An investigation of the validity of virtual reality for shopper research.

Submission for the degree of Doctor of Philosophy

Katherine Anderson

M.Bus (Research), B. Management (Marketing), B. Arts (International Studies)

Supervisors

Professor Rachel Kennedy

Professor Larry Lockshin

Associate Professor Anne Sharp

Nevertheless, she persisted.

Mitch McConnell

ACKNOWLEDGEMENTS

I would like to thank the following companies for providing data for this thesis:

Advanced Simulations
Kantar TNS
Kimberly-Clark
Nielsen
InContext Solutions
Mars-Wrigleys Inc.
Mai Research
Mondelez International, Inc.
The Sun Products Corporation

Thanks also to the industry practitioners who generously shared their time and expertise in-depth interviews. Your input and candor were greatly appreciated.

Thank you to the University of South Australia and the Ehrenberg-Bass Institute for Marketing Science for financial support that assisted me to undertake this research.

To Associate Professor Erica Van Herpen and Associate Professor David Bednall, thank you for the kind and insightful suggestions you made as examiners of my thesis.

To my supervisors, Rachel Kennedy, Anne Sharp and Larry Lockshin, thank you for your support, understanding and unwavering trust in my abilities. You read many, many versions of my work, supported me in late night (and early morning) conference calls and provided sound counsel. For your support and patience, I am grateful.

To my family, my husband Luca, and my friends and colleagues in Melbourne and Adelaide, thank you for your patience, humour and support as I completed this challenge. You made me feel like I wasn't alone.

Otto and Obi von Schnapps, my best Schnauzer boys, thanks for keeping an eye out for intruders as I typed and for taking me out for walks in the sunshine.

ABSTRACT

A computer-generated 'virtual' simulation of a real store is an innovative research method that requires further validation to support its use in marketing. This research investigates how accurately consumers' usual brand and product choices are captured in virtual reality (VR) simulations of familiar retail environments. It also investigates how the fielding mode and protocols of VR studies affect the validity of results.

Data from eight VR studies (n = 4463 VR purchases) is benchmarked against purchase data collected in real stores and multivariate modelling is used to confirm what biases are consistently evident. Additionally, the experiences and concerns of 40 marketing managers and market researchers that have used VR for marketing research are discussed.

This research establishes that a number of biases affect VR purchase data and that contrary to expectations, half of all participants select products or packs sizes they do not usually buy in real stores despite being instructed to shop normally.

Regression analysis establishes that five variables, including the relative price, size and popularity of products, relate to biases in the selection of products.

Descriptive and regression analyses demonstrate that deploying VR studies with practice trips, as recommended by prior research, does little to reduce these biases in VR studies for frequently purchased hedonic, or utilitarian, grocery products. On average the shares of products differed by just one or two percent between studies that simulated one or more 'normal' shopping trips, and the number of trips simulated was not significant predictor of bias in regression analyses. Similarly, analyses demonstrate that fielding studies with specialised equipment and in-person researcher supervision, does not affect the magnitude of the biases evident. Modelling found that the VR shares of products were overestimated in VR studies fielded in specialised facilities and in studies fielded online using panel members' home computers.

Despite these issues, VR may still be a good method for testing how changes to pack designs, product assortments or shelf layouts in real stores may affect shoppers' purchase choices. However, they do indicate that studies must be carefully designed and that calibration models to correct for the biases evident in VR data, should be devised.

Additionally, this research demonstrates that VR simulations limited to one area of the store, such as the confectionery aisle, and displayed on normal computer screens can replicate the effects of shelf layout and assortment organisation changes in real stores, using data from two studies. The studies examined provided sales estimates within 1% of sales estimates from real stores, suggesting that the extensive simulations and specialised

equipment used in some studies (e.g. Burke and Leykin 2014, van Herpen, van den Broek et al. 2016, Waterlander, Jiang et al. 2015) are not required to obtain useful results.

A key limitation of this research is that protocols which may improve the realism of VR purchase data, for example having participants shop for products they actually need or introducing real financial consequences to experiments, were not tested. Future research should address this gap.

TABLE OF CONTENTS

CHAPTER 1 ..... **13**
INTRODUCTION 13
THE USE OF VIRTUAL REALITY FOR MARKETING RESEARCH 13
KEY TERMS 14
FOCUS OF THIS RESEARCH 14
KEY FINDINGS 18
STRUCTURE OF THIS THESIS 20

CHAPTER 2 ..... **24**
REVIEW OF PUBLISHED FINDINGS 24
PART 1: VALIDATION RESEARCH FOR VR 24
PART 2: METHODS RESEARCH FOR VR 35

CHAPTER 3 ..... **49**
INVESTIGATION OF INDUSTRY KNOWLEDGE 49
QUALITATIVE RESEARCH: APPROACH 49
QUALITATIVE RESEARCH: FINDINGS 51

CHAPTER 4 ..... **65**
RESEARCH QUESTIONS 65

CHAPTER 5 ..... **69**
QUANTITATIVE DATA SOURCES & RESEARCH APPROACH 69
NATURAL EXPERIMENT BASED ON INDUSTRY DATA 69
ANALYSIS APPROACH 89
LIMITATIONS OF THE RESEARCH APPROACH 94

CHAPTER 6 ..... **98**
STUDY 1: BIASES FOR BRANDS 98
OVERVIEW OF THEORY 98
STRUCTURED INVESTIGATIONS 100
DATA SOURCES & METHOD: OVERVIEW OF PERTINENT DETAILS 103
RESULTS: CONSISTENCY OF PREMIUM BRAND BIAS ACROSS STUDIES 107
RESULTS: CAUSES AND EXTENT OF UPSCALING BIASES 108
CHAPTER SUMMARY 122

CHAPTER 7 ..... **124**
STUDY 2: BIASES IN PACK CHOICES 124
OVERVIEW OF THEORY 124
STRUCTURED INVESTIGATIONS 125
DATA SOURCES & METHOD: OVERVIEW OF PERTINENT DETAILS 128
RESULTS: CONSISTENCY OF LARGE PACK BIAS ACROSS STUDIES 132
RESULTS: POTENTIAL CAUSES OF PACK SIZE BIASES 133
CHAPTER SUMMARY 147

CHAPTER 8 ..... **149**

STAGE 2: CONFIRMATORY MODELLING 149
INTRODUCTION & OVERVIEW OF THEORY 149
DESCRIPTION OF DATA 151
EXPLORATORY MODELLING 154
RESULTS: REGRESSION MODELLING 160
CHAPTER 9 169
STUDY 3: SHELF POSITION RESEARCH 169
OVERVIEW OF THEORY 169
STRUCTURED INVESTIGATIONS 170
VR STUDIES: PERTINENT DETAILS 170
RESULTS: CONSISTENCY OF REAL AND VIRTUAL EFFECTS 174
CHAPTER SUMMARY 179
CHAPTER 10 180
SUMMARY AND DISCUSSION OF FINDINGS 180
SUMMARY OF QUANTITATIVE FINDINGS 181
DISCUSSION OF KEY FINDINGS 183
IMPLICATIONS FOR THE USE OF VR FOR SHOPPER RESEARCH 187
IMPLICATIONS FOR THE DESIGN OF VR STUDIES 190
LIMITATIONS & RECOMMENDATIONS FOR FUTURE RESEARCH 195
APPENDIX 1 202
COMPANIES USING VR FOR RESEARCH 202
APPENDIX 2 205
IN-DEPTH INTERVIEW MATERIALS 205
RECRUITMENT MATERIAL 205
PARTICIPANT INFORMATION SHEET 208
INTERVIEW DISCUSSION GUIDE 209
CHARACTERISTICS OF INDUSTRY PARTICIPANTS 211
APPENDIX 3 213
STRUCTURE OF VR DATA & METRICS AVAILABLE 213
APPENDIX 4 217
DETAILS OF DATA SETS 217
APPENDIX 5 226
EXAMPLES OF QUESTIONNAIRES USED 226
APPENDIX 6 244
DESIGN CONSIDERATIONS FOR VR STUDIES 244
APPENDIX 7 253
ADDITIONAL RESULTS FOR CHAPTERS 6,7 AND 8 253
REFERENCES 268

LIST OF FIGURES

Figure 1: Example of a virtual reality environment used by marketers in 2011 to test new shelf signage14
Figure 2: Examples of different types of products and retail environments simulated..... 16
Figure 3: Examples of environments used in pre-2000 and post-2010 VR research 25
Figure 4: Examples of 2D and 3D product images..... 38
Figure 5: An example of how price tags were displayed in 3D VR projects..... 38
Figure 6: Examples of how single and multi-screen setups compare..... 40
Figure 7: Example of head-mounted VR displays and the view of participants 40
Figure 8: Example of VR simulation where multiple categories could be shopped 46
Figure 9: Example of a large screen used to display VR environments to retailers..... 60
Figure 10: Shoppers could control their view of the shelf and select products to view..... 76
Figure 11: Example of a less realistic virtual environment 77
Figure 12: Examples of retail environments not validated..... 78
Figure 13: Illustrating poor resolution of shelf layout information available for some studies..... 79
Figure 14: Example of how starting positions were varied in some studies, afffecting which products were seen first..... 81
Figure 15: Example of a study where large packs were more visible than small packs 127
Figure 16: Example of a study where smaller multi-packs packs were more visible than large packs 128
Figure 17: How pack and product size is conveyed in VR 138
Figure 18: Assumed attention to bagged confectionery packs..... 143
Figure 19: Results of colour composition analysis for VR confectionery environment 145
Figure 20: Scatter plots for *VR bias* and *Real share* biases 156
Figure 21: Distribution of model residuals: actual and normal distributions 158
Figure 22: Normal probability plot of actual and theoretical data 159
Figure 23: Plot of residuals against the predicted values of *VR Bias*..... 159
Figure 24: Different shelf layouts tested in Study 6 171
Figure 25: Example shelf layout tested in Study 7 173
Figure 26: Example of a VR rig in which participants walk to move forward..... 199
Figure 27: Structure of VR data..... 213
Figure 28: Alternative starting positions in a category-focused virtual simulation..... 244
Figure 29: View of virtual shelf adjusted for a moderately tall shopper 246
Figure 30: Purchase quantities from multi and single category simulations 250
Figure 32: Residual plots for IVs..... 262

LIST OF TABLES

Table 1: Sample of prior findings regarding the real-world validity of VR data 26
Table 2: Similarity of real and VR purchase data across different product categories 28
Table 3: Findings regarding the value of 3D visualisation in VR studies 37
Table 4: Amount of time spent shopping and size of VR simulations 45
Table 5: Product range for each VR simulation selected for analysis 72
Table 6: Changes tested in shelf layout studies available 81
Table 7: Scale of VR studies selected for analysis 82
Table 8: Computing interfaces of VR studies selected for analysis 82
Table 9: Summary of studies used to address each RQ 84
Table 10: Summary of studies used to address each RQ 84
Table 11: Details of the VR studies selected for analysis 86
Table 12: Validation data at the store level 88
Table 13: Example of how results are analysed and reported 91
Table 14: Relationship between number of brands included and av. deviation figures 92
Table 15: Purchase shares of premium grocery brands in VR studies 107
Table 16: Validation of the sample for VR Study 1, Confectionery 108
Table 17: Validation of the sample for VR Study 5, Paper towels 109
Table 18: VR purchase choices of different demographic groups – price tiers 110
Table 19: Individual level consistencies in real-virtual brand choices 111
Table 20: Results for brands with large and small price premiums 112
Table 21: Biases for premium brands in virtual simulations: Brand level results 114
Table 22: Results for premium brands in hedonic product categories 115
Table 23: Results for premium brands in utilitarian product categories 116
Table 24: Results for large-discount PL brands in virtual studies 117
Table 25: Results for small-discount PL brands in virtual studies 118
Table 26: Results for PL brands in simulations fielded in person and online 119
Table 27: Results for premium brands in simulations fielded in person and online 120
Table 28: Shares of price tiers on consumers' first and second trips 121
Table 29: Purchase shares of larger pack sizes in real and virtual studies 132
Table 30: VR purchase choices of different demographic groups 134
Table 31: Validation of the sample for VR Study for baby nappies 135
Table 32: Claimed purchasing of different pack sizes VR purchases 136
Table 33: Purchase shares of larger pack sizes in-store, online and in VR 137
Table 34: Proportion of consumers who rotate a product they selected 139
Table 35: Results for pack sizes in simulations fielded with different screen sizes 140
Table 36: Shares of larger pack sizes in consumers' first and second trips 140

Table 37: Purchase shares of larger pack sizes across categories 141
Table 38: The influence of visual salience on purchase choices 146
Table 39: Pack size and price values across (product) categories 153
Table 40: Impact of aggregation on correlation coefficients 157
Table 41: Regression model: coefficients and *significance* 161
Table 42: Regression model from selected product categories: coefficients and *significance* 163
Table 43: Regression model to estimate precisely the significance of the number of trips simulated 165
Table 44: Penetration different product types in the sample selected 171
Table 45: Impact of organising the category 174
Table 46: Real-VR purchasing of different product types 175
Table 47: Impact of moving products in aisle 176
Table 48: Impact of moving brands to eye level shelf positions in Study 6 177
Table 49: Impact of moving brands to eye level shelf positions in Study 7 for feminine care products 177
Table 50: Summary of quantitative investigations 181
Table 51: Group 2 Market researchers (n = 16) 211
Table 52: Group 3 Industry experts (n = 3) 211
Table 53: Group 1 Marketing managers (n = 21) 212
Table 54: List of metrics available from a typical VR study 215
Table 55: Impact of viewing angle on purchase choices in a confectionery study 247
Table 56: Impact of viewing angle on purchase choices in a petcare study 247
Table 57: Purchase choice results from a VR simulation with multiple categories shown 249
Table 58: Results from single category simulations with (slightly) different instructions 251
Table 59: Proportion of buyers and sales represented by VR recruitment criteria 252
Table 60: Statistical tests for premium grocery brands 253
Table 61: Premium brand shares of shoppers of different ages in VR study 253
Table 62: Statistical tests for PL grocery brands 254
Table 63: Statistical tests for brands fielded online and in person 254
Table 64: Statistical tests for pack size choices 254
Table 65: Results from 'seek and find' tasks 256
Table 66: Colour analysis results 256
Table 67: Descriptive statistics for variables included in regression analysis 260
Table 68: Correlation matrix for all variables (*n* = 182) 261
Table 69: Comparability of samples for shelf layout study reported in Chapter 9 267

ACRONYMS & ABBREVIATIONS

2D	Two dimensional, i.e. lacking depth
3D	Three dimensional
C.I.	Confidence interval
C.L.	Confidence level
CPG	Consumer packaged good
ct	Count of items in package
DV	Dependent variable
FMCG	Fast moving consumer good
HCI	Human computer interaction
IQR	Interquartile range
IV	Independent variable
M	Mean
MR	Market research
MS	Market share
ms	Millisecond
N/A	Not available
NB	National brand
N.S.	Not statistically significant i.e. $p > .10$
oz	Ounces in package
PL	Private label brand
SD	Social desirability bias
SKU	Stock keeping unit i.e. a product with a distinct barcode
StD	Standard deviation

Stat.Sig. Statistically significant

VR Virtual reality

WIMP Window, icon, menu, pointing device, a term that encompasses the keyboards, mice and touch screens with icons and menus

X-Large Extra large, generally in reference to a computer screen or pack size

CHAPTER 1
INTRODUCTION

Overview

This thesis investigates the validity of virtual reality shopping simulations for shopper research. This chapter outlines why this research was conducted, the key questions investigated and the most important findings.

The use of virtual reality for marketing research

VR simulations enable marketers and marketing researchers to recreate the atmosphere of a real shopping environment using images and computer programming. In 1996 the Harvard Business Review published an article hailing virtual reality (VR) as a "breakthrough" in marketing research (Burke 1996) and in 2009 to major industry bodies, the Advertising Research Foundation (ARF) and the Path to Purchase Marketing Institute published a report promoting VR as a market research tool.

When VR simulations were first introduced in the 1990s they often used video footage from real stores and relied on shoppers verbalising what they would buy. By the early 2000s, virtual environment and environments could be created with computer modelling (see for example Figure 1) and shoppers could click on products to view close-ups. This enabled marketers to use VR simulations for a range of pricing, merchandising, retail and packaging tests and test retail environments before they were built (Bock and Treiber 2003, Vrechopoulos, O'Keefe et al. 2004, Berneburg 2007, Wasserman 2009). Since 2000, VR technology advanced to include head mounted displays (HMD) that participants wear, ultra-wide screens that fill shoppers' entire field-of-view and avatars to help participants feel present in VR environments (Porcheddu and Venturi 2011, van Herpen, van den Broek et al. 2016, Wright, Schnack et al. 2017).

Key Terms

The following section defines key terms used in this thesis.

Virtual reality (VR) is as a computer-generated representation of a three-dimensional space, designed to generate the feeling of being physically present in the space visualised. In VR, users view the space from a first person perspective and control the environment in real time (Hale and Stanney 2014). The ability for shoppers to interact in real time is one of the key differences between VR and other methods of research that use computer-generated images such as choice studies, shelf projection studies and pack tests. In all these methods, the products that shoppers view, is pre-determined by researchers.

The **virtual environment** is the place or space depicted; it can be a real location, imagined, or a combination of both. Virtual environments can be large in scale, for example, e.g. a 'virtual world' comprised of many scenes through which a user can travel (e.g. Second Life, Assassins Creed); or small in scale, for example a single room or area.

In the VR literature, **presence** refers to the psychological sensation of feeling 'present' in a virtual environment when physically situated elsewhere (Witmer and Singer 1998). While some VR environments are more successful in inducing feelings of presence, all VR environments aim to evoke it to some extent.

Figure 1: Example of a virtual reality environment used by marketers in 2011 to test new shelf signage



Focus of this research

This research investigates the validity of data collected in VR studies for marketing research and how the fielding and design of studies affects the validity of VR data. While a number of studies have shown how VR can be used to visualise new products, merchandising concepts and store designs and to increase consumer engagement with e-commerce and virtual

world sites (Rosenberger and Chernatony 1995, Burke 2002, Vrechopoulos, O'Keefe et al. 2004, Manganari, Siomkos et al. 2011, Gadalla, Keeling et al. 2013, Hansen 2013, Papagiannidis, Pantano et al. 2013, Scaraboto, Carter-Schneider et al. 2013, WARC 2016a), high quality research demonstrating that VR collects realistic shopper data that can be generalised to real purchase situations is scarce.

There are a variety of metrics and behaviours that must be validated if VR is to be trusted for a variety of marketing and shopper research but in the 25 years since VR's introduction, only 16 published studies have compared VR data to data from real purchase situations and these have been limited to a small number of categories and retail contexts. Explorations with industry users of VR found that only nine have conducted validation research and have done so using a variety of approaches.

Given the technological advances achieved over the past 25 years (Cohen and Gadd 1996, Porcheddu and Venturi 2011, van Herpen, van den Broek et al. 2016, Wright, Schnack et al. 2017), research testing and benchmarking the contribution of each element of VR is also essential. Less than half of the VR validation studies published in the marketing literature tested elements of technology or study design, so many gaps in knowledge remain, even when literature from other domains of research, such as Human-Computer Interaction (HCI), is utilised.

This research aims to contribute to the body of knowledge for VR research and improve industry practices by addressing six key areas.

RQ1: How accurately are consumers' real purchase choices captured in VR?

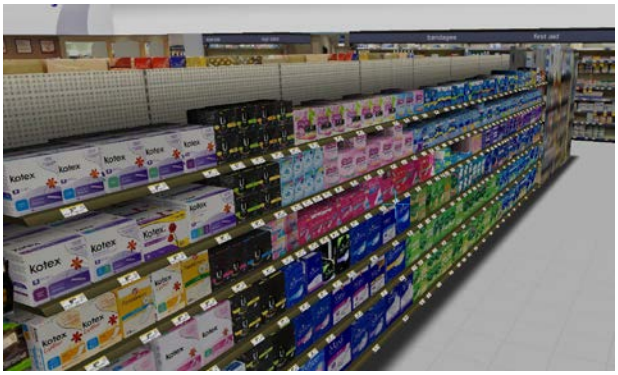
Shoppers' brand, product and pack size choices are compared to survey data collected prior to four VR studies commissioned by consumer packaged goods companies, to establish how representative VR data is of shoppers' usual purchase repertoires. Prior research assumed that most shoppers selected products from their usual purchase repertoires because the correlation between the real and VR purchase shares of products was high (e.g. $r = .80$) or because participants claimed to have chosen realistically in post-shopping surveys (Cohen and Gadd 1996, Campo, Gijsbrechts et al. 1999, Tjostheim and Haugland 2005, Leister, Tjøstheim et al. 2007). This research establishes what proportion of shoppers chose products consistent with their real brand, pack size and product choices in four VR studies.

RQ2: Are VR purchase data for some brands, pack sizes and types of products consistently biased?

Product choice data from eight VR studies (n = 3963 VR purchases) is benchmarked descriptively to sales data from real stores and analysed via regression analysis to establish what biases are consistently evident across VR studies.

Past studies have reported that there are differences between the real and VR shares of *some* national brands and *some* private label brands in VR studies for grocery stores (e.g. Burke, Harlam et al. 1992, Higaki 2004, van Herpen, van den Broek et al. 2016). However, past studies did not establish whether different types of grocery products are equally affected by biases for national and store brands or shoppers tailor their purchases to different types of retail environments.

Figure 2: Examples of different types of products and retail environments simulated





RQ6: Does VR provide valid shelf layout results when shown on a normal computer screen?

Another key area of interest for marketers is whether VR can be used for shelf layout tests (Breen 2009, Porcheddu and Venturi 2011, Massara, Porcheddu et al. 2014, van Herpen, van den Broek et al. 2016). Data from VR studies that tested substantial changes to the arrangement of products are compared to benchmarks from studies in real physical stores to determine if VR simulations provide valid estimates of the impact of shelf layout changes on brand and category sales. While other studies have investigated this question, they have used unrealistic shelf layouts or head mounted VR displays (van Herpen, van den Broek et al. 2016) (Porcheddu and Venturi 2011) (Massara, Porcheddu et al. 2014), methods not typically used by industry researchers when researching shelf changes due to their cost and concerns about ecological validity (Treiber and Needel 2000, Breen 2009).

RQ3: Why are biases apparent? To what extent are they a consequence of how VR studies are conducted and the technology utilised?

Visual attention, product interaction and purchase data from five VR studies is used to investigate why differences between real and VR purchase estimates are evident, a question not adequately investigated by prior research. Product choice data from real online shopping trips, an environment where shoppers cannot feel, smell or touch products, is compared to VR data to establish if the similar absence of sensory information in VR studies is likely to be responsible for some of the biases evident. Additionally, regression

modelling is used to validate if fielding and technological differences between studies are significant sources of bias.

RQ4: Does fielding VR studies online affect the generalisability of VR data to real physical retail environments?

Data from VR studies conducted using online panel members' home computers and collected in specialised research facilities in shopping centres is benchmarked to sales data from real physical stores. Fielding VR studies with pre-recruited online research panels is more cost effective than recruiting shoppers in public places (Breen 2009). However, as the majority of grocery purchases still take place in physical stores (Nielsen 2015), establishing the generalisability of VR shopping data collected online is important.

RQ5: Are multiple trips necessary to collect accurate purchase choice data?

Purchase data from VR studies where one or more 'normal' shopping trips were simulated, is compared to sales data from real stores to determine whether additional trips improve the realism of VR purchase choice data. Additionally, the number of shopping trips simulated is included in regression analysis to factors that are significant predictors of the size of biases across studies.

While past studies concluded that multiple trips are necessary to collect accurate purchase data (e.g. Burke 1996, Campo, Gijsbrechts et al. 1999, Sharpe, Staelin et al. 2008), they recruitment and fielding costs substantially (Needel 2005), can be burdensome to participants (Mueller, Lockshin et al. 2010) and may not be necessary as consumer acceptance of VR increases (WARC 2016b).

Key findings

The key findings from this thesis are now discussed to highlight the contributions of this research.

Finding 1: Many shoppers select brands, products and pack sizes in VR simulations that they do not typically buy in real stores

While past studies presumed that consumers selected products from their usual purchase repertoires in VR simulations because correlations between real and VR purchase shares were strong, direct comparisons found that 36% of shoppers selected a pack size they did not typically buy when shown a VR product selection and instructed to "shop normally". Additionally, 51% chose a brand they had not purchased in the last month and 22% selected an different type of product i.e. microwavable, than they typically purchased.

Additionally, in VR studies demographic and behavioural groups that do not usually buy large packs or premium brands, such as small households and middle-aged consumers (Dick, Jain et al. 1995, Ailawadi, Neslin et al. 2001, Nielsen 2016), commonly selected these products. Household panel data was not available to confirm that shoppers' simulated purchase repertoires were not representative of typical category buying patterns but this could be tested in future studies.

Finding 2: VR purchase data is consistently biased for some brands and groups of products

This study found that:

- The price of products, relative to others included in the study, had a substantial effect on the size of VR purchase biases.
- Biases were consistently evident for private label (PL) products in VR data. Across eight independent VR studies, sales of generic, store and independently branded PL products were underestimated by between 55% and 75%. Whether a product was a national brand (NB) or PL was found to be a significant factor in predicting the magnitude of biases, independent of their price.
- The VR purchase shares of popular product ranges with market shares in excess 10% were more accurate and less biased than the purchase shares of product ranges with small market shares (e.g. <5%).
- Whether a product's packaging featured a colour not commonly used in that product category, had a significant effect (p = .01) on the size of biases.

Only one of these factors, the prices of products, was previously incorporated into modelling for VR data.

Finding 3: The causes of bias are different than assumed by academic and industry users

Previous studies hypothesised that biases in shoppers' purchase choices were a consequence of consumers being unable to handle virtual products (Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016), feeling self-conscious about their purchase choices (Burke, Harlam et al. 1992, Waterlander, Jiang et al. 2015) or being unfamiliar with VR technology (Burke 1996, Cohen and Gadd 1996, Schnak, Holdershaw et al. 2016).

This study finds that:

- Fielding VR studies online, an approach that should reduce the influence of social desirability on shoppers' choices (Colombotos 1969, Krysan, Schuman et al. 1994, Holbrook, Green et al. 2003, Duffy, Smith et al. 2005, Chang and Krosnick 2010), had no effect on the size of purchase biases in the studies examined.
- Whether one or two normal shopping trips were undertaken by VR shoppers was not a significant factor in predicting the magnitude of biases and in VR studies with an additional shopping trip, biases were still evident for expensive products, PL products and uniquely coloured packs.
- The biases for larger packs and more premium products that were evident in VR data were *not* evident in real online shopping data where participants paid for the products they choose. This suggests that the lack of real purchase consequences, rather than the inability to handle or smell products, contributes to the biases observed.
- The behavioural and demographic characteristics of the VR samples are unlikely to have caused the biases for larger pack sizes, premium brands and particular product types.
- In a number of studies, there were not significant differences in the VR product and pack size choices of shoppers with different (real) incomes and household sizes, suggesting that real economic concerns and consumption rates fail to inform shoppers' VR purchase choices in experiments that are hypothetical.
- Less than 10% of shoppers pay visual attention to price and pack size information in VR studies. This suggests that financial and storage considerations do not deter shoppers from choosing larger and more premium product alternatives than they usually buy in VR studies that do not have real purchase consequences.

Finding 4: Fielding mode does not have a significant impact on the validity of VR purchase data

Regression modelling found that whether VR data is collected in specialised research facilities or via pre-recruited online research panels is not a significant predictor of the size of biases for grocery products ($p > .37$). This suggests that accurate VR data can collected online using pre-recruited consumer research panels if issues of study design (e.g. the lack of real consequences) are addressed.

Finding 5: Multiple trips are not necessary for VR purchase choice studies

While past studies recommend that VR studies be fielded with three or four practice trips to acclimatise participants to shopping in a VR purchase environment and normalise purchase quantities (Burke, Harlam et al. 1992, Burke 1996, Campo, Gijsbrechts et al. 1999), descriptive and regression analyses demonstrate that the number of trips simulated does not have a significant impact ($p > .28$) on the size of biases for grocery products. The experiments constructed had sufficient power to detect a small effect, had it been evident.

This suggests that repeated shopping trips, a procedure that can add significantly to respondent remuneration costs, particularly if data from shoppers' initial trips is not utilised in analyses, are not a good use of research funds.

Finding 6: VR simulations displayed on normal computer screens can provide valid shelf layout results

VR shopping simulations displayed on normal computer screens can predict the directional and sales impacts of reorganising products by type and location. Benchmarking found that:

- Moving products to the start of the aisle significantly increase the shares of some products (+23%). Results were consistent in magnitude with sales experiments in real stores ($Mean_{effect}$ 17%; Dreze, Hoch et al. 1995, Chung, Schmit et al. 2007, van Nierop, Fok et al. 2008).
- Moving products to eye-level significantly increased the shares of some products (*M* significant effect 36%, StD 26%), a boost similar to sales experiments testing similar changes in real stores ($Mean_{effect}$ 39%, StD = 30% N.S. different; Hubbard 1969, Dreze, Hoch et al. 1995, van Nierop, Fok et al. 2008).
- Organising the category by product type increased sales of more expensive products types (+39%).

This finding extends past studies that examined whether shoppers select items from similar areas of the shelf in real and VR studies and used more sophisticated VR equipment. Showing that that the sales effects predicted by VR are consistent in magnitude and direction with results from real stores provides a more concrete answer about the validity of VR for shelf layout testing.

Structure of this thesis

This thesis is comprised of nine chapters; a brief overview of each chapter is provided below.

Chapter 2 – Review of published findings

Chapter 2 reviews the published studies that have investigated the real-world validity of VR shopping simulations, the accuracy of VR shopping data or how VR studies should be conducted. It discusses the key contributions and limitations of these studies and identifies some gaps in the available knowledge.

Chapter 3 – Discussion of industry knowledge

Chapter 3 discusses findings from 40 in-depth interviews conducted with marketing managers and industry researchers to access unpublished knowledge about how VR studies are used by marketers and marketing researchers, how studies should be conducted, how results vary across product categories and what biases exist in VR data. These were conducted prior to developing the research questions and method of research.

Chapter 4 – Research questions

Chapter 4 outlines the research questions for this thesis. They focus on determining how accurate VR simulations are for product choice and shelf layout research, what biases consistently occur and how VR studies should be conducted to ensure the accuracy of results.

Chapter 5 – Quantitative data sources & research approach

Chapter 5 discusses how data from high-quality VR studies was obtained, collected and used to address the research questions. As data collection procedures for commercial market research (MR) studies are not typically disclosed, detailing how the VR studies used in this research were fielded, makes a substantial contribution. The data sources used to establish the real-world accuracy of choice and shelf layout information are also discussed. These are all data sources readily available to marketing managers.

Chapter 6 – Study 1: Biases for brands

Chapter 6 is the first of two chapters reporting descriptive results and identifying variables to be included in regression analysis. It reports on the realism of shoppers' brand choices, the magnitude of biases for brands and investigates which types of product categories are most affected by biases for cheaper and premium brands. It also reports on studies fielded online and in-person, and where different numbers of shopping trips were simulated and provides evidence that the samples validated were not biased.

Chapter 7 – Study 2: Biases in pack choices

Chapter 7 is the second descriptive results chapter. It reports on the product and pack sizes shoppers choose in VR simulations and the realism of shoppers' purchase choices based on their demographic and behavioural characteristics and claimed purchases. It also explores the likely causes of the biases observed for larger and brighter packs as a precursor to

regression analysis in Chapter 8 and provides evidence that sample biases were not the causes of purchase biases. Chapter 6 and Chapter 7 use data from eight VR studies conducted for packaged grocery products.

Chapter 8 – Stage 2: Confirmatory modelling

Chapter 8 investigates whether the biases identified in Chapter 6 and 7, are significant predictors of the size of biases in VR when the effects of these variables are considered simultaneously with product category covariates. It also confirms that number of trips simulated and mode of fielding do not have significant effects on the size of biases in VR studies for commonly purchased packaged grocery products.

Chapter 9 – Study 3: Trustworthiness for shelf position research

Chapter 9 is the final quantitative results chapter. It examines whether VR simulations provide results consistent with what happens in real stores when shelf layouts are changed. It utilises data from one VR study designed to test the sales effects of reorganising a packaged grocery category by product type and altering the shelf layout to improve sales of certain brands and products. It compares VR data to data from sales experiments in real stores testing similar changes and discusses the generalisability of the VR results and reports on the types of products shoppers choose. It demonstrates that VR purchase estimates may deviate substantially from in-market estimates for more expensive product types within a category e.g. convenience products, highlighting a factor that should be included in future studies.

Chapter 10 – Discussion of key findings

Chapter 9 brings together the key findings from this thesis, discusses their implications for the validity of VR for marketing research (MR) and the conduct of VR studies.

This chapter also discusses the limitations of this thesis, including the limited number of VR studies and approaches validated, and recommends areas for future research.

Appendices

Appendix 1: Lists 33 consumer goods companies that were identified as users of VR.

Appendix 2: Shows the materials used to recruit participants for exploratory and validation research and conduct interviews.

Appendix 3: Provides information about the structure of VR data files and metrics typically collected.

Appendix 4: Provides details about the data sets validated including their sampling frames, sample sizes and computer interfaces.

Appendix 5: Contains examples of the pre and post-shopping surveys used in VR studies to collect demographic and behavioural information.

Appendix 6: Provides details of preliminary analyses conducted to inform the inclusion criteria for studies. These are discussed in Chapter 5.

Appendix 7: Provides statistics e.g. *p* values, *t* values etc. for analyses presented in Chapters 6, 7, 8 and 9. These are summarised in the main body of the thesis in either tables or text.

CHAPTER 2

REVIEW OF PUBLISHED FINDINGS

Overview

Part 1 of this chapter discusses studies that have investigated the real-world validity of VR shopping simulations for shopper research and Part 2 discusses studies that have investigated how particular aspects such as technology, simulation realism, novelty and other factors affect this validity. This chapter includes studies published in industry magazines such as WARC, in marketing, health, consumer behaviour and computing journals and self-published by industry researchers to provide a complete picture of the evidence base for VR in marketing research.

Part 1: Validation research for VR

In total 16 studies in the marketing and consumer behaviour literature have investigated the real-world validity of VR shopping simulations by comparing data from VR studies to data or findings from real stores. This approach has been used establish the validity of other previously innovative research methods including Simulated Test Markets, Discrete Choice Experiments and online surveys (Silk and Urban 1978, Yeager, Krosnick et al. 2009, Mueller, Lockshin et al. 2010). Additionally, five studies have examined how real VR environments feel to consumers, an indirect measure of validity commonly used in psychology and other fields of research that utilise VR environments (Steuer 1992, Schuemie, Van Der Straaten et al. 2001, Sutcliffe, Gault et al. 2005).

Key studies and findings

The first study that examined the validity of VR shopping data tested whether a basic VR shopping simulation, an example of which is shown in Figure 3, could capture shoppers' grocery choices, product repertoires and responses to price promotions (Burke, Harlam et al. 1992). It found that VR data correlated strongly to data from real stores across three product categories (see Table 2 for results) but that the impact of price promotions on consumers' purchase choices were overestimated and the variety of products shoppers chose on real shopping trips was underestimated (Burke, Harlam et al. 1992). This study compared the VR purchase choices of 16 female participants recruited via a church mailing list to their purchase choices over 12 weeks in real stores. Since this first validation study, studies have utilised increasingly realistic VR purchase environments (see Figure 3) and investigated the validity of a range of behaviours including brand loyalty (Needel and Bean 2002), attention to price and promotional signage (Berneburg 2007, Harwood and Jones 2011, Burke and Leykin 2014) and category spend (Cohen and Gadd 1996, Higaki 2004, Waterlander, Jiang et al. 2015).

Figure 3: Examples of environments used in pre-2000 and post-2010 VR research





Images from: Burke 1996 InContext Solutions, 2012

Validation research has focused on three key areas:

(1) The ability of VR shopping simulations to capture real shoppers' brand and product choices (e.g. Cohen and Gadd 1996, Tjostheim and Haugland 2005, Waterlander, Jiang et al. 2015). This has been a key area of research because choice experiments, focus groups, panel surveys and other common methods of shopper research, have difficulty accounting for the influence of in-store factors on shoppers' choices (Burke, Harlam et al. 1992, Young 2005).

(2) If VR shopping simulations can forecast how price and promotional changes in stores will affect sales (e.g. Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016).

(3) If VR shopping simulations can predict the impact of shelf layout changes on grocery purchases (e.g. Precourt 2012, van Herpen, van den Broek et al. 2016).

Results from 16 studies are summarised in Table 1 to provide an indication of the breadth of validation research conducted for VR's use in marketing. All of these studies conducted concluded that VR was of use for at least some types of research.

Table 1: Sample of prior findings regarding the real-world validity of VR data

Study	Key results relevant to this thesis
Burke, Harlam et al. 1992	VR purchase shares within 5% points of products' real market shares for 64% established grocery brands tested; promotional responsiveness of shoppers within 5% points
Burke 1996	VR estimated predicted which shelf layout would perform best in real stores (specific results not reported)
Cohen and Gadd 1996	60% of shoppers chose the brand they usually buy (claimed) in the VR (n = 96)
Campo, Gijsbrechts et al. 1999	VR estimated propensity of shoppers to switch brands over 12 purchase occasions to within 20% and the shares of two out of five grocery brands accurately; promotional responsiveness of shoppers overestimated
Needel and Bean 2002	VR predicted the impact of assortment changes on brand and category shares and the loyalty of shoppers to particular product attributes (specific results not reported)
Higaki 2004	VR predicted the purchase shares of new products to within 1% point for 75% of brands following calibration
Bock and Treiber 2003	VR estimated the purchase shares of products in a product category not previously researched (ice-cream); specific results not reported
Tjostheim and Haugland 2005	64% of consumers selected similar products on a real and virtual shopping trip
Berneburg 2007	VR performed as well as a physical prototype in selecting which pack design and price point to implement in real stores (13% difference in conjoint scores on average)
Harwood and Jones 2011	VR provides valid data regarding shoppers' visual attention patterns in real stores although quantity of attention devoted to different elements different substantially to real store
Sharpe, Staelin et al. 2008	VR predicted accurately how changes to the range of product sizes offered would affect shoppers' purchase choices (virtual results within 1% point of real share for most pack sizes)
Precourt 2012	VR predicted that grouping products together would increase sales; validated by in market implementation, specific results not reported
Needel 2013	VR estimated the purchase shares of different brands and the impact of assortment changes on brand shares; specific results not reported
Waterlander, Jiang et al. 2015	VR can predict the share of wallet for different grocery product categories; 50% of VR shares within 25% of real shares
van Herpen, van den Broek et al. 2016	VR can identify which promotion leads to the largest consumer response and which shelves are most popular in real stores

Area 1: Product and brand choice findings

A variety of approaches have been used to determine the validity of VR purchase choice data, including comparing the real and VR purchase choices of small groups of consumers (Burke, Harlam et al. 1992, Cohen and Gadd 1996, Tjostheim and Haugland 2005, Waterlander, Jiang et al. 2015), comparing VR data to national sales data (Higaki 2004, Precourt 2012, Needel 2013) and to sales data from the same retail outlets as were simulated with VR (Campo, Gijsbrechts et al. 1999). Studies have tended to report similar metrics and results (see Table 2) enabling some generalisable findings to be established despite absence of structured replications or meta analyses. The most common validation metrics reported are (a) overall correlation between real and VR data and (b) average deviation between real and VR sales estimates.

It is important to note that the studies listed in Table 2 utilised a variety of different software, instructions and research designs. Needel (2013) for example recruited participants online via Internet research panels; most other studies recruited and executed studies in person (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999), a factor that may affect the accuracy of results. Findings regarding the impact of fielding mode are discussed in Part 2 of this chapter. While the diversity of fielding and validation approaches used in past studies strengthens the robustness of the overall conclusion that VR shopping simulations can provide accurate purchase data, it makes it difficult to draw conclusions about the accuracy of VR purchase data across product categories.

All of the studies reported in Table 2 focused on the accuracy of VR for commonly purchased grocery items such as ice-cream, beer and soft-drinks. While VR has been used to conduct shopper research for a range of other consumer products including cosmetics (Breen 2009), cigarettes (Kim, Nonnemaker et al. 2014), apparel (Papagiannidis, Pantano et al. 2013, Burke and Leykin 2014), tires (Burke 1996), fresh fruit and vegetables (Waterlander, Jiang et al. 2015), and cars (Harwood and Jones 2011), validation metrics comparable to those in Table 2 have not been reported for these categories.

Table 2: Similarity of real and VR purchase data across different product categories

Study	Product category	Correlation *r*	Av. Deviation
Needel 2013	Salad dressing	.98	-
Higaki 2004	Beer	.96*	1.4% pts
Needel 2013	Crisps	.95	-
Burke 1996	Household cleaners	.94	-
Needel 2013	Salted nuts	.94	-
Needel 2013	Laundry detergent	.94	-
Burke 1996	Health & beauty	.90	-
Needel 2013	Shampoo	.90	-
Needel 2013	Cereal	.89	-
Needel 2013	Pasta sauce	.86	-
Needel 2013	Toothpaste	.86	-
Needel 2013	Soft drinks	.84	-
Burke et al. 1992	Soft drinks	.83	3.0% pts
Desmet et al. 2013	Coffee	.81	8.1% pts
Burke et al. 1992	Paper towels	.80	8.0% pts
Needel 2013	Tea	.80	-
Burke et al. 1992	Canned tuna	.67	12.0% pts
Campo et al. 1999	Jam	.67	6.6% pts
Total	17 product categories	.86	6.5% pts

* Post-calibration result. Without calibration *r* = .79.

The similarity of products' real and VR market shares ranges between 67% and 98% across grocery products; this is a substantive difference.

Studies that have examined the correspondence between the real and virtual purchase choices of individuals (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Sharpe, Staelin et al. 2008, Burke and Leykin 2014), an approach that has been relatively rare due to the time and organisation required for participants to shop both the real and VR environments, have tended to provide more insight into what factors influence the realism of consumers' purchase choices, than studies that have compared the aggregated purchase choices of groups of consumers. In consumer research, stability at the aggregate, market level often belies variation for individuals across purchase occasions (Bass 1974, Sharp and Romaniuk 2002, Wright and Stocchi 2010), so validating the behaviours of individuals across purchase occasions is important.

In summary, studies reporting validation results for VR purchase data have found that:

- VR results are generally fairly accurate, deviating by between 1% and 10% from in-market sales estimates (average deviation, 6.5% points). This is similar to the accuracy of small scale 'mock' stores (i.e. average +/- 6% Gabor, Granger et al. 1970, Bressoud 2013, Desmet, Bordenave et al. 2013) and product lists/images in predicting the in-market performance of grocery products (average deviation typically +/-9% points; Burke, Harman et al. 1992, van Herpen, van den Broek et al. 2016). That VR performs similarly to these methods is not surprising; all three use hypothetical purchase situations and images/prototypes to understand shoppers' choices, although they differ substantially in establishment costs (Tonkin, Ouzts et al. 2011).
- The accuracy of VR purchase estimates varies across products categories (see Table 2). Most of these studies only examined two or three grocery categories. Past studies have concluded that product categories with frequent promotions (e.g. soft drinks) produce more accurate product choice estimates than product categories with less frequent price promotions (e.g. canned tuna, Burke, Harlam et al. 1992). However, analyses have failed to explain why the accuracy of VR data varies across product categories that are likely to be similar in this regard e.g. paper towels and household cleaning products (Burke 1996, Needel 2013).
- VR can accurately capture the prevalence of decisions made in-store without advance planning (i.e. impulse purchase decisions). VR studies have found that 64% of shoppers chose at least one product they had not planned to buy (Tjostheim and Haugland 2005), a similar proportion to shopping trips in real stores (60% to 70% Underhill 1999, 66% Bell, Corsten et al. 2011).
- 60% to 80% of shoppers felt that the VR environments looked like the store(s) they usually shop (Cohen and Gadd 1996, Waterlander, Scarpa et al. 2011, Waterlander, Jiang et al. 2015, Schnak, Holdershaw et al. 2016), providing substantial scope for improvement.
- Purchases of more expensive brands, including national brands and premium-quality grocery brands, are typically overestimated in VR studies by between 10 and 200%

(Campo, Gijsbrechts et al. 1999, Higaki 2004, Needel 2013, van Herpen, van den Broek et al. 2016). Similar biases have been observed in 'mock' store studies where shelves of products are erected in shopping malls or research laboratories (Nevin 1974, Desmet, Bordenave et al. 2013), suggesting they may be a consequence of research artificiality.

- Purchases of generic (i.e. no-brand), store and private label (PL) products are often underestimated in VR studies by 50% or more (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013, van Herpen, van den Broek et al. 2016), although in some categories of one study, shoppers selected more store brands than expected (Burke, Harlam et al. 1992).
- Purchases of new brands and products are often overestimated by 50% or more (Burke 1998, Bressoud 2013), similar biases have been observed in mock store studies (Klink and Smith 2001, Bressoud 2013).

These biases may be a result of differences in the cognitive processes of consumers tasked with shopping simulated environments for research purposes and the cognitive processes of consumers shopping in real stores for consumption purposes. In real stores, purchase decisions are typically characterised as involving a period of orientation, a phase of option identification, evaluation and elimination, a product selection, and for some purchases/consumers, a short period of visual and/or cognitive verification of their choice (Bettman 1979, Russo and Leclerc 1994, Senter and Wedell 1999, Shams, Wästlund et al. 2012, Shi, Wedel et al. 2013). In VR experiments it is possible that some or all of these stages differ. For example, the orientation phase may be longer because VR environments are novel (WARC 2017) and lack some of the sensorial information that is important to shoppers e.g. smell, sound, physical space (Sommer and Aitkens 1982, Davies, Kooijmanb et al. 2003, Hall, Johansson et al. 2010). These differences may mean that consumers rely more on store signage or familiar brands to orientate themselves in VR environment than in real stores (Harwood and Jones 2011). Similarly, consumers may use different criteria to evaluate and eliminate options; for example, being less influenced by price, past experience or their households' consumption requirements than in real stores, particularly if VR simulations fail to feel real or lack real purchase consequences (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Sharpe, Staelin et al. 2008). Consumers may also spend longer identifying and evaluating product options because they know they are being paid and observed by researchers (Sawyer 1975) or because they intend to select multiple items (Abdul-Muhmin 1999). These differences may make consumers more likely to notice and select new brands in VR experiments than in real stores and less likely to select cheap or undesirable brands (Nevin 1974, Norwood and Lusk 2011), biases that have been observed in past simulations.

Conversely, consumers may allow their visual attention to be more visually and less goal directed than in real stores because they know they do not have to consume the products they select (Orquin and Scholderer 2011, Burke and Leykin 2014). It is also possible that in

VR simulations without budgets or real purchase consequences, the elimination and verification phases are skipped or condensed, biasing purchase estimates and choices (Campo, Gijsbrechts et al. 1999, Burke and Leykin 2014). It is also possible for example that as recognition of PL packs is lower than for national brands, particularly where PL packaging is plain (Romaniuk 2016) and PL brands have not invested advertising (Nenycz-Thiel and Romaniuk 2014), consumers are less likely to notice these brands in VR settings if they pay less attention to products' prices (Desmet, Bordenave et al. 2013).

Tracking shoppers' eye-movements has been used to provide insight into the phases and duration of decision making in real stores (Russo and Leclerc 1994, Pieters and Warlop 1997, Van Der Lans, Pieters et al. 2008, Tonkin, Ouzts et al. 2011, Sørensen, Clement et al. 2012, Hendrickson and Ailawadi 2014). While it has not been used for this purpose in VR studies, it has been used in a number of VR studies to determine what elements of the environment attract shopper attention (Harwood and Jones 2011, Tonkin, Ouzts et al. 2011, Burke and Leykin 2014, Binge, Llinares et al. 2015).

Visual attention data can be collected in VR settings using *(1)* infrared cameras attached to a computer screen or *(2)* eye-tracking glasses worn by participants. Both methods calculate the wearers' likely point of gaze by bouncing a small beam of infrared light off the back of the wearer's right eye to determine their point of gaze. Visual attention is normally comprised of a series of short fixations in which the eye is stable and information is extracted from the scene, and rapid movements (saccades) between points of visual attention (Pieters and Warlop 1997, Wedel and Pieters 2000, Duchowski 2007). Analysis of these saccades and fixations can be used to identify when the purchase process commences, how much attention shoppers pay to price and to determine consumers' choice sets (Russo and Leclerc 1994, Shams, Wästlund et al. 2012, Lindberg, Lindström et al. 2013, Orquin and Loose 2013). Eye tracking measurements are taken many times a second, enabling shoppers' paths of gaze to be determined and decision-making process to be interpreted. It is important to note that cognitive attention does not accompany all visual attention; many tasks, including routine shopping, can be completed automatically with little or no conscious processing of information (Shiffrin and Schneider 1977, Graham 1997, Scamell-Katz 2012).

Visual attention data can also be collected by having participants click on the areas of the screen or elements of the environment that capture their conscious attention (Kleinschmit, Reid et al. 2012, Perception Research Services 2014). VR is seen as a promising addition to eye-tracking studies because it enables eye-tracking data to be collected in controlled conditions with realistic visual prompts (Harwood and Jones 2011, Tonkin, Ouzts et al. 2011, Binge, Llinares et al. 2015).

VR research utilising eye-tracking has found that consumers acquire information from virtual environments in similar ways to real stores, focusing initially on colours and shapes during orientation then progressing to focus on particular elements of the scene that are eye-catching or relevant to their search or purchase task (Tonkin, Ouzts et al. 2011, Harwood, Jones et al. 2013, Burke and Leykin 2014). However, two studies have concluded shoppers pay more attention to visually interesting elements of the environment e.g. bright colours, special shelf signage in VR simulations than in real stores (Harwood, Jones et al. 2013, Burke and Leykin 2014). No research has investigated whether the stages of decision making are the same.

VR studies have also found that very few shoppers (i.e. <10%) look at information displayed on the back or side of product packaging when selecting products (Bressoud 2013) and that very few (i.e. <9%) recall looking at the prices of products (Desmet, Bordenave et al. 2013). In real stores at least 15% of shoppers look directly at the prices of products (Lindberg, Lindström et al. 2013) and at least 50% recall accurately the price paid for grocery items, knowledge that is acquired through experience and visual attention (Dickson and Sawyer 1990, Vanhuele and Dreze 2002).

Additionally, results from three different VR studies suggest that consumers browse categories more thoroughly, spend longer shopping and consider more products than typical when shopping in real stores (Kleinschmit, Reid et al. 2012, Bressoud 2013, Desmet, Bordenave et al. 2013). One of these studies found for example that average, participants spent more than three minutes shopping for oral care products (i.e. toothpaste) and closely examined between, six and seven products (Kleinschmit, Reid et al. 2012), a level of examination rarely observed on real shopping trips where shoppers typically examine one or two products and spend between 10 and 20 seconds shopping each product category (Dickson and Sawyer 1990, Le Boutillier, Le Boutillier et al. 1994). Interestingly, studies using images of shelves projected onto a large screen also found that consumers took longer complete a familiar shopping task in a real stores (Tonkin, Ouzts et al. 2011), suggesting that the novel environment affects shopper behaviour.

The differences observed between real and VR behaviour in eye-tracking studies may provide alternative explanations for some of the biases observed in VR studies. For example, participants may select fewer generic or private label brands because the packaging of these products is often less visually interesting than the packaging of well-known national brands (Romaniuk 2016).

A limitation of past studies, including those reported in Table 2, is that most failed to consider that the accuracy of shoppers purchase choices may vary across participants. A recent study found that participants who reported strong feelings of presence while shopping a VR grocery store provided data more consistent with their real-world purchase

choices than participants who reported weaker feelings of presence in the VR simulation (15% improvement, $p < .01$ Waterlander, Jiang et al. 2015) and three shopper studies have found that younger, tertiary educated respondents report higher levels of presence in VR studies than other groups of participants (Waterlander, Scarpa et al. 2011, Waterlander, Jiang et al. 2015, Schnak, Holdershaw et al. 2016).

'Presence' is a concept used extensively in the VR literature to determine how realistic a VR environment feels to consumers (Sheridan 1992, Lombard and Ditton 1997, Slater and Wilbur 1997). Studies in human-computer interaction have found that participants' capacity for presence has more effect on their ability to successfully perform tasks in 3D VR environments than the hardware used to display and control VR environments (Roupé, Bosch-Sijtsema et al. 2014), suggesting it is an important factor to consider. However as discussed, relatively few studies in the marketing literature considered the impact of presence.

Area 2: Price change and promotional effects

The second area investigated by prior validation studies is if VR shopping simulations can accurately forecast the impact of price and quantity discounts on the purchase choices of consumers. This has been investigated by three studies (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016), two of which were designed as replication/extension studies (Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016). Despite using different validation sources, recruitment approaches and product categories, all three concluded that price promotions have more impact on shoppers' purchase choices in VR simulations than real stores (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016), even when they are not highlighted to participants with shelf tags, shading or additional text (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999).

All three studies concluded that price promotions *(1)* had 10% to 50% more impact on which products shoppers' chose, a larger increase in product sales than in real stores and *(2)* induced stockpiling effects, resulting in more multi-item purchases in VR simulations than real stores (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016). In all of three studies, stockpiling had more impact on purchase quantities than product choices. However, Burke et al. (1992) found that the sales effects of price promotions were overestimated by 10% to 20%, a fairly small error, whereas Campo, Gijsbrechts et al. (1999) found that the impact of price promotions was overestimated by between 200% and 300%. It is possible that differences in the depth, frequency or signposting of the promotions tested, brands and product categories tested or validation approaches were responsible for the difference in results between studies. Promotional depth, frequency, signposting and category effects have all been shown to influence the

responsiveness of consumers of promotions in real stores (Tellis 1988, Bijmolt, van Heerde et al. 2005).

Previous studies hypothesised that price promotions have more impact in VR environments because they are less sensorially simulating than real environments—there are no staff, children or other shoppers to contend with (Campo, Gijsbrechts et al. 1999, Breugelmans and Campo 2011, van Herpen, van den Broek et al. 2016), no smells or sounds (Burke, Harlam et al. 1992, Cohen and Gadd 1996) and participants generally shop for a smaller basket of products than they do in real stores (Burke, Harlam et al. 1992, Waterlander, Jiang et al. 2015). There is evidence from real online shopping studies to support the conclusion that price promotions have more impact in computer mediated shopping environments because they are less stimulating (Chu, Arce-Urriza et al. 2010, Breugelmans and Campo 2011, Breugelmans and Campo 2016).

However, past studies also concluded that shoppers do not consider their real usage rates, inventory requirements and capacity to store products when deciding how many products to purchase (Campo, Gijsbrechts et al. 1999). This may be because shoppers do not physically handle products, inducing large stockpiling effects in VR studies irrespective of the size of the price promotions offered.

Area 3: Shelf position research

The third area investigated by prior validation studies is whether vertical and horizontal shelf position effects are captured accurately in VR shopping simulations. This has been a key area of investigation as these effects are difficult to measure accurately without conducting experiments in real stores, something that is often difficult to achieve (Burke 1996, Wasserman 2009, Sorensen and Suher 2010).

While a number of consumer packaged goods (CPG) companies have reportedly used VR for shelf layout research (see Breen 2009, Precourt 2012), results from only three studies have been reported in the literature (Porcheddu and Venturi 2011, Massara, Porcheddu et al. 2014, van Herpen, van den Broek et al. 2016). These studies concluded in VR studies shoppers select more products from between eye and waist level and from the first section of the aisle/assortment than from other areas of the shelf/aisle (Porcheddu and Venturi 2011, van Herpen, van den Broek et al. 2016), a finding that is consistent with results from real grocery stores (Drèze, Hoch et al. 1994, Sorensen 2009, Porcheddu and Venturi 2011, Scamell-Katz 2012, Burke and Leykin 2014).

However, these studies also found that the magnitude of horizontal and vertical shelf position effects was estimated poorly by VR studies (Porcheddu and Venturi 2011, van Herpen, van den Broek et al. 2016). Both of the studies that examined this found that:

- The sales benefits of placing a product near the start of an assortment or aisle was overestimated by a VR environment. Porcheddu and Venturi (2011) found that 55% of products were selected from the start of the aisle, 22% more than expected and van Herpen, van den Broek et al. (2016) found that 53% of shoppers selected items from this area, 18% more than in real stores.
- Sales estimates for other areas of the shelf deviated substantially from estimates for real stores. However, results were not consistent across studies. In van Herpen, van den Broek et al. (2016), a study that used multiple screens to display the VR environment, 12% more shoppers than expected selecting items from the top shelf. The same bias was not evident in Porcheddu and Venturi (2011) which used a head mounted VR display. In Porcheddu and Venturi (2011), fewer shoppers selected items from the top shelf than expected, favouring products on eye and waist level shelves. How the technology used in these studies may have contributed to the differences observed is discussed on page 39.

Part 2: Methods research for VR

Over the last 25 years a small body of research (i.e. eight studies) has investigated the technology and research design necessary to collect valid VR shopper data. Additionally, research in the fields of human-computer interaction and psychology utilising different VR approaches has focused on tasks of relevance to marketing including navigation and comprehension.[1] Research has focused on four key elements: *(1)* the visual realism of VR environment, *(2)* the hardware used, *(3)* the instructions given to participants, and (4) the novelty of VR environments. As most of these studies have had small samples (e.g. $n = 30$), this Part 2 focuses on findings consistent in multiple studies.

Area 1: Visual realism of VR environments

Five studies have examined how the realism of VR environments contributes to the accuracy of VR data using a variety of approaches including questioning participants about how realistic different elements of the simulation felt and comparing data from experiments with different levels of realism (Burke, Harlam et al. 1992, Berneburg 2007, Waterlander, Jiang et al. 2015, van Herpen, van den Broek et al. 2016).

Studies that asked participants to evaluate the realism of the simulation used a variety of approaches including:

- Asking participants "imagine they were doing their real-life grocery shopping" in the VR environment (Burke, Harlam et al. 1992).

[1]This is key as the benefits of different visualisation and interaction technologies to VR data has been shown to be task-specific. For example, using large screens improves the ability of subjects to complete navigation tasks quickly and accurately in 3D VR environments but does not improve performance of comprehension tasks (Ni, Bowman and Chen, 2006).

- Asking participants if the VR was "similar to their usual shopping experience" (Stone and Cogdon, 2013).
- Administering the Witmer and Singer (2002) Presence and Immersive Tendencies questionnaire (Berneburg 2007, van Herpen, van den Broek et al. 2016, Waterlander, Jiang et al. 2015).

Generally, these studies have found that VR environments that look more realistic (Burke, Harlam et al. 1992, Berneburg 2007, van Herpen, van den Broek et al. 2016) or feel more realistic to participants (Waterlander, Jiang et al. 2015, Stone and Cogdon, 2013), provide more accurate data. However, the accuracy of consumer self reports of their purchases is questionable, as is the validity of the Witmer and Singer immersion questionnaire. A recent study by van Herpen, van den Broek et al. (2016) reported that participants felt more immersed in the virtual environment than a real environment, something that should not be possible given the scale constructs. The Witmer and Singer questionnaire has also been criticised by other researchers for poor construction (Usoh, Catena et al. 2000).

Most commonly studies investigating how the visual realism of VR environments affects the validity of results have investigated whether making products look three-dimensional (3D) improves the accuracy of results (Li, Daugherty et al. 2003, Berneburg 2007, Kleinschmit, Reid et al. 2012, van Herpen, van den Broek et al. 2016).

A number of other studies tested different approaches did not attempt to pinpoint which elements of the environment made the VR appear more realistic; a major limitation as not all elements contribute to how realistic a simulation feels. Background noises for example have been shown to not affect how realistic VR environments *feel* to participants (Schnack, Holdershaw et al. 2016) but spatial control has been shown to affect perceptions of realism substantially (Stone and Cogdon, 2013).

Figure 4 provides examples of how VR simulations look when products look two-dimensional (2D) and 3D. Making products look 3D requires more time and technical expertise as products need to be photographed from the front, back and sides, and photographs linked together so that packages appear to have depth when displayed on screen (van Herpen, van den Broek et al. 2016). Making products look 2D is much more straightforward and requires only a front-on image of each product (Berneburg 2007). Often these can be sourced from product image libraries such as www.dlibrary.com.au.

Four studies have compared results from similar 2D and 3D VR shopping environments (Li, Daugherty et al. 2003, Berneburg 2007, Kleinschmit, Reid et al. 2012, van Herpen, van den Broek et al. 2016). While most of these investigated different metrics, all four studies concluded that 3D rendering improved the validity of results to at least some degree. Specific results from these studies are summarised in Table 3.

Table 3: Findings regarding the value of 3D visualisation in VR studies

Study	Key findings
Li, Daugherty et al. 2003	20% improvement in shoppers' knowledge of products when shown 3D rather than 2D prototypes and 16% improvement in consumers' confidence in their purchase decisions; results consistent across three products categories.
Berneburg 2007	50% improvement in predictions regarding which pack would perform best in real stores; estimates of the importance of packaging and price on shoppers' purchase choices 38% and 13% more similar with real data.
Kleinschmit, Reid et al. 2012	3D visualisation increased shopper engagement with the environment by 40 - 66% over 2D visualisation. Did not affect which items shoppers chose by a statistically significant amount but small differences may have been evident.
van Herpen, van den Broek et al. 2016	Estimates of category spend, purchase quantity and popular shelf locations more similar to physical store when VR 3D rather than 2D. Incidence of national and PL brand buying 18% more accurate in 2D than 3D environment.
Average impact	**31% impact (StD 21%)**

Making products looks 3D improves the accuracy of VR data by 31% on average in VR shopping studies.

The results in Table 3 indicate that that 3D rendering affects the validity of VR data by approximately 30% although the magnitude of the effect varies substantially across studies and metrics (StD 21%, Li, Daugherty et al. 2003, Berneburg 2007, Kleinschmit, Reid et al. 2012, van Herpen, van den Broek et al. 2016). It is important to note that some studies such as van Herpen, van den Broek et al. (2016) varied factors other than the factors other than the level of rendering between executions.

The studies reported in Table 3 suggest that 3D rendering affects the accuracy of pack design metrics more than price-related metrics (Li, Daugherty et al. 2003, Berneburg 2007), an unsurprising finding given that packages for grocery items look more different than price tags when 3D (see Figure 5). 3D rendering may also be important for conveying information about the price and prestige of products, other factors shoppers use in decision making (Allenby, Fennell et al. 2005).

Figure 4: Examples of 2D and 3D product images



The 3D packages on the right look more realistic but take more time to program.

Figure 5: An example of how price tags were displayed in 3D VR projects



Packages look more different than price tags when 3D.

A key area of contention is whether 3D rendering improves or even influences the validity of shoppers' VR product choices. Kleinschmit, Rizzo et al. (2012), an industry study detailed in Table 3 reported minimal differences in brand and SKU choices between a 2D and 3D versions of a VR shelf (n = 1800). In contrast, an academic study with a smaller sample (n = 90), found that 3D visualisation increased the proportion of shoppers selecting expensive national brands and reduced the proportion selecting private label products by almost 18% (van Herpen, van den Broek et al. 2016). These findings suggest more research is needed.

Interestingly, the results in Table 3 are directionally consistent with findings from the human-computer interaction literature regarding the value of 3D visualisation, although these concluded that the display hardware used (e.g. HMD, normal computer screen, high resolution large screen) has twice as much impact on the ability of participants to perform and recall movements in 3D environments than whether the environment was 2D or 3D (Slater, Linakis et al. 1996, Tan, Gergle et al. 2004, Tyndiuk, Lespinet-Najib et al. 2004). Generally, these studies had better experimental design/controls than studies published in the marketing literature such as Kleinschmit, Rizzo et al. (2012), which varied multiple factors across executions.

Area 2: Hardware used to field VR studies and fielding mode

Another element that may affect the realism of VR data that has been researched is the hardware used to display and interact with VR simulations. To date three types of hardware have been used to field VR in the literature: *(1)* normal computer screens (Burke, Harlam et al. 1992, Burke and Leykin 2014, van Herpen, van den Broek et al. 2016, Waterlander, Jiang et al. 2015), *(2)* large curved screens that provide a 180° field-of-view (Stone and Cogdon, 2013, Binge, LIinares et al. 2015, van Herpen, van den Broek et al. 2016) and *(3)* wearable head-mounted displays (HMD) that fill shoppers' field of view (Porcheddu and Venturi 2011). Images of each type are shown in Figure 6 and Figure 7.

Research in other fields has shown that multi-screen setups and HMDs elicit stronger emotional and physiological responses from participants than normal computer screens (Sutcliffe, Gault et al. 2005, Lee, Kim et al. 2013, Grinberg, Careaga et al. 2014), presumably because they create a stronger sense of 'presence' (Slater, Linakis et al. 1996, Tyndiuk, Lespinet-Najib et al. 2004, Chen, Sun et al. 2009). However, research in marketing has been limited to seeing which combinations of hardware and software when varied together provide more accurate purchase data or presence (see for example, Van Herpen, van den Broek et al. 2016, Schnak, Holdershaw et al. 2016).

For example, a recent study compared the accuracy of shelf position, promotional lift and product choice results collected using one computer screen or multiple computer screens (van Herpen, van den Broek et al. 2016). This study concluded that the multi-screen system (see Figure 6) provided shelf position, purchase quantity and purchase choice data that was 30% more accurate. However, this study varied the functionality of the software and how products looked (2D, 3D) across the two VR versions tested. Similarly, a recent study concluded that shoppers reported greater feelings of presence, engagement and shopping ease when experiencing a VR shopping simulation on head-mounted display (HMD) compared to a normal computer screen (Schnak, Holdershaw et al. 2016) but also varied the software used across the technology combinations tested. Marketing studies that have been stricter in their research design (e.g. Stone and Cogdon, 2013) have found that wide screens make it easier for consumers to shop large assortments quickly.

Figure 6: Examples of how single and multi-screen setups compare



Figure 7: Example of head-mounted VR displays and the view of participants



Studies in other fields have found that consumers engage with screens, even large wall-size screens, differently to real scenes, focusing on different areas and elements of the environment than in real physical environments (Mannan, Ruddock et al. 1997, Parkhurst, Law et al. 2002, Parkhurst and Niebur 2003, Bindemann 2010). Research in marketing has confirmed this (see for example van Herpen, van den Broek et al. 2016).

Evidence suggests that head mounted displays provide more realistic positional data. A 2011 study utilising a basic HMD VR environment (Porcheddu and Venturi, 2011), found that 93% of shoppers selected an item from shelf positions at eye and hand level, a larger proportion than in screen-based studies (Raghubir and Valenzuela 2006, Atalay, Bodur et al. 2012, van Herpen, van den Broek et al. 2016 60%).[2] This may be because head mounted displays replicate the physical requirements of selecting items from top or bottom shelves more accurately than computer screens.

In the computing literature, interaction devices tend to be classified as either "window, icon, menu, pointing device" (WIMP), a term that encompasses the interaction devices with which we are most familiar e.g. keyboards, mice, touch screens etc., or "advanced devices",

[2] These studies seem to have biases towards the selection of products from the upper and lower shelves.

a term that includes a variety of different interaction technologies including motion sensors, haptic feedback gloves and eye-movement controls (Roupé, Bosch-Sijtsema et al. 2014). Most of the VR studies reported in the literature including Burke, Harlam et al. (1992), Desmet, Bordenave et al. (2013) and van Herpen, van den Broek et al. (2016) have used WIMP devices such as keyboards and mice.

Throughout the 1990s and 2000s VR studies were fielded using specialised hardware designed to mimic the motion of pushing a shopping trolley (see Figure 6 for example, from Breen, 2009). While no studies demonstrating the value of this approach were published, there are likely to be significant advantages of this approach over WIMP technology (Tal and Wansink 2011). Studies in other fields have demonstrated that head mounted displays with motion-tracking help participants navigate 3D VR environments twice as fast as hand-driven steering devices such as joysticks and expend less cognitive effort (Roupé, Bosch-Sijtsema et al. 2014), perhaps because they are more reflective of real-world interaction modes and movements. Similar approaches could have been used in marketing, although the transition towards fielding VR studies over the Internet to pre-recruited groups of consumers (Bock and Treiber 2003, Breen 2009) is likely to have limited their usage.

Fielding studies online enabled VR studies to be conducted with larger samples (Advertising Research Foundation 2009, Needel 2013) and to be integrated into online surveys, a popular mode of market research (GreenBook 2015). Three studies have reported data from VR studies fielded online (Higaki 2004, Needel 2013, Waterlander, Jiang et al. 2015), although other studies are likely to have been conducted by industry researchers given the reported popularity of the method (Breen 2009, Greenbook 2015). One of these studies, Higaki 2004, concluded that accurate purchase forecasts for new and established products could be collected online using consumers' WIMP devices (r = .96, mean deviation 1% point, see Table 2 for results from in-person studies). However, there are some indications that either the display and interaction interface or the mode of fielding (i.e. online with no personal contact between researchers and participants) influences the data collected.

Needel (2013) for example found that session times were substantially longer in VR simulations fielded online than in studies fielded in specialised facilities (M_{time} = 4 minutes vs 1 minute 40 seconds) across a range of grocery categories. Longer session times is consistent with the effects of small screens and less institutive interaction devices, both of which increase the time required to view and navigate VR environments (Slater, Linakis et al. 1996, Tyndiuk, Lespinet-Najib et al. 2004, Stone and Congdon 2007, Roupé, Bosch-Sijtsema et al. 2014). It is also possible that consumers' home internet connections were slower than those used in the in person facilities, that the videos used to instruct participants how to navigate the VR and select products are less helpful than the tailored demonstrations given during in-person studies or that unsupervised consumers may treat a research task, particularly a VR one, as 'entertainment' and select too many products, a

concern raised by a number of researchers (Desmet, Bordenave et al. 2013, Papagiannidis, Pantano et al. 2013, Perception Research Services 2014). Computer lag time may also contribute to biases, making products time consuming to compare (Roupé, Bosch-Sijtsema et al. 2014) and resulting in most participants (i.e. 85%) selected fewer total items than they normally buy on a real shopping trip (Waterlander, Jiang et al. 2015), a result that is highly unusual for VR studies (Burke 1996, Campo, Gijsbrechts et al. 1999, Higaki 2004).

Area 3: Instructions given to participants & research protocols

A number of studies in the social sciences have established that the wording of participant and task instructions can impact the validity of experimental data (Krosnick 1999, Aadland and Caplan 2003, Groves and Lyberg 2010). VR studies also have provided anecdotal evidence that the instructions given to participants affect the validity of VR data. For example, Campo, Gijbrechts et al. (1999) concluded based on point comparisons to a previous study (Burke, Harlam et al. 1992) that whether participants are shown how to switch brands in the introduction influences the realism of brand switching metrics by up to 40%. Similarly, early studies reported that through trial and error they found that consumers behaved more naturally when told they were trialing a new home shopping service and researchers' interests in their purchase choices were not mentioned (Burke 1996, Cohen and Gadd 1996). However, none of the VR studies reported in the literature since 2000 employed this tactic and most of these had good overall accuracy (see Table 2).

Past studies have consistently found that assigning budgets and specific shopping tasks diminishes the realism of VR purchase data, even when they are tailored to the self-reported purchase quantities of shoppers (Burke, Harlam et al. 1992, Cohen and Gadd 1996, Campo, Gijsbrechts et al. 1999, Waterlander, Scarpa et al. 2011). Most of the studies reported in the literature since 2000, simply asked participants to "shop as they would normally shop" (Bock and Treiber 2003) or "make all their planned purchases for the week" (Waterlander, Jiang et al. 2015). Past studies do not appear to have experimented with applying time pressure, a factor that impacts shopper behaviour in real stores (Pieters and Warlop 1999, Silayoi and Speece 2004, Reutskaja, Nagel et al. 2011), suggesting that research into the protocols of VR studies is limited.

Most of the studies reported in the literature have provided participants with a small monetary incentive for participation (e.g. €7.50, NZ$20 or a donation to their favorite charity), a factor that may have affected shopper's purchase choices (Sawyer, Worthing et al. 1979). No reported studies have experimented with having consumers pay for or actually receive the products they choose and none made shopper's purchase choices their incentive, a tactic commonly employed in other methods of research to improve the realism of consumers' purchase choices (Lusk and Schroeder 2004, Norwood and Lusk 2011). In fact, some recent VR studies excluded from analyses shoppers who indicated they tried to

arrive at the value of their incentive when selecting products (van Herpen, van den Broek et al. 2016). This study did not report what impact this had on the validity of results, suggesting again that research into the protocols of VR studies is limited.

Studies testing other aspects of the research protocols not been reported in the literature, but it is possible they have been conducted by industry researchers. When the studies in Table 2 are arranged by date, a weak relationship (Pearsons r = .43) is evident between the year of publication and the accuracy of VR purchase data, suggesting that researchers have learnt how to conduct VR studies more effectively since the first of these studies was published in 1992. However, improvements in the visual realism of VR simulations (see Figure 3) may also be responsible for the improvement in results.

Area 4: Novelty of VR environments

The fourth area investigated previously is how the novelty of VR as a research tool affects the validity of VR shopping data. Findings from the HCI literature have demonstrated that practice sessions have a strong impact on the ability to consumers to accurately navigate computer-mediated environments (Slater, Linakis et al. 1996). In a similar vein, a number of studies in the marketing literature hypothesised that consumers familiar with VR from playing video games may be *less* likely to spend too much time exploring VR environments or select too many products (Campo, Gijsbrechts et al. 1999, Waterlander, Scarpa et al. 2011, Papagiannidis, Pantano et al. 2013, Schnak, Holdershaw et al. 2016). However, this theory has been tested and the results indicate that experience with VR in an entertainment context does not make participants good shopper research subjects. Shoppers who indicated they play video games once a week or more were *less* likely to report trying to choose the products they normally buy in the VR grocery shopping study than shoppers who do not play video games (16% reduction in self-reported accuracy, stat. sig. $p < .10$ Leister, Tjostheim et al. 2007).

The second approach tested to reduce the novelty of VR is repeated exposure; an approach consistent with the findings of the HCI literature regarding the value of practice for computer-mediated environments (Slater, Linakis et al. 1996, Sparto, Whitney et al. 2004, Zygourisa, Giakoumisb et al. 2014). Marketing studies have consistently found that having shoppers repeat their trips through a VR environment, halves the amount of time participants spend shopping for products (Kleinschmit, Reid et al. 2012) and how many products shoppers select by between 6% and 33%, (Burke, Harlam et al. 1992, Burke 1996, Campo, Gijsbrechts et al. 1999, Kleinschmit, Reid et al. 2012). However, some of these studies have found that simulating eight shopping trips in a single session without the benefit of consumption experiences between reduces the variety different brands shoppers normally select across purchase occasions by 42% or more (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999). Past studies have tended to simulate between three and 12

shopping trips in a single research session (Burke, Harlam et al. 1992, Burke 1996, Campo, Gijsbrechts et al. 1999, Kleinschmit, Reid et al. 2012).

That additional shopping trips improve the realism of VR purchase data is not surprising—observational research has found that shoppers require one or two sessions to learn how to navigate even basic online shopping platforms (Anesbury, Nenycz-Thiel et al. 2016), a technology that is likely to be more familiar to most shoppers in its structure than computer-mediated 3D VR environments. Based on the improvements evident in purchase quantity and session time metrics when data from multiple trips is compared, researchers have hypothesised that multiple trips are necessary are necessity to reduce the novelty of VR and get participants to shop with the same low level of attention to the shelf this is characteristic of real shopping trips (Burke 1996, Tal and Wansink 2011).

More recent studies have continued to assign participants multiple shopping trips. However, in these studies data from all of the trips conducted was pooled (Burke and Leykin 2014, Waterlander, Jiang et al. 2015), suggesting that multiple trips were conducted to enlarge the sample and rather than normalise purchase metrics. No studies reported in the literature have compared visual, cognitive or product attention data from successive shopping trips.

The third way in which the novelty of the VR environment may affect the accuracy of VR data is via the range of products and number of categories that participants shop. This factor has not been tested in past studies but its impact is evident when results from studies reported in the literature are compared directly (see Table 4).

Table 4: Amount of time spent shopping and size of VR simulations

Study	Product category(s)	VR environment	Number of products	Shopping time *Mean*
(Kleinschmit, Reid et al. 2012)	Toothpaste & mouthwash	Portion of an aisle	66	216 seconds
(Bressoud 2013)	Breakfast cereal	One aisle	168	86 seconds
(Desmet, Bordenave et al. 2013)	Ground coffee	One aisle	182	34 seconds
	Average	One category	134	112 seconds
(Burke and Leykin 2014)	Range of packaged grocery products including cereal bars	Multi-category, Small grocery store	2,000+	12 seconds

In VR studies where only one product category was shown, consumers spent more time shopping the product selection for each category than in VR studies with lots of product categories.

In studies where participants shop a small range of products (i.e. < 200) they seem to spend longer considering each purchase (i.e. Table 4: 34 seconds vs 12 seconds). They also select more products from each product category than expected (Cohen and Gadd 1996) and examine in more detail the range of products shown, picking up and examining for example six different oral care products (Kleinschmit, Reid et al. 2012). Thus, it seems that the realism of browsing and shopping time metrics are influenced by the scale of the VR simulations. Figure 8 provides an example of a VR simulation in which with multiple categories could be shopped.

Results from two multi-category studies reported in the literature—Burke and Leykin (2014) and Waterlander, Jiang et al. (2015)—suggest that when VR simulations include a large range of products from multiple product categories, purchase rate and shopping time data is more likely to resemble data from real stores. In real stores shoppers typically spend less than 20 seconds shopping for grocery products; Mean time (*M*) = 12 seconds, majority of purchases between nine and 18 seconds (Hoyer 1984, Cobb and Hoyer 1986, Dickson and Sawyer 1990, Le Boutillier, Le Boutillier et al. 1994). This is not significantly different to the results for Burke and Leykin (2014) cited in Table 4. However, it is not clear showing multiple categories improves the accuracy of purchase choice data as studies reporting this metric have not been published.

Figure 8: Example of VR simulation where multiple categories could be shopped

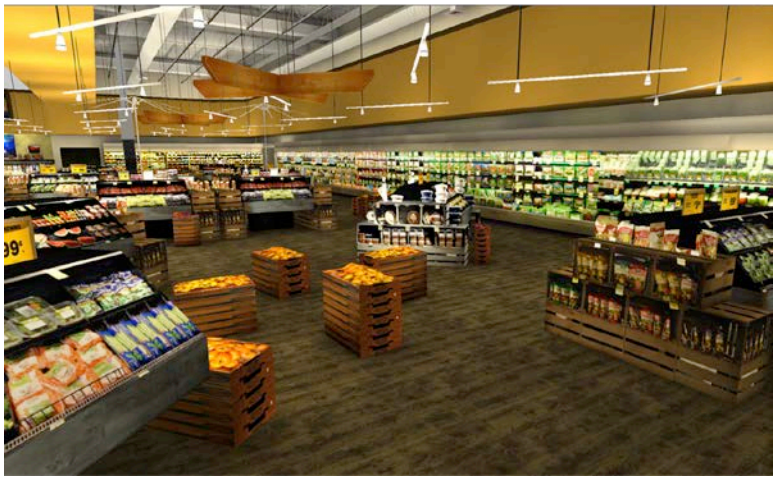


Image from: InContext Solutions

In this simulation participants could shop for many products, similar to Waterlander, Jiang et al. (2015).

However, simulating multiple categories is likely to add significantly to the time and cost of a VR study and most of the studies reported in Table 2 with high correlations between real and VR data, restricted shopping to a single product category.

Chapter Summary

Finding 1 from the literature: VR data has some known biases

While prior studies have generally concluded that VR is a useful research tool that produces valid shopper data, prior studies have found real and VR data to differ in a number of key ways, including:

- Shoppers focused on and selected items from different areas of the store shelf in VR simulations than real stores, for example, select more products from the start of the aisle in VR studies than in real stores (Porcheddu and Venturi 2011, van Herpen, van den Broek et al. 2016).
- Shoppers selected more expensive national brands in VR studies and fewer private label (store) brands than on real shopping trips (Campo, Gijsbrechts et al. 1999, Higaki 2004, Needel 2013, van Herpen, van den Broek et al. 2016). In some studies these biases were substantial with demand for expensive grocery brands being overestimated by between 45% and 300% (Campo, Gijsbrechts et al. 1999, Higaki 2004, Needel 2013) and demand for private label brands being underestimated by between 3% and 60% (Campo, Gijsbrechts et al. 1999, Higaki 2004, Needel 2013, van Herpen, van den Broek et al. 2016).
- Participants considered, examined and selected a greater number of products from each category than they usually do in real stores (Burke, Harlam et al. 1992, Cohen and Gadd

1996, Cohen and Gadd 1996, Klienschmit, Reid et al. 2012, Waterlander, Jiang et al. 2015).

- Consumers spent longer shopping for commonly purchased grocery products than is realistic (Kleinschmit, Reid et al. 2012, Bressoud 2013, Desmet, Bordenave et al. 2013, Needel 2013, Burke and Leykin 2014). This bias was particularly pronounced in VR studies that focused on a small number of product categories (Kleinschmit, Reid et al. 2012, Bressoud 2013, Desmet, Bordenave et al. 2013) and did not expose participants to the VR environment before collecting the data used in analyses (Kleinschmit, Reid et al. 2012, Needel 2013, Burke and Leykin 2014).
- Temporary price promotions have substantially more impact on shoppers' product choices and purchase quantities than in real stores (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999).
- Some of the approaches used in past studies to establish the validity of VR data, such as the Witmer and Singer immersion questionnaire may be problematic Usoh, Catena et al. 2000, van Herpen, van den Broek et al. 2016).

All of these biases suggest that further refinement of the method and more research is required.

Finding 2 from the literature: Limited range of product categories validated and replication studiers

The second finding evident from the literature is that a very limited number of replication studies or studies reporting comparable results have been reported (Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016). The absence of well designed, close replication studies has made it difficult to develop generalisable findings (Lindsay and Ehrenberg 1993, Hubbard and Armstrong 1994, Hunter 2001, Uncles and Wright 2004, Uncles 2011).

Furthermore, common validation metrics have only been reported for a limited range of grocery categories (see Table 2), making it impossible to develop generalisable benchmarks.

Finding 3 from the literature: Well-designed research lacking in some areas

Incidental methods research has demonstrated that a number of factors affect the validity of VR data including whether 3D product visuals are used (Li, Daugherty et al. 2003, Berneberg 2007, van Herpen, van den Broek et al. 2016, et al.), the instructions given to participants (Campo, Gijsbrechts et al. 1999) and the respondents recruited (Waterlander, Scarpa et al. 2011, Schnack, Holdershaw et al. 2016, Leister, Tjosthem et al. 2007). This is lacking in a number of areas including what hardware required for pack testing, visual

attention and price/promotional studies, whether 3D visuals improve the accuracy of shelf and/or category placement research and whether more intuitive controls improve the accuracy of purchase choice data. While findings from the Human-Computer Interaction literature and other areas of marketing research can corroborate findings and aid hypothesis development, well designed, task specific studies are essential (Ni, Bowman and Chen, 2006).

Industry users may have investigated some of these issues, having reportedly used VR for a great variety of shopper research (Treiber and Needel 2000, Breen 2009). The following chapter addresses this gap by interviewing industry practitioners that have used VR for shopper research.

CHAPTER *3*

INVESTIGATION OF INDUSTRY KNOWLEDGE

Overview

This chapter discusses the findings from 40 in-depth interviews conducted with marketing managers and market researchers to access their knowledge. This chapter begins by discussing how respondents were selected and interviews conducted.

Qualitative research: Approach

How access to industry knowledge was obtained

Chapter 2 identified that in addition to academic researchers, knowledge about the validity of VR simulations for shopper research has come from three groups:

(1) Marketing managers responsible for commissioning VR research.

(2) Market researchers (MR) who have fielded VR studies on behalf of clients.

(3) Industry experts who have evaluated the method or advised CPG clients on investment decisions for VR.

Over eighty individuals from these groups were identified in published reports, company press releases, LinkedIn, presentations, blogs, websites and personal networking. The companies identified as users of VR are listed in Appendix 1. Individuals from each of these companies were contacted using email and LinkedIn with personalised messages and the flyer in Appendix 2 which explains the objectives of the project. TNS Kantar Retail also assisted in distributing this flyer to CPG companies. A third of the individuals contacted did not respond (*n* = 30, 36%) and 13% (*n* = 10) declined to participate. Additional follow-up emails and the endorsement of senior managers within their companies were typically required before individuals agreed to participate.

Non-response rates were highest amongst industry experts (70%), particularly those employed by universities. This may be because we were unable to secure personal introductions to this group or because of the high volume of email most academics receive requesting their expertise. Marketing managers had the highest response rate (61%), followed by market researchers (44%) as I was able to leverage my professional networks to secure email introductions to most of the targeted individuals. The marketing managers that declined to be interviewed provided a range of reasons, including: their company had not used VR recently (n = 2), their company had ceased using VR (*n* = 1) or they had not personally executed a VR project and had no relevant knowledge to contribute (*n* = 2). The

market researchers that declined to be interviewed generally cited confidentiality concerns (i.e. not comfortable discussing findings from client projects, unable to because of confidentiality agreements) or felt that reports already in the public domain they had contributed to (i.e. Shaping Retail) were sufficient to address concerns about validity.

The job titles and number of years of experience with VR of participants are outlined in Appendix 2. Forty interviews was sufficient to reach 'saturation', a criteria commonly used to establish the adequacy of qualitative samples (Charmaz 2006, Bowen 2008, Corbin and Strauss 2008).

How interviews were conducted

Collectively individuals from 23 different CPG companies and market research agencies were interviewed. It was necessary to interview multiple individuals from organisations because of their multi-national, multi-category structure. Where possible (n = 10), individuals from the same organisation were typically at the same time to facilitate the contrasting of experiences.

Interviews were conducted by the PhD candidate, an experienced qualitative researcher with training from the Australian Market and Social Research Society (AMSRS) and the University of South Australia. Interviews were typically 60 to 75 minutes long and were conducted in person (n = 7), by telephone/video-chat (n = 29) and via email (n = 4). Remote methods were used as most respondents live outside Australia.

Interviews were supervised by two experienced qualitative researchers, the ethics committee of the University of South Australia and the Ehrenberg-Bass Institute, a certified market research agency (ISO 20252 certification). The project information sheet provided to participants is detailed in Appendix 2. A discussion guide was developed through eight preliminary interviews conducted between October 2011 and June 2012 (see Appendix 2). Only questions relevant to each users' experiences and responsibilities were asked.

The managers and researchers interviewed were typically experienced with VR, having used it for between one and 22 years (M = 5 years, StD = 4). Twenty one were marketing managers (53% of sample), 16 were market researchers (40% of sample) and the remainder were industry experts (n = 3). The sample included more marketers from the US and UK as VR is more common in these regions. Collectively 20% of the sample was from Continental Europe, Australia and Asia, and three of the marketers interviewed were responsible for shopper research globally.

Verbatim quotes from interviews are used through this chapter to illuminate findings with the identity of participants disguised.

Qualitative research: Findings

Product categories and issues industry users have researched with VR

Collectively, the marketers and market researchers interviewed have used VR to research packaged grocery products including soft drinks and confectionery, stationery, household cleaning products, health and beauty products, alcohol, non-prescription medicines and pet care items. They have used VR to research a variety of marketing issues including special displays and signage at shelf, changes to shelf layouts, product assortments, packaging materials, pack graphics and prices.

> *We do a VR test for virtually every pack change to check that if you change the pack, are people still going to buy? It's a good disaster check.*
>
> *US Marketing Manager*

> *VR is a tried and true method for understanding the impact of category or pack changes on sales, reach and volume.*
>
> VP global research agency

Amongst the group sampled, VR simulations were most commonly used for pack design, product choice and shelf layout research. The majority of the companies interviewed have used it for A/B testing where new pack designs, product assortment or shelf layouts are tested against other designs and what is currently available in real stores. In this context it has been used primarily to determine *if* the new pack designs, product assortments or shelf layouts were likely to increase sales and *which* layout or design was most likely to increase brand and/or category sales. It was not typically used estimate how *much* impact a pack, assortment or layout change would have in real stores.

> *I would never tell a retailer you can make X amount of money by implementing a change we tested with virtual.*
>
> US category manager

VR users commonly believed this based on the inaccuracies in impact estimates reported in Burke, Harlam et al. (1992) and Campo, Gijsbrechts et al. (1999) and their own experiences some of which are discussed on page 52.

Discussions identified that while most VR studies include pricing information (see for example Image 5), VR was not commonly used by marketers in industry for pricing research for two reasons. Firstly, temporary price changes were often not tested prior to introduction (Szabo, Bogomolova et al. 2012) and secondly, many industry users believed VR will not provide accurate price elasticity results because shoppers do not spend their own money

and because of the differences reported in past studies comparing VR price promotions to tests in real stores (e.g. Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999).

> *Any method where consumers don't spend real money…you cannot expect to get it [price impact estimates] right.*
>
> Insights Manager, Global position

Participants reported using VR mainly to test changes to shelves, layouts and products in categories. VR simulations were not used to test how changes to the location of products or product displays outside the main aisle (i.e., seasonal displays or tie-in displays in other parts of the store) will affect shopping and buying due to the difficulty of constructing realistic scenarios that involve shopping the whole store (see Cohen and Gadd 1996 for further discussion) and the cost of fielding VR studies with multiple tasks. For example:

> *We told people to shop a whole store and tests would run for 50 minutes …It cost us a fortune in respondent time.*
>
> Insights Manager, Global position

> *It's difficult to construct a scenario that involves the whole store and capture the things people do normally. For example, seeing a product on the end-cap and then purchasing it from the shelf. We focus on the category instead.*
>
> Industry Expert, UK role

Validation research industry users have conducted

Despite being responsible for over 40 VR shopper projects, only nine of the industry users interviewed had attempted to ascertain the real-world validity of their data. They reported doing so using a variety of methods, including:

- Comparing VR sales data to sales data from real stores (*n* = 2*)*. These studies found that results for most brands were "pretty good" i.e. within +/-10%, but that some differences were evident. These are discussed on page 54. One industry expert reported conducting similar research and reaching similar conclusions.
- Comparing VR data to the results of in-store tests of trials (*n* = 4*)*. These studies all found that VR correctly predicted how changes would affect the shares of brands but did not predict accurately how much impact the change had in real physical stores (e.g. volume change or sales lift). Two of these were shelf layout studies; one was a range reduction study and one was for a new product. Managers concluded that:

I don't trust virtual to estimate how high is high. For that you need a controlled in-store test.

US category manager

For the new product study, sales estimates were within 25% of real sales, but this study estimated sales from a small proportion of the sample, those with positive purchase intention.

- One company had conducted a parallel study where participants shopped a VR convenience store followed by a real convenience store with a similar layout. They found that shoppers spent a similar amount of time in both stores, visited similar areas and looked at a similar number of products in the store (e.g. 20) but that shoppers selected more items in the VR simulation (M_{items} = 3.8, 20% more than in real store, statistical significance not reported) and spent longer shopping the VR store than the real convenience store. Additionally, they found that the additional items shoppers selected came primarily from 'unhealthy' product categories such as confectionery and soft drinks, a finding that is consistent with the results of one published study (Waterlander, Scarpa et al. 2011).
- One company compared data from a VR study to observational data from real stores for the same product category. They found that participants spent longer shopping (e.g. 3+ minutes), handled more and selected more products from the category in the VR simulation (M_{items} = 4.5) than in real stores (M_{items} = 1.6, StDs not reported).
- A marketing manager in China compared the results of a VR study to consumer panel data and found that some important findings regarding shopper loyalty differed. The VR study concluded that consumers were very brand loyal and consistently bought the same brands and types of products, whereas purchase panel data showed that most shoppers switch randomly between a range of different brands and product types.

In addition to these studies, three respondents mentioned that they had conducted informal checks to confirm that the VR shares for "their products were about right" but this type of research is likely to miss systematic biases in the shares of products as most product categories are highly fragmented with each product having very small market shares.

A number of marketing managers who had not conducted validation studies expressed a desire for "more proof that VR really works" but had different ideas regarding what metrics should be validated and how validation studies should be conducted. While some felt that point estimates from VR data should be validated to scanner data from real stores, others felt that it was more important to confirm VR helped them make the right decision" or that "shoppers navigated the store the same way and choose the same brands as they normally buy".

A number of marketers expressed they had found it difficult to validate VR data using data from in-store trials, the method recommended by Burke (1996) and the *Shaping Retail* industry report. While the report presumed that in store roll-outs of the ideas tested in VR would provide good validation data, marketers reported that "practically everybody changes something between the time they test and the time it goes to market" and that often changes that test well are not implemented in stores.

While the experts interviewed were generally more confident that VR provides trustworthy results, they also felt that more validation research would be worthwhile.

A number of the market researchers interviewed expressed a desire to conduct more validation research, including research to confirm that "individual shoppers navigated the store the same way and choose the same brands as they normally buy", but had found it difficult to obtain sales data from marketers. This may be because sales and observational data is typically the purview of sales or account management teams in CPG companies; a group with whom market researcher agencies do not report directly.

Differences industry users have observed between real and VR behaviours

Of the 21 marketing managers interviewed, 80% were aware of differences between real and VR behaviours. For example, a number had observed that shoppers spent longer than normal browsing the category in VR, paid more attention to point-of-sales displays and selected more products than shoppers typically purchase on shopping trip. While these are issues that have been reported in the literature (Harwood and Jones 2011, Kleinschmit, Reid et al. 2012), experienced marketers were generally aware of these issues from their own observations.

Marketers had differing opinions about why these differences occur and how they can be rectified. Some thought that shoppers selected too many products because they were trying to be 'helpful' by showing researchers all of the products they would consider or occasionally purchase; others thought these biases were a consequence of shoppers not spending real money or physically handling products or of shoppers being observed by researchers. One of the market researchers interviewed hypothesised that over-purchasing biases were a consequence of VR studies including too many heavy buyers (i.e. individuals who buy the category frequently) and researchers failing to ask participants how many items they usually purchase before conducting VR shopping tasks, a practice he believes primes participants to shop more normally. However, a study published in the literature has disproved this last hypothesis, showing that because shoppers overestimate how many items they purchase in a month, they also purchase too many in the VR simulation (Campo et al. 1999).

Generally, the researchers interviewed were not concerned by these biases because they calibrate (i.e. discount) purchase quantity data or focus only on which retail environment maximised sales (i.e. performs "best"), not specific purchase estimates. However, some researchers were concerned that these biases may indicate there are differences in virtual and real decision-making processes:

> *Results might be okay at the aggregate level but at the individual level there might be considerable variation… similarities in market shares might be masking differences in the behaviours that really matter.*
>
> Global director, research agency

Biases industry users have observed in VR data

Three market researchers (19%) mentioned that they had noticed "a tendency for shoppers to buy more branded products" in VR simulations and fewer private label products. Most attributed this bias to shoppers not wanting to "admit that they buy PL products". Researchers were divided as to whether this bias was evident in all VR studies or only those that used PL products with made-up packaging. One of these researchers commented that they generally over-sample PL buyers or exclude shoppers not willing to purchase PL products to ensure the sample of PL purchases will be similar to real stores in studies where this bias is of concern.

Some marketers had also observed other biases in VR data that were not mentioned in the literature. One of the managers interviewed had found that the shares of larger pack sizes (e.g. family and value packs) were 10-20% larger than expected in a study they conducted to investigate which pack sizes they should remove from their range. A very experienced market researcher confirmed that he had observed this pattern in at least one study for breakfast cereal. In this study, for approximately 80% of brands (n ~ 50 products) the large 'value pack' size was more commonly selected than the regular pack size, something that he commented rarely occurs in real stores. This experienced industry user thought it was possible that this bias had been apparent in other VR studies but missed as managers generally request brand, product or category-level results in reporting, not pack size splits. This was confirmed by my examination of managerial reports for 40 industry studies for analyses in Chapter 5.

It is certainly plausible that purchases of larger pack sizes could be overestimated in VR studies. Not only do shoppers tend to ignore realistic purchase limits in VR studies (Campo, Gijsbrechts et al. 1999), it is easy for shoppers to underestimate the size of products when viewing them without the capacity to feel their size and/or weight. Neurological studies have found that participants often miscalculate the size of objects when relying solely on visual inputs (Aglioti, DeSouza et al. 1995). Furthermore, a number of previous studies

(Cohen and Gadd, 1999, Campo et al. 1999, Waterlander et al. 2015) showed a single pack size for each study, so this bias may have been missed.

Findings about the accuracy of VR data across product categories

While most marketers and market researchers were aware that the accuracy of VR data is likely to be influenced by the product categories researched, most felt that it was suitable for the product categories they were responsible for researching. All of these were packaged products that do not require sensory information for decision-making.

While most of the managers interviewed had not conducted validation research, a few had observed differences in the accuracy of VR across studies:

- One recalled a project for canned tuna that failed to provide accurate market share data. They attributed this to a brand with substantial market share being omitted from the study (i.e. 'bad shelf set').
- One recalled a project for potato crisps that failed to provide accurate market share estimates. They attributed this to the product images not being high enough resolution for shoppers to read flavour information accurately when shopping the VR.
- One had noted that VR data was accurate for canned soft drinks than oral care products such as toothpaste, mouthwash etc. The marketer who contributed this knowledge did not posit any explanations for these results. It is possible VR results were more accurate for soft drinks because brand preferences were more entrenched, the explanation posited by Burke et al. (1992) for differences in the accuracy of VR data across similar categories.
- One marketer had observed that VR data was less accurate in categories with a wide range of prices (e.g. "luxury categories like alcoholic beverages"). They attributed this to the tendency for shoppers to choose more expensive products than they usually buy in VR situations. This tendency may also explain why results were less accurate for oral care products than soft drinks; there are often large variances in the prices of electric and manual toothbrushes and regular toothpastes and teeth-whitening toothpastes.

In addition to demonstrating an awareness of biases, these comments suggest that industry users have learnt through trial and error that realistic shelf sets and clear product images affect the accuracy of results, learnings that mirror the findings of studies reported in the literature (Burke, Harlam et al. 1992, Kleinschmit, Reid et al. 2012) (Waterlander, Jiang et al. 2015).

How VR simulations are conducted

Over 90% of marketing managers expressed a desire for more methodological research into the conduct of VR studies. A number mentioned that a 'set of guidelines' or a 'handbook' would be useful and many felt that more unbiased (i.e. non-agency research) research was

needed as they felt that market researchers "often just justify their own approach" or "get us to spend more money" or are "flying blind" when designing VR studies.

Some researchers did admit that little formal research or experimentation has been conducted to validate the instructions used. Nielsen and Vision Critical are exceptions, having publishing findings from some small experiments. However, most researchers did not advocate for more methodological research, believing that:

> *Over the years we have learnt how to get the data to settle down.*
>
> Market researcher, global firm

When queried, a number of researchers were concerned that a handbook would erode their competitive advantage and constrain and innovation.

Scale and dimensions of VR simulations

Discussions confirmed that the majority of VR studies focus on one area, product category or aisle of the store (see Figure 2). This is consistent with industry VR studies reported in the literature (Burke 1996, Kleinschmit, Reid et al. 2012).

While all of the marketers interviewed felt that simulating an entire shopping trip to collect data for one product category was wasteful,[3] many had had faced pressure from market researchers to do so:

> *There is a tendency among suppliers to want to take you on a walk through a large part of the store … most of the time it is unnecessary and adds significantly to the cost of a study.*
>
> Director, Research Agency

> *I think it is critical for immersing consumers and getting them to shop normally. I can talk about why intuitively it makes sense but don't have data to show.*
>
> Research director, VR market research company

Amongst the marketers interviewed, their preferred approach was to have shoppers imagine they are making a normal shopping trip that would cover different areas of the store, but commence the VR shopping trip at the start of the aisle or area they are interested in researching (see for example Figure 1). This approach reduces the number of product images that need to be uploaded or updated for a VR project and the amount of incentive required for participants, compared to simulating a longer shopping trip.

[3] This a key difference between academic and industry research studies; academic researchers are interested in shopping generally whereas marketers are only interested in shopper behaviour in the product category or area of the store they manage.

Two of the researchers interviewed mentioned that they maintain 'product image libraries' so that when a project is commissioned they do not need to obtain product images. However, both mentioned that keeping these libraries up to date is a challenge given that "the average category has between 100 and 150 items and 10 or 15 percent change every 6 months." Researchers also mentioned that basic VR models of different retail environments (e.g. a Walmart, or a fast food outlet) are often tailored to show specific products and areas of the store. Using basic models speeds up substantially the deployment of projects.

The industry experts interviewed were confident that reasonably accurate data could be obtained from VR simulations that cover just a small area of the store as long the products shown were familiar to shoppers and the section of the store shown was visually realistic. Most of the market researchers interviewed felt that context beyond the category or aisle of interest was necessary to produce accurate results. For example:

> *Over the last five years we've realised that the extent to which you can bring respondents into the store environment can really influence the data collected. How categories are displayed and how shoppers approach the aisle can have a big impact [on results].*
>
> Market Researcher, Global agency

The two approaches mentioned to achieve this were: *(1)* have participants shop for a basket of items that includes the product category of interest and *(2)* use a short video to "fly" participants from the entrance of the store to the start of the aisle of interest. However, none of the market researchers interviewed were able to provide evidence that either of these approaches improve the accuracy of category-specific purchase data:

Discussions highlighted that VR studies were always fielded with a "control" group that shows the pricing, shelf layout and assortment of products currently available in real physical stores. All three groups of respondents agreed this was necessary because as VR results cannot be generalised directly to real stores.

Both marketers and market researchers expressed the belief that VR simulations that looked three-dimensional would provide more accurate data. While none were able to provide evidence to support their belief, it is consistent with the findings of published studies discussed in Chapter 2. Some market researchers also believed it was necessary to adjust shoppers' view of the shelf to match their real height, i.e. for short shoppers to look at VR shelves from a different perspective to tall shoppers. In general, the market researchers interviewed express the belief that "the closer VR looks to real life, the better the data" and that contemporary VR simulations are look very realistic:

I'm dealing with the same objections I was 10 years ago… it's not realistic enough, people don't use real money, it's not high resolution enough… which is ironic because the technology is so much better.

US Insights manager with global responsibilities

Interestingly, not all of the industry experts interviewed agreed that improvements to the visual realism of VR simulations were necessary. Most were confident that as long as VR simulations were "a decent approximation of reality", they would achieve "fairly accurate results". One of the experts interviewed had compared basic and more detailed VR environments and found they produced similar results, a finding that parallels published research (see for example Burke, Harlam et al. 1992, Kleinschmit, Reid et al. 2012, van Herpen, van den Broek et al. 2016). While the marketing managers interviewed were unaware of this research, some did comment that improvements in VR research have been technology rather than research-need driven and that there was a strong desire amongst category management teams and research agencies to use very realistic-looking environments.

There is a tendency to sell in a gold solution when bronze solution would suffice.

UK Insights Manager

Mode of fielding and infrastructure used

Interviews found that the majority of VR studies fielded by industry researchers were displayed on normal computer screens or large (i.e. 27") touch-screens. At the time of this research (2012-14), none of the large shopper research agencies were employing head-mounted displays for research and only three managers had used head-mounted displays to conduct research. All experienced issues participants becoming disorientated, lost or motion-sick because of the displays.

None of the industry users interviewed had used large (i.e. 12-foot) screens to collect VR data although a number had used them to show VR data and environments to retailers and sales staff (see Figure 9 for example). One of the marketing managers interviewed felt that VR data collected using large screens would be more accurate because products would be would be displayed life size or larger but had not used this approach because their screen was located at their company headquarters, making it difficult to disguise the objective of the research when recruiting participants.

Figure 9: Example of a large screen used to display VR environments to retailers



Image from: business.nab.com.au

When asked, most of the marketing managers interviewed felt that large computer monitors (i.e. 22 inch) were sufficient to collect VR purchase and pack choice data but some had concerns about the use of laptops and tablet computers with smaller screens (i.e. 10 to 14 inches). These managers were concerned about screens being sufficiently large that all participants, including those with glasses or poor eyesight, could read the information displayed on product packages.

Most of the VR studies discussed in interviews used conventional point and click tools i.e. keyboard and mouse or joystick. While two managers had used motion sensors or haptic feedback gloves, they both found that these tools were difficult to calibrate and "freaked" consumers out. The only other tool that users reported using was a joystick modelled on a shopping cart/trolley handle (see Figure 6). Versions of this tool where shoppers stood and leant on a trolley handle were also trialled. An example of this interface is provided in the *Shaping Retail* report.

At the time these interviews were conducted, 50% of marketers had fielded a VR study online using a pre-recruited research panel. Marketing managers were generally comfortable using pre-recruited online panels for VR research having used them for survey research for many years. The marketers who had fielded projects online had done so to obtain "better value for money" (i.e. larger samples for less money), reduce fielding windows and obtain more geographically representative samples:

We used to be limited to 10 or 12 research sites…now we can get respondents from all over the country.

Insights manager, US position

Quotas that used to take a month to fill we can get in a week now.

Market Researcher, Shopper research agency

Some of the marketers interviewed expressed concerns that VR data collected through online panels may be less accurate because participants are not supervised or assisted by researchers and because it is harder to verify the characteristics of participants (e.g. age, gender etc.). Some marketers were also concerned that VR environments would have to be smaller in scope and resolution to fit consumers' personal computer screens and internet bandwidths and that shoppers may not have the right "mindset" when participating online.

We're intercepting subjects on a real [shopping] mission and asking them to shop as though they were in the real store so behaviours should be more similar. [Participants were intercepted in a grocery store car park.]

Researcher, global agency

Concern regarding the "mindsets" of online panel participants was echoed by a few marketing researchers. However, most of the market researchers interviewed were reluctant to express concerns regarding online fielding, perhaps because they are finding "in person fielding increasingly difficult to sell" and they all expected to conduct more VR research online in the future. Most of the market researchers interviewed relied on specialist panel providers (e.g. SSI) to procure respondents for VR studies. None had compared the accuracy of shopper data from different research panels prior to selecting a provider. Most selected research panels based on costs and/or company affiliations (i.e. used the panel of a partner or subsidiary company). Furthermore, none had commissioned research to compare VR data collected in person with VR data collected online before moving offering online fielding to clients. This suggests that their decisions to collect VR data using online panels was motivated primarily by logistical and cost concerns.

In contrast, the majority of VR studies reported in the literature were fielded in person. It is possible that academics prefer to conduct studies in-person so they can observe and interact with participants and are less concerned about fielding costs.

Fielding protocols including the number of trips simulated

Similar fielding protocols, recruitment approaches and instructions were used across most of the CPG and research companies interviewed. Companies did not typically use budgets, time pressures or specific shopping scenarios (i.e. imagine you are shopping for a home movie night) as researchers found that these are difficult for some shoppers to understand,

a finding that is consistent with the literature (Cohen and Gadd 1996) (Burke, Harlam et al. 1992). Interestingly, one of the few companies that had experimented with assigning shoppers' purchase scenarios (e.g. imagine you are shopping for snacks for a home movie night, imagine you are shopping for snacks to take to work etc.) found that these had little to no impact on the products shoppers selected. Most of the companies interviewed wanted participants to "shop as they normally would".

None of the researchers or managers interviewed had utilised presence or immersion questionnaires to determine how successful VR simulations were in inducing feelings of presence. Some had asked shoppers if they "behaved normally", whether "prices were realistic" and "if the virtual store looked like the store you usually shop" in post-shopping surveys, but this was unusual. More commonly participants reported that:

> *Most shoppers have never seen anything like it…They impress all but the most tech-savvy of respondents.*
>
> Researcher, global agency

Interviews found that industry users do not typically field VR studies with multiple shopping trips to habituate shoppers to the VR environment, the approach recommended by prior research. CPG studies are generally fielded with:

(1) an introduction to the VR environment and interface, either verbal, written or in a video,

(2) a practice task to confirm that shoppers can find and select items accurately,

(3) a 'normal' shopping trip and,

(4) a 'seek and find' task where shoppers locate a specific product as quickly as possible.

Researchers who had conducted studies with multiple 'normal' shopping trips had found that shoppers were 'hyper-attuned' to differences in the virtual environments, even if there were none, and that this had undermined the validity of the data collected.

> *We've not had good success with sequential monadic designs … the problem is subjects spend the whole time trying to figure out what changed … we end up having to throw out the second set of data.*
>
> US Research director, VR research provider

Secondly, some researchers felt additional trips were not needed in studies to pick which layouts or packs should be trialled in real stores. This is likely to be a misunderstanding of the literature—additional trips were recommended to improve the accuracy of shoppers' choices and behaviours by desensitising them to the novelty of the VR environment (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999). Additionally, some market researchers

were concerned that "throwing out" the data from shoppers' first few trips, the approach recommended by prior studies, would be perceived as a "waste of money" by their clients.

The only researchers who had conducted studies with multiple shopping trips had done so to *(a)* collect more data by having participants shop multiple layouts or *(b)* to test how stock-outs affected shoppers' product choices. However, these types of studies were not common amongst the marketers interviewed and some market researchers had concerns about repeated purchase exercises:

> *Repeating purchasing exercises does not give a good indication of repeat/loyalty rates without a physical product experience between to influence the shoppers' choices.*
>
> Senior market researcher, US research agency

These discussions suggest that industry users have different motivations and action standards to academic researchers when conducting research and are unconcerned about the accuracy of purchase quantity data but are concerned by the realism of shoppers' product choices and behaviours. These discussions also highlight practical issues with conducting studies as recommended by the literature.

Chapter Summary

A broad cross-section of marketers, market researchers and industry experts that have used VR for research were interviewed to gain access to industry knowledge and established that relatively few marketers (n = 9, 23% of the sample) have conducted research examining the real-world validity of VR data.

Most industry-based methodological research has been informal—through trial and error, industry users have learnt that VR shopping simulations must:

- Include all products that have substantial (i.e. > 5%) market share if they are to provide useful product choice data.
- Have a control cell that replicates the pricing, shelf layout and product assortment of real stores.
- Use images with enough resolution that participants can read flavour, size/weight and ingredient information from product packaging. Using pack images is likely to be particular important for the selection of PL products as consumers may be less able to recall the names of these products.
- While it is certainly possible that customer recognition of PL packs is lower than for national brands, particularly where PL packaging is plain (Romaniuk 2016) and PL brands have not invested advertising (Nenycz-Thiel and Romaniuk 2014), these factors would be expected to affect attention in real stores equally.

- Focus on one or a small range of product categories to be cost effective.

Discussions also identified that VR simulations were most commonly used to test new shelf layouts, packages and product assortments, areas that have not been the focus of validation studies in the literature. They have tended to focus on the suitability of VR for pricing, promotion and shopper repertoire research e.g. (Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016); areas of that are of less concern to marketing managers. This underscores the importance of engaging with industry users before conducting research.

The validity of VR data, particularly in regards to the realism of shoppers' product and brand choices, was a concern of a number of industry users. Marketers and market researchers have observed many of the same biases that are reported in the literature, as well as some additional biases including a tendency for shoppers to select larger pack sizes than they usually buy. While some of the industry users interviewed agree with academic researchers that these are likely to be a consequence of social desirability or shoppers not spending real money (Cohen and Gadd 1996, Campo, Gijsbrechts et al. 1999), other factors including the instructions used, the samples recruited and screens sizes for VR studies were mentioned as possible causes. That the same biases have been identified in industry and academic studies but attributed to different causes, suggests more research is needed.

Interviews also identified that most marketing managers employ different fielding approaches and VR technologies than have been examined in published research. Studies are typically displayed on normal computer screens, focus on a single product category or area of a store and fielded to pre-recruited online research panels, methods that have not been the focus published studies. Furthermore, VR studies typically simulate just one 'normal' shopping trip, an approach not recommended in the literature (Burke, Harlam et al. 1992, Burke 1996, Campo, Gijsbrechts et al. 1999).

The next chapter brings together the findings of these interviews with and the literature to discuss the most critical areas for research.

CHAPTER 4
RESEARCH QUESTIONS

Overview

This chapter presents six research questions that address identified gaps in knowledge regarding VR. These questions relate to the accuