promotion, and more importantly, whether it was profitable to save them, at the level of product categories. The high resolution CLV can also be combined with the framework by Wang and Hong (2006), which proposed strategies to address customers depending on the temporal dynamics of their CLV - i.e. whether their CLV was rising/falling, consistent across time, or high vs. low – by extending it into the product category level, and adding the dimension of category consistency at the person level. The contribution of the thesis thus enables a retailer to address not only whether a customer should be retained or marketed to, but also in which category it would be most appropriate. The models presented can additionally be extended to allow a CRM or store manager to predict the effect of a promotion; given a dataset up to the present day, the parameters of the presented category-level model of CLV can be estimated, and with a known discount and associated marginal cost, the system can use this model to evaluate whether an additional discounted transaction would actually raise CLV for that customer.

The study is not without limitations and provides several opportunities for further research. First, the effect of discounts was analyzed as though the product categories are independent without interaction effects. A future study should investigate the interaction effect to determine when there is a tradeoff between fostering one category at a time vs. “sacrificing” one category with a discount to gain transactional traffic with another, and whether the overall effect of driving foot traffic to the store offsets these losses. Secondly, loyalty card holders might be different from the larger population of all customers and that a given loyalty card might represent a household with different people with different tastes. Since smartphones allow person-level identification, it might be interesting to compare CLV on a household level vs. CLV on a person level. Furthermore, it is well acknowledged that consumers often split their purchases between multiple retailers (Leszczyc et al. 2000; Rhee and Bell 2002); these external
purchases are in practice unobservable to the retailer. Although retailers only have data about their own loyalty card holders, it might be interesting to consider cross-retailer data in a CLV model for developing a customer profitability management strategy that actively considers store loyalty. A final limitation is that a constant margin of 5% per transaction was assumed for simplicity; when using the presented CLV model in practice, the margin has to be instantiated with the actual margin to provide useful insights in the CRM and to provide a more precise CLV estimate for measuring the effectiveness of marketing actions.

5.5 Conclusion and Future Research
This chapter showed that there is significant room for improvement when it comes to strategies for managing customer profitability. The CLV model was extended for higher data resolution in two aspects. First, this chapter considered a higher spatial resolution by extending the current person-level CLV model to a product category level. The approach was validated with a data set of over 638,000 actual PoS transactions and found that the current person-level approach overestimates the CLV on average by 43%. This illustrates with how much "noise" current CLV approaches operate today and how valuable it is to consider heterogeneity in categories explicitly in the model. These findings will help increase the efficiency of marketing actions significantly. Second, this chapter evaluated the repeated application of the higher spatial resolution model at a higher temporal frequency and showed how this can be used to diagnose marketing actions like promotions. These findings showed how to immediately assess whether a sacrificed margin had a positive long-term effect on retaining customers. Accordingly, it is the position of this thesis that the current practice of infrequent CLV calculations will evolve more towards real-time monitoring, which in turn complements nicely with evaluating personalized marketing measures deployed via smartphones. Taken together, RQ3 is thus addressed; the expansion of the Fader model to the product category level allows the effectiveness of high resolution marketing actions to be evaluated, and acknowledges the customers’ heterogeneous preferences.

Future work could involve building an extensive information system based on the ideas highlighted in this chapter, and deploying it with a retailer’s CRM system for evaluating whether this highly granular customer-category approach of addressing customer marketing show an economical improvement over the existing method.
6. General Discussion and Conclusions

The previous three chapters presented analytical and marketing methods which address key challenges and gaps in the implementation of a multi-channel marketing scheme for physical retailers. This chapter summarizes the main contributions of the previous chapters and discusses the main implications for both research and practice, and closes with a discussion on the thesis’ limitations and directions for future research.

6.1 Motivation and Summary of the Thesis

The starting point of this dissertation was that highly granular marketing is more effective than applying the same marketing policy to all customers (Changchien et al. 2004; Jiang and Tuzhilin 2006), thus motivating the contributions in the thesis. The benefits to retailers include greater customer loyalty and a greater likelihood of repeated purchases at the retailer (Mittal and Kamakura 2001; Olsen 2002; Shankar et al. 2003) that result from the a better targeted and more satisfying offer to the customer (Gummesson 2002; Liang et al. 2007; Xiao and Benbasat 2007). Furthermore, in addition to enjoying spillover effects from improving customer outcomes, with the appropriate analysis, retailers can target specifically customers of value according to their business strategy (Kumar et al. 2006), either to further develop loyal segments or prevent switchers from defecting to another retailer (Gedenk et al. 2010; Wang and Hong 2006).

The core obstacles for physical retailers to enable highly granular marketing is the retailer’s capability to (1) collect customer transaction data, (2) transform the data into insights and (3) operationalize the results with individual level marketing (Arora et al. 2008; Neslin et al. 2006; Oliver et al. 1998; Sheth et al. 2000). With loyalty cards and customer relationship management systems, physical retailers have the means to collect the data necessary for high resolution marketing, but with only direct mail available, they lacked a timely method for operationalizing the results (Gedenk et al. 2010), resulting in mass marketing (Ailawadi et al. 2001; Gedenk et al. 2010) or non-reactive forms of rewarding customer purchases, such as loyalty card bonus point collection schemes (Kumar and Shah 2004; Kumar et al. 2006). Recently, with the advent of smartphones and tablets, physical retailers have also gained the timely method for operationalizing the results (Gedenk et al. 2010). Therefore, of the three major obstacles, the core remaining one is the ability to transform the data into insights. Already, the previously separate channels of E-commerce and physical stores are converging (Enders and Jelassi 2000) into what are now multichannel “bricks-and-clicks” stores (Gulati and Garino 2000), with the dominant pattern in such a setting being browsing for products online and purchasing in-store (Verhoef et al. 2007). Retailers already benefit from an increase in in-store sales even in the
absence of a customization campaign (Avery et al. 2012; Verhoef et al. 2007); a personalized extension with “smartphone-enabled individualized marketing, in-store purchase” therefore shows promise for retailers.

Accordingly, to the core research question of the thesis “how can high-resolution multi-channel marketing be enabled for physical grocery stores?”, the thesis addressed the aforementioned potential of high resolution marketing enabled by smartphones. It specifically addressed the gap of transforming customer data into insights by focusing on specific steps in the multichannel customer management decision (MCMD) framework (Neslin and Shankar 2009). The MCMD framework describes a multichannel strategy consisting of multiple stages: analyzing the customers, developing a multichannel method, designing channels, implementing the marketing and evaluating the results. Since new technologies have already enabled efficient data collection, the thesis focused on the MCMD steps of transforming of data to insights, operationalizing the results, and enabling its evaluation. The thesis therefore provided methods for (1) analyzing the customers to identify a customer base, (2) marketing products using the online channel to promote sales in-store, and (3) evaluating the results of such targeted marketing. Thus, Chapter 3 tackled the customer analysis by presented a variety seeking segmentation scheme for complete marketing personalization - that is, segmentation that consists of only one customer – in order to identify a customers’ heterogeneous preferences, and thus form the basis of finer grained marketing (Dickson and Ginter 1987). Chapter 4 addressed how to leverage the observational learning marketing method common in E-commerce for driving in-store purchases for physical retailers. Finally, Chapter 5 presented a high resolution CLV model as a method for evaluating the results of a marketing method and also for identifying the value of customers in the product categories they shop in. The resulting mapping between the MCMD framework and the contributions of the thesis is illustrated schematically in Figure 28.
6.2 Key Findings and Implications of the Thesis

The present dissertation fills the following research gaps with three distinct empirical studies, the results of which are summarized by answering the research questions raised at the beginning of this thesis.

6.2.1 How Retailers Could Segment Customers in Multichannel Grocery Retailing

**RQ1:** How can a relevant high resolution segmentation scheme be developed for multichannel marketing for physical grocery stores?

Variety seeking, which forms a relevant segmentation base for multichannel marketing for physical grocery stores, can be described by a Poisson distribution: In segmentation, the highest resolution would be the identification of a customer’s unique tastes and (Dickson and Ginter 1987) and targeting them accordingly in “segment of one” marketing (Arora et al. 2008).

Drawing on the literature which observed that customers have strong habits in the domain of grocery retailing (Hoyer 1984; Liu-thompkins and Tarn 2013; Wood and Neal 2009) and typically do not change their purchase patterns until they become satiated with their experiences in a product category (Chintagunta 1999; McAlister and Pessemier 1982), this thesis proposed...
accordingly that the identification of category level variety seeking would therefore be a good psychological basis for finding out where a customer may break out from this habitual pattern, and therefore be open to new offers, recommendations and marketing actions. Using PoS transaction data widely available to retailers as a basis for estimation and analysis, the thesis found that the variety of goods purchased in a product category for a customer fits a Poisson distribution. This relationship holds true even when the analysis was conducted across different granularity levels of product categories and also when the model was estimated at different time points. As such, the relationship is robust against the arbitrariness of product categories as set by the retailer and also the sampling window, and therefore can be generally applied in different grocery retail settings. The Poisson “storyline” is also intuitive; individuals have differing rates of variety seeking and satiation, subject to some stochastic process, and some probability of consuming a certain amount of variety relative to the rest of the population- in-line with the evidence on satiation (Chintagunta 1999; Lattin and McAlister 1985).

The representation of variety seeking as a “variety index” allows an objective and intuitive mathematical comparisons of variety seeking between individuals and between categories and is readily implementable as a basis for segmentation. The Poisson cumulative distribution function (CDF) describes the probability someone bought less than or equal to k products. Since the relationship between the number of unique products bought and the variety index is non-linear, by transforming the Poisson distribution estimated for each category to a corresponding CDF, the value returned by a CDF enables an exact and intuitive mathematical definition of variety seeking of an individual relative to the overall population. This value, dubbed the “variety index”, ranges from 0 to 1 and can be readily compared between categories and also within an individual as well, since it indicates exactly how much variety a person seeks relative to the rest of the population. Thus, the interpurchase times and natural quantities bought which differ between categories are all mathematically captured by the \( \lambda_{\text{Poisson}} \) computed for each product category in the estimated CDFs, which is carried over in the computation of the variety index.

The data-driven estimation of variety seeking can be easily operationalized. It was a key aim of this thesis that the presented results can be transformed into insights and operationalized into marketing results by practitioners. To this aim, this thesis expanded on the variety seeking literature by providing a data driven method which allows the estimation of variety seeking for every product category, without having to ask a customer with psychometric questionnaires how much variety he seeks and in which categories (Baumgartner and Steenkamp 1996; Van Trijp et al. 1996), thus avoiding the practical problems associated with survey non-participation (Nulty
2008). Furthermore, unlike the Trivedi model (1999) for numerically evaluating variety seeking, a customer’s stated preference of the different brands are not required. Since the variety index is a mathematically objective definition of variety, it is straightforward for the retailer to interpret. As per the MCMD framework, analysis of the variety index can thus serve as an initial step of identifying in which customers are interested in trying out new products, and in which category.

Application of the presented segmentation scheme revealed strong heterogeneity in variety seeking between categories within an individual, confirmed the habitual consistency of variety seeking, and therefore provides a method for retailers to identify for which categories to provide an offer to a given customer. The thesis confirmed the body of literature which noted heterogeneity in customer buying preferences for different product categories (Allenby and Rossi 1998; Leszczyc and Bass 1998; Lim et al. 2005). Since previous work on variety seeking only characterized a person’s overall extent of variety seeking, it ignored the heterogeneity within a customer in terms of which categories he would choose to seek variety. The variety seeking segmentation scheme presented in the thesis showed that variety seeking differed strongly between categories in an individual, thus confirming and extending the literature on customer-category heterogeneity to the variety seeking research stream. The between category consistency with which a customer sought variety was characterized in the study with the Gini index, which adds the dimension of consistency to a customer’s overall typology of variety seeking. Furthermore, in addition to within-customer product-category heterogeneity, the study also discovered that the categories in which a customer seeks variety was consistent with time. Therefore, the segmentation scheme simultaneously confirmed two perspectives on customer behavior in grocery retailing: that customers have strong habits in shopping even in terms of which categories they seek variety (temporal homogeneity), while exhibiting heterogeneity in terms of how much variety they seek between their shopping categories (category heterogeneity). Since variety seeking is time consistent, from an implementation point of view, only a one-time estimation is needed; the results showed that even a quarter of a year is sufficient for estimation. Although both questionnaires and the results of this thesis show similar temporal consistency, the difference between the infrequently deployed questionnaires is that the estimation method presented here is data-driven and can be scaled up easily to a large base of customers; applying a variety seeking questionnaire for every category in which a customer shops, and for every customer in a retailer is, in contrast, a difficult endeavor. The data-driven estimation method presented can also be applied regularly as new transactions are recorded, to
see if variety seeking would change over a time horizon longer than the one-year of data studied in this thesis.

Taken together, the combination of variety seeking and the category level heterogeneity gives a method for retailers to identify for which categories to provide an offer to a given customer; an example from Chapter 3 is elaborated on in Figure 29:

![Figure 29: Overall and category level variety indices of four customers (a) and the corresponding typologies and retailer promotion strategy (b)](image)

Figure 29 show the variety seeking indices of four customers in the top five categories in terms of purchase frequency. The customer’s mean variety indices and their variety seeking Gini are also given. Given the customers’ mean variety indices and Ginis, they are mapped onto the typologies and corresponding strategies given in Figure 29 (b). The selection of which category becomes particularly important for those who exhibit a high Gini but only moderate variety seeking (ex. Customer B), since for these customers there are only a few categories where he seeks variety (in this example, Category 2 and 5), and thus, an offer applied without this knowledge could run the risk of rejection by the customer. This is an advantage over the old questionnaire based framework for identifying overall variety seeking; Customer B would appear to be a mid-range variety seeker, but it would not have been possible to identify which in categories. For customers who are more consistent (Customers A, C and D), the general recommended strategy for a retailer would be to target product categories where the variety index is high (ex. Category 5).
The results show, from a product category point of view, which categories attract variety seeking. Additionally, not only can the data-driven method identify individual variety seeking at the category level, but conversely, also quantify comprehensively across all retailer product categories where variety is actually sought. The study revealed that the top categories where customers seek variety are soft drinks, yoghurt, cheese, baked goods, milk and cream; for all remaining categories, the amount of variety sought by the population was also quantified and the results therefore give a quantifiable basis for the selection of specific product groups for future variety seeking studies. In contrast, past studies in variety seeking focused on arbitrary and often merely convenient samples of product purchase data available to the researchers. For example, the study by Kahn et al. (1986) examined sandwich bags, wraps, margarine, cereals and soft drinks while the study by Trijp, Hoyer & J. Inman (1996) focused on beer, coffee, hand rolled tobacco and cigarettes, and finally the study by Trivedi (1999) focused on hypothetical purchases of cola. Therefore, the study presented in the thesis gives insights on the acknowledged research gap of product-category influences on variety seeking (Michaelidou and Dibb 2009; Roehm and Roehm 2004; Tang and Chin 2007). It also gives retailers insights on which categories to focus on for invoking a variety-seeking based marketing – on categories where customers seek little variety, such as toilet paper, it may be a waste of marketing resources to induce customers to try new products. Similarly, for categories where variety is sought, a retailer might leverage this information to target these categories for new product development, possibly with alternate product lines which command a higher margin.

6.2.2 How Retailers Could Leverage Marketing Methods in One Channel to Enhance Performance in Another

RQ2: How can social learning marketing be developed for benefiting physical grocery stores?

Social learning in the form of sales velocity increases the likelihood of a purchase and help retailers promote mid-tail products.

The dominant pattern in multi-channel retailing is browsing for products online and purchasing in-store (Verhoef et al. 2007); by extension, this implies that marketing techniques – such as social learning marketing - which have previously been used only in single channel, pure online environments like E-commerce portals (where both marketing and purchases are online) can also be leveraged to benefit physical grocery retailers. E-commerce retailers have utilized to great effect lists of top ranked products to promote product sales; the higher the sales rank, the more likely customers buy that product (Chen, Wang, et al. 2011; Duan et al. 2009; Tucker and
Zhang 2011). This influence to buy, based on observing what others bought is known as observational learning OL, a form of social learning (Bikhchandani et al. 1998). Thus observational learning has a potential for benefiting physical grocery retailers. However, by examining a set of PoS data, it was shown in this thesis that currently popular products – those which have a high current sales rank – sell well week-to-week and tend to stay that way throughout the year. Therefore, the method of recommending the currently popular product, which was appropriate for infrequently bought durable goods in E-commerce, is not appropriate for grocery retailing, since the recommended product will be items customers already buy, which do not change week to week.

Since observing the static popularity of a product was not appropriate for physical grocery retailers, the thesis investigated the idea of sales velocity. Operationalized as the change in sales rank of the product, the results show that increasing sales velocity positively influences the likelihood of purchases, and that this result is robust across different products. The positive signaling from an improving product was even able to reverse the preference seen in the control group for a higher ranked one – even in the face of the more common sales rank heuristic. For practitioners, the thesis showed that products in the mid-tail of a long-tail distribution exhibit high sales velocity, and therefore, using a list of products sorted by their sales velocity could systematically promote these mid-tail products. In the case of the physical grocery retailer partner examined in this thesis, mid-tail products consist of 41% of the retailer’s revenue, so this is a managerial relevant segment of goods.

Theoretically, the thesis extended the observational learning research stream to account for its velocity effect: since observational learning was defined as the “inference resulting from rational processing of information gained by observing others” (Bikhchandani, Hirshleifer, and Welch 1998, pg. 153), sales velocity can be conceptualized as learning from observing the rate at which the actions occur - the rate change of the actions can be thought of as the “velocity”. It also reaffirmed the literature on intertemporal choice, which observed that that customers have a preference for a sequence of improving outcomes (Loewenstein and Prelec 1993); that is they prefer things that get better and better, and that this effect gets stronger with the speed with which the improvement occurs over time (Hsee and Abelson 1991; Hsee et al. 1991, 1994).

**Sales velocity is robust across four highly different product categories and stronger for hedonic and fast moving goods.** The sales velocity effect was tested in this thesis in three studies, across four product categories. The four product categories were sampled from two common product
classification paradigms; the hedonic paradigm and the non-durable goods paradigm. Products from the extreme points of these two paradigms were chosen: chocolate (high degree of hedonism, non-durable fast moving goods), laundry detergent (low degree of hedonism, non-durable fast moving goods), vacuum cleaners (non-hedonic, durable good) and televisions (hedonic, durable goods). In spite of their differences, the sales velocity effect was shown in increasing the likelihood of a future purchase for all categories. The results also suggest that additionally, the effect is stronger for hedonic and frequently purchased goods (chocolate) and weakest for its diametric opposite (the non-hedonic, infrequently purchased vacuum cleaner). Therefore, for practitioners, sales velocity may be appropriate in promoting grocery retailers’ assortment of hedonic products. This is in line with the benefit-congruency framework by Chandon et al. (2000), which showed that the extent of perceived hedonism in the product affects how customers respond to the type of sales promotions (which they classified as utilitarian – ex. price cuts - or hedonistic like free gifts); the promotion is most effective when the promotion and product type are matched in hedonism. Accordingly, the results imply that sales velocity is congruent with hedonic products.

The results also theoretically confirmed that the selected products, consistent with literature, were perceived distinctly for their hedonic / non-hedonic and durable / non-durable properties (Derbaix 1983; Harlam et al. 1995; Holbrook and Hirschman 1982; Laurent and Kapferer 1985; Voss et al. 2003; Yeo and Park 2006; Zheng and Kivetz 2009). Furthermore, the results complemented the economic perspective of observational learning, which held that the positive signal that is inferred from observing others’ choices should only be a function of how many others have chosen that same product (Chen, Wang, et al. 2011; Duan et al. 2009); product characteristics should play no role in the observation and interpretation of other shopper’s actions.

The studies reveal practical guidelines on a field deployment of sales velocity. The studies conducted have also explored additional factors which influence the strength of the sales velocity effect, which provide some guidelines for retailers on how to practically implement it in the field. In the study on the numerosity effect, it was shown that the numerical framing of the sales velocity can strengthen the sales velocity effect on the likelihood of purchases, by changing the perception of the rank change. In particular, a % change representation led to a larger numerical representation of sales velocity, which in turn boosted the perception of the rank change. Therefore retailers could use alternative units to achieve a stronger sales velocity effect, keeping in mind that the alternative units may expose different products in the long tail. Thus,
which representation is ideal depends on which products the retailer wants to promote. In addition to the numerosity study, additional guidelines on implementation were revealed in the negative sales velocity study. It was shown that a product’s decrease in sales velocity has a stronger positive effect on the likelihood of purchase of the non-decreasing product than a corresponding increase in sales velocity. This means that although negative sales velocity could be used to nudge customers away from the negatively affected product towards another product, it is better altogether for practitioners to avoid displaying negative sales velocity. This can be achieved by displaying the results in a list of products sorted by sales velocity, showing only the results for the top n items.

Theoretically, the contribution of the thesis reaffirms the work on numerosity (Cheema and Bagchi 2011; Chen and Rao 2007; Kruger and Vargas 2008; Monga and Bagchi 2012; Pandelaere et al. 2011; Zhang and Schwarz 2012) and extends it by showing that it could be leveraged on non-price metrics like sales velocity to boost sales. Furthermore, the study on negative sales velocity affirmed the literature on loss aversion, which suggested that losses are perceived to be stronger than gains of equivalent size and that the perception of the loss is relative to the position from where the loss occurred (Hardie et al. 1993; Kahneman and Tversky 1984; Tversky and Kahneman 1991).

Sales velocity can be integrated in a wide range of marketing tools, replacing price cuts. The use of observational learning as marketing is an example of leveraging the strengths particular to the online channel; the sales rank and their change over time, which need to be rapidly computed and updated, are easy to implement online, but difficult to implement in print in physical stores. In the form of sales velocity, they can also benefit grocery retailers and is readily implementable into smartphones: three schematics of how it can be implemented are shown in Figure 30.
Figure 30: Sales velocity embedded as a featured product (a), within a recommendation (b) or as a list of products sorted by sales velocity (c)

Figure 30 (a) features targeted products prominently on a website’s front page, like a normal advertisement, but justifying the product’s presence by stating its sales velocity attribute. Figure 30 (b) features sales velocity embedded within a recommendation, in the context of a variety seeking framing. In other words, sales velocity is used to complement or boost high resolution marketing, tying in with the insights of Chapter 3. In contrast to the schemes proposed in (a) and (b), where the sales velocity information is essentially pushed to the customer, even if they were not actively looking for this information in their decision-making, Figure 30 (c) shows sales velocity within a top ten list of products sorted by their rise in popularity, as a separate tab when customers are exploring products, thereby making it the salient attribute in decision making. These methods can potentially replace the common application of discounts; in (a), it replaces a featured discount and in (c), it replaces a list of discounted products. These methods can save margin in comparison to discounts; while discounts incur a per-use cost to the retailer, the marketing method of social learning persuades on information alone and do not carry this disadvantage.
6.2.3 The Need for High Resolution Customer Relationship Management

RQ3: How can the effectiveness of high resolution marketing actions be evaluated, given customers’ heterogeneous preferences and the capability of retailers to act on these preferences in a timely and precise manner?

Existing models of CLV for evaluating customer profitability ignore the strong between-category heterogeneity in CLV found in this thesis, hence motivating this thesis’ category-level CLV model for evaluating fine grained marketing. As identified by Kumar et al. (2006, pg. 281), one important question in both research and practice regarding CRM profitability is “what is the right metric to manage customer loyalty”? A common method of evaluating customers is customer lifetime value (CLV), which factors in not only the retention aspect (i.e. reduce churn), but whether it is actually profitable to have spent the marketing effort on keeping the customer when projected over the expected lifetime of the customer. The focus of this thesis was on extensions of the model by Fader et al. (2005), which derives a probabilistic expression for the average spending per transaction and an expression for the future number of transactions based on the NBD/Pareto model by Morrison and Schmittlein (1988). The advantage of the Fader model is that it can be readily estimated with point of sale data from retailers, needing only recency, frequency and monetary value as inputs. Past research found that the model could empirically track customer spending and the number of transactions; however, this evaluated customers only at a customer-store level, thus ignoring the possibility that customers can defect in select categories even if overall they have not defected at the retailer. This thesis presented a category-level model of CLV, based on the Fader model. From this model, the CLV index proposed in the thesis offered a dimensionless way of comparing the customer’s extent of lifetime value relative to the rest of the population and also between categories, thus factoring out the differences in natural reference prices between categories. It emerged that even within a customer, they can differ significantly in their CLV indices between categories; in fact the majority exhibit heterogeneity between categories – and therefore, a different propensity to defect.

Therefore, it is possible to identify at the category level which consumers are valuable, and at a person-aggregated level, how consistent (via the Gini) that customer is across categories. The CLV indices and the actual CLV model complement each other; the indices allow the retailer to uncover categories where the customer might be at the risk of defecting or to be rewarded for loyalty, and the actual CLV model allows the retailer to predict whether a marketing investment in this product category would pay off in lifetime value. On a theoretical level, this heterogeneity
in CLV also lends further support to the attribute view of products, which posit that different product categories have different attributes which are “consumed” (McAlister and Pessemier 1982), and given a customer’s individual tastes to the attributes, they would have different purchasing patterns from category to category, and hence lifetime value. Furthermore, since consumers often split their purchases between multiple retailers (Rhee & Bell 2002; Leszczyc et al. 2000), they may have greater lifetime value in certain categories at one retailer while lower lifetime value in others.

In addition to the heterogeneity in lifetime value of customers, the thesis confirmed also that product categories differed significantly in interpurchase times, as well as in the average transaction value. In doing so, the study showed that CLV models at a customer-store level are ignoring credible differences between categories, and therefore obscure the dynamics and behavioral differences of customers between product categories. The results also confirm the observations in literature that the natural interpurchase time of different product categories are different (Neslin et al. 1985).

**Existing models of person-level CLV overestimate a person’s lifetime value.** The traditional CLV at a customer-store level overestimates a customer’s overall lifetime value on average by 43% compared to cumulated lifetime value from each of his individual product categories. This result lends support to the observation that the customer-store level CLV, which aggregates all purchase of the customer together in a given store visit, ignores that the possibility that in specific categories, the customer may have already defected and therefore the actual lifetime value for that category would actually be lower. Therefore the person-CLV is too optimistic and can provide a false positive to the retailer. The category-level CLV complements the existing work by providing a more conservative evaluation of a customer’s lifetime value.

**Application of the Category-level CLV model complements the temporal framework by Wang and Hong (2006).** The framework by Wang and Hong (2006) was one of the first to consider the idea of choosing an appropriate marketing action depending on the lifetime value’s amount as well as two dimensions of time: trend and volatility – that is the change in customer profitability and the temporal fluctuation of the CLV in a time period. The framework by Wang and Hong (2006) proposes what strategies to pursue depending on the temporal trends of the customer overall; the CLV-Consistency framework in this thesis proposes “where” or in which categories to apply it. Intuitively, if a person has a high CLV in a few categories but not others, for this person to have a high CLV, relative to the population he must shop at a sufficiently high frequency, and
thus this presents the opportunity for him to be converted or saved from defecting in his laggard categories – an insight that was not visible when considering the Wang and Hong (2006) framework alone. The combination of the two frameworks for guiding retailers is illustrated in Table 16.
Table 16: Insights from the Wang and Hong Framework combined with CLV-Consistency Framework of this Thesis

<table>
<thead>
<tr>
<th>Combinations of CLV Dimensions</th>
<th>Classification</th>
<th>Corresponding Tactic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{t-1}$ = High</td>
<td>Defecting</td>
<td>Win back tactics: Win back customer share</td>
</tr>
<tr>
<td>$T_t$ = Down</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t \neq$ Loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t$ ≥ Medium</td>
<td>Growing</td>
<td>Upgrade tactics: increase customer share</td>
</tr>
<tr>
<td>$T_t$ ≠ Down</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$V_t$ = Safe</td>
<td>Steady</td>
<td>Loyalty tactics: broaden market spaces</td>
</tr>
<tr>
<td>$P_t$ ≤ Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_t$ ≠ Up</td>
<td>Inactive</td>
<td>Reactivation tactics: Increase customer base</td>
</tr>
<tr>
<td>$P_t$ = Loss</td>
<td>Unprofitable</td>
<td>Cost down tactics: decrease contact cost</td>
</tr>
<tr>
<td>$T_t$ ≠ Up</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) Use CLV-Consistency framework in selecting subset of categories for further investigation

| Gini is high                  | Overall decline in $P$ is due to select low CLV categories. Investigate these. | Overall increase in $P$ is due to select high categories. Investigate low CLV categories for possibilities to upgrade the share. | Has some high CLV categories but overall not valuable. Target the low CLV categories if category CLV model predicts profitability. | Run category level CLV models to see if there are any categories where customer is still profitable; apply model to predict if interventions in these |
| Gini is low                   | Since decline is uniform, investigate any profitable category for targeting. | It means all categories are increasing together. Any category can be investigated for targeting. | Consistently low value across categories. Reactivate a category only if category CLV model predicts profitability. |

(3) Of the subset of categories selected for investigation, apply the CLV-Category model to determine which categories can be profitably developed given a marketing investment and its margin, and the predicted lifetime of the customer.

(4) Drop non-profitable categories from consideration

(5) Apply the Wang and Hong framework to the remaining categories to determine which marketing tactic to apply to each category

As Table 16 illustrates, take for example the case where the overall CLV is decreasing for a customer. The Wand and Hong framework would call for a win back strategy; the question is, in which category? As per the CLV-Consistency framework, if the Gini coefficient of the category level CLV is high, it means the decline can be attributed to the decrease in CLV in a few select categories with low CLV. So this subset should be selected for further investigation. The CLV-Category model should then be applied to determine which of the categories in the selected
subset could actually be profitably developed given a marketing investment and its margin, and the predicted lifetime of the customer. Categories determined to be non-profitable should then be dropped from further consideration. The Wang and Hong framework can then be applied again, this time at the category level to make the final decision on what strategy to apply to these low CLV categories. For a given category, if the CLV has been decreasing but the previous CLV was high, then a win back strategy is applicable. Otherwise a reactivation strategy applies. As another example, in the case where the overall CLV is at a loss, the Wang and Hong framework would call for a decrease in contact cost; again, application of the CLV-Consistency framework would reveal if certain categories have a positive CLV and are still salvageable and profitable, in spite of the overall non-profitability. The CLV-Category model can then be applied to predict whether a marketing action, with a certain margin, could boost these categories, and whether the total CLV is brought back into a positive number as a result.

6.3 Implications for Practitioners: An Integrated Method of High Resolution Marketing

The objective of the thesis was to provide an analytical and methodological groundwork for marketing practitioners to address their customers at a high resolution, motivated by the opportunities of smartphones. The focus was on identifying variety seeking, predicting lifetime value, and leveraging sales velocity. Although is up to practitioners to decide how to use such knowledge for their specific marketing needs, some guidelines on how to apply the ideas are as follows. In this thesis, the multichannel customer management decision (MCMD) framework (Neslin and Shankar 2009) structured the topics and motivated the research gaps addressed by the thesis. The previous chapters showed how to individually address the research gaps framed by the MCMD framework; in alignment with the goal of the thesis to have actionable and implementable results by marketing practitioners, Figure 31 shows the structured analytical steps, enabled by the thesis, which the retailer can take in doing the groundwork analysis required for crafting a high resolution marketing strategy.
Since each retailer may differ in terms of their marketing objective and strategy, it is beyond the scope of the thesis to dictate exact marketing campaigns that can be done with the ideas of this

**Figure 31: Link between the MCMD framework and the concrete analytical and marketing steps to take, as guided by the thesis chapters**

- **Given transaction history of customer:**
  - Pre-selecting profitable customers-category pairs based on their category-level CLV for targeting.
  - (Chapter 5)

- **Given customers-categories selected for targeting:** Check variety seeking tendencies in those categories. Non-discount marketing such as social learning or recommendations may be appropriate for high variety-seeking categories.
  - (Chapter 3)

- **Apply marketing method** - which could include sales velocity
  - (Chapter 4)

- **Determine effectiveness of marketing action.**
  - (Chapter 5)

- **Compute the overall CLV and trend**
  - Use the Wang & Hong framework to select marketing tactic based on overall CLV and trend

- **Compute the category level CLV and GiniS**
  - Use the CLV-Consistency Framework in selecting subset of categories for investigation

- **Of the subset of categories selected for investigation, apply the CLV-Category model to determine which categories can be profitably developed given a market investment**

- **From transaction data, estimate variety indices and GiniS for each customer at each product category**

- **Consider profitable variety seeking categories for targeting with non-discount marketing**

- **Apply sales velocity embedded in a recommendation, or in a list of sorted products, or in a banner advertisement**

- **Recompute the category-CLV post-marketing action, to evaluate the effectiveness of the marketing**
thesis. According, a high level guideline is as follows. Taking the process outlined in Figure 31, a retailer would first use the insights of Chapter 5 to preselect customers for targeted marketing based on expected profitability – retailers should avoid wasting results on unprofitable customers. First, a retailer would compute the CLV at the person level using the Fader model. The Wang and Hong framework can then be applied to determine the high level marketing strategy: a win-back, upgrade, loyalty, reactivation or cost-down tactic. From there, the category level CLV can be computed for individual customers, to select a subset of customer-category pairs for further investigation.

While Chapter 5 reveals which categories are profitable to be developed by retailers, the methods of Chapter 3 can be used to determine in which of these subcategories do customers seek variety, and therefore, indicate customer interest in trying new goods. The high variety seeking, high profit categories would be good candidates for further targeted marketing, particularly on the sales velocity marketing methods of Chapter 4 or with recommendations.

Finally the marketing effectiveness can be evaluated by re-applying the CLV-Consistency framework developed in Chapter 5; at a category level, it can be investigated whether a customer’s CLV increases or decreases as a result of a marketing action.

6.4 Implication for Managers

While the previous sections laid the analytical groundwork for high resolution marketing and gave guidelines on how practitioners might apply the methods for their marketing needs, this section discusses some considerations for managers who wish to evaluate the feasibility and benefits of deploying the ideas discussed in the thesis. With these considerations, a manager in the retail industry can then build a business case according to their specific needs.

6.4.1 Pre-existing Infrastructure

The thesis results are most readily usable for retailers which already have at least a loyalty card, CRM infrastructure and mobile phone channel in place, as it reduces the capital cost of establishing such a system, and more importantly, the existing capability of collecting of such transaction data allows the retailer to benchmark and quantify the impact of implementing the models found in the thesis. Such retailers also benefit from a shorter lead-time in deploying the ideas in the thesis. For retailers which have all of these in place, pilots and business cases for implementing the ideas in the thesis would be straightforward to analyze. For the CLV analysis, for example, one can determine the potential savings that arise if the wasted discounts, as calculated in Chapter 5, were replaced instead with a discount-free promotion such as a
recommendation. The potential can then be tested in a field experiment; out of a pool of customers where discounts were previously shown to have been a waste, one group could be given a discount, another group given nothing, and a final group given a recommendation. The actual conversion rate of the recommendation, relative to its cost of implementation, can then be compared against the conversion rate of the coupon, relative to the coupon cost. As a benchmark for managers, typical coupon redemption rates in the USA were 1% in the offline channel and 10% for online coupons (Esber and Walter 2011). Additionally, for the sales velocity idea, if a retailer has an existing app or website already, then the cost of presenting this information is rather small, and the result on product click-through (Farris et al. 2010) or purchases can be evaluated. For the variety seeking estimation, the cost of deploying a questionnaire to every individual customer for every product category where he shops can be easily compared to the maintenance cost of collecting the CRM transaction data and employing analysts to apply the models in the thesis. For the grocery retailer partner examined in this thesis, there were 146 product categories examined along with 848 customers. If this retailer wanted to map out the variety seeking preferences of these customers for every category, the worst case would be deploying 123,808 questionnaires. As noted by Nulty (2008), participation rate is roughly 30% for online surveys, so at best, the retailer could only identify the variety seeking preferences for 30% of the customer-category pairs.

For retailers who do not have a loyalty card and CRM system, nor a website, it would be more difficult to evaluate the financial gain of implementing the high resolution marketing ideas in the thesis, since they lack data about their customer base at an individual level. In this case, the volume of sales of the retailer would probably play the greatest role in such an evaluation. This is discussed as follows.

6.4.2 Appropriate Types of Physical Grocery Retailers

Regardless of whether the retailer has an existing loyalty card and CRM system, it is expected that the results of this thesis are more helpful to retailers the larger their sales volume and the larger their costs associated with discounts. For large volume retailers, this is because the ideas in the thesis are scalable with technologies like smartphones; namely, the ability to identify the CLV and variety seeking at a customer-category level allows for individual targeting without much additional cost per customer-category. The initial infrastructure cost is expected to be offset the larger the sales volume of the retailer, which would correspond to a greater number of customer-category pairs available for targeted marketing. Particularly in chapter 5, even with a small pool of 848 customers, the total wasted discounts in a year associated with these
customers was 11866.38€. For retailers with a larger pool of customers, this cost would scale up, which can be addressed with high resolution marketing. For example, in the study from Singh et al. (2006), he describes the established retailer under study from “a small town on the East Coast” of the USA to have 10,000 loyalty card holders, who cover 85% of all sales in the store. As a rough estimation, if one assumes this retailer has a similar per customer wastage as the retailer studied in this thesis, a simple linear scaling of the 848 customers to 10,000 would lead to potentially 139,930.78€ in wasted discounts; retailers in even larger markets could potentially incur a larger cost. Similarly, it would be expected that the ability to offer a sales-velocity motivated promotion, that replaces a discount, scales to greater savings the higher the cost of discounts for that retailer. Additionally, the application of sales velocity or a recommendation in the place of a discount might also remove the well documented negative side effects of discounts, such as stockpiling (Gedenk et al. 2010). As this is an unexplored field, a retailer interested in using sales velocity should thus pilot and test it in field in order to determine its success rate and to what extent it can substitute discounts.

In the scientific literature, the retailer’s strategy with respect to every day pricing, sales volume and application of discounts can be classified by the EDLP – HiLo taxonomy (Bell and Lattin 1998; Haans and Gijsbrechts 2011; Hoch et al. 1994; Moreau et al. 2001). Namely, two classes of retailers which emerge are those who offer every-day low pricing (EDLP) for a wide assortment of products, characterized by high volume sales at low prices and low profit margins, and the stores which overall charge higher prices for their products and earn a higher margin, but offer temporary deep discounts in select categories (called HiLo). Thus, within the EDLP and HiLo taxonomy, the thesis results are expected to help EDLP retailers because their high volume allow for the scaling of benefits, and the results can help HiLo retailers due to the deep cost of discounts. Since the models of this thesis were estimated from loyalty card data of a HiLo retailer, it is likely that the pattern of shopping behavior found in the thesis holds greater similarity with other HiLo retailers as opposed to EDLP retailers. For EDLP retailers, since they by definition typically do not focus on discounts at all, rather than replacing a discount, sales velocity should instead be used as a method of increasing the basket size of the customer – i.e. targeting products which they might not normally buy, but might try for the sake of variety. This use case can be assisted with the variety seeking estimation as shown in the thesis. Furthermore, since margins are low - typically 1-2% in the USA market (Singh et al. 2006) - and a large customer base is needed to make a EDLP business viable, tracking category-level CLV might be a valuable method of maintaining proper volume, by focusing on customer-category retention.
Note that the CLV computation presented in thesis already has a built in margin parameter, which would then allow the EDLP retailer determine whether it makes sense to spend money retain the customer. In certain markets it should be noted, however that – for example, in Switzerland - discounters (such as Aldi, Lidl and Denner) who offer EDLP do not deploy loyalty cards, while all of the HiLo retailers such as Migros, Coop and Manor all have a loyalty card. Without the loyalty card or similar CRM-linked system, it becomes difficult, if not impossible to implement the variety seeking and CLV estimation in this thesis. Such types of EDLP retailers might not benefit fully from the ideas of this thesis.

6.5 Limitations and Future Research

To summarize, the research in this thesis contributes to laying the analytical and marketing groundwork for high resolution marketing as enabled by ubiquitous technology such as smartphones. The thesis also lays the groundwork for future projects and developments, both in terms of follow-up research studies and also for large-scale multi-channel deployments with a physical grocery retailer. They are summarized below.

Compare the variety index with past questionnaire measures in quantifying individual-level variety seeking. The results in the thesis offered a mathematically objective measure of a person’s variety, namely, as a statistical measure of variety seeking relative to the population at the category level. In order to close the link with existing psychometric work (Baumgartner and Steenkamp 1996; Van Trijp et al. 1996), future work should estimate variety seeking on a consistent pool of customers using both the method of variety indices proposed in this thesis and the existing psychometric questionnaires. The results can then be compared to their extent of agreement. Practically, however, given the large number of categories, the non-participation issues of questionnaires as noted by Nulty (2008), and the need for a willing retailer to deploy such a questionnaire, this validation is likely only possible by sampling a subset of products and customers.

Investigating the link between variety seeking and customer behavior in response to recommendations and social learning. Past research suggest that customers who pursued higher variety tend to increase overall consumption quantity (i.e. variety consumption has an additive rather than a substitutive effect) (Kahn and Wansink 2004; Read et al. 1995; Simonson 1990). Ailawadi et al. (2001) established that customers who seek variety are particularly open to new offers and experiences, which suggests that they may be particularly receptive to discovering other products such as recommendations and social learning – thus potentially
allowing an alternative to costly price discounts for this subset of customers. However, prior to this thesis, variety seeking was typically measured with questionnaires, rather than from an information system, and thus it was not easy to simultaneously measure variety seeking and evaluate a person’s receptiveness to recommendations or social learning - as such, verifying this potential link was difficult. With the results of this thesis, it is now possible to quantify variety seeking without a questionnaire, and therefore, investigate this link. An ideal setup, therefore, for future research in this area would be a retailer with point of sale data and also a recommender system and social marketing measures implemented in their online and mobile channels; the variety seeking model can then be used to classify customers and an empirical link can be established between their extent of variety seeking and their receptiveness to recommendations and social marketing. Such a result would benefit practitioners, in that they could validate the value of measuring variety seeking, and it would also benefit researchers since it would intersect a previously pure marketing topic (variety seeking) with what was a topic purely studied from the information retrieval literature (recommendation systems).

**Estimation of Variety Seeking and CLV with other data pools.** The results in this thesis showed that there were strong category differences within-customers in terms of their lifetime value and extent of variety seeking. While there is no theoretical reason to suggest why this would be different with any other retailer, for generalizability of the conclusions made in the thesis, in future studies it would be desirable to repeat the estimation of Variety Seeking and CLV with other PoS data sources to check whether the extent of within-person category-level heterogeneity holds true in other retailers and domains as well.

**Validation of the sales velocity effect in the field across different product categories with real monetary cost associated with decisions.** The sales velocity effect was tested in this thesis in three studies, across four product categories. The four product categories were sampled from two common product classification paradigms; the hedonic paradigm and the non-durable goods paradigm. Products from the extreme points of these two paradigms were chosen – but in spite of their differences, the sales velocity effect was shown in increasing the likelihood of a future purchase for all categories. Although promising, the studies did not test sales velocity in the presence of price cues nor with monetary consequences in the decisions of the participants. Thus as a first step, future work should test whether the sales velocity effect is diminished in the presence of price and actual monetary loss, similar to the incentive-aligned studies (Miller et al. 2011). This can be achieved by having choices which have real monetary consequences. Furthermore, the sales velocity should be deployed in a retailer’s smartphone app or E-
commerce portal in order to study in field - with real choices - its effect on customer decision making. The design of such a field experiment in terms of how to display the sales velocity information could use some of the guidelines presented in the thesis.

**Combine all ideas of the thesis and evaluate them in a field deployment.** The developments in this thesis suggest that they could help retailers enable high resolution marketing. The ideas were evaluated independently in three separate studies. However, since the ideas of the thesis are meant to work together for a multi-channel retailer, a full validation should therefore attempt to fully integrate all of the ideas of the thesis within a retailer’s multi-channel CRM and marketing strategy, as per the idea outlined in Figure 31. In this manner, the goal of the thesis, which was to have actionable and implementable results by practitioners, can be really evaluated in terms of a real-world implementation. However, a field deployment – whether partial or full - where the ideas of the thesis are integrated with the technology and processes of a physical grocery retailer is not trivial and could take years, involving key stakeholders in the IT, software development, marketing and business analytics departments of such companies.
7. Bibliography


Appendix: Questions and Measures of Chapter 4 Studies

A.1 Main Dependent Measures (used in all three Chapter 4 studies)

<table>
<thead>
<tr>
<th>Question</th>
<th>No Difference</th>
<th>Definitely Offered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Which offer would you probably buy if you were in the market for this product?</td>
<td></td>
</tr>
</tbody>
</table>

A.2 Manipulation, Attention Checks and Confounding Checks (used in all three Chapter 4 studies)

<table>
<thead>
<tr>
<th>Question</th>
<th>No Difference</th>
<th>Definitely Offered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2</td>
<td>How would you compare the two offers in terms of price?</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>How would you compare the two offers in terms of sales rank change?</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>How would you compare the two offers in terms of current sales rank?</td>
<td></td>
</tr>
</tbody>
</table>

A.2.1 Degree of Involvement (used in Study 1 of Chapter 4):

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly Disagree</th>
<th>Neutral</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5</td>
<td>I am particularly interested in chocolates.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Q6</td>
<td>Given my personal interests, chocolates are not very relevant to me.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Q7</td>
<td>Overall, I am quite involved when I am purchasing chocolates for personal use.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Q8</td>
<td>Overall I am quite involved when I am purchasing chocolates for others.</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Q9: How different do you consider the two offer descriptions? (1 = “very small”, 7 = “very large”).
A.2.2 Extent of Hedonism and Frequency of Purchase (used in Study 2 and 3 of Chapter 4):

Question Text: For each statement below, indicate how close to the adjective that you believe best describes your feelings about <Product category>. The more appropriate the adjective seems, the closer you should place your mark to it.

<table>
<thead>
<tr>
<th>Q10</th>
<th>Unenjoyable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Enjoyable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q11</td>
<td>Not Frequently Purchased</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>Frequently Purchased</td>
</tr>
</tbody>
</table>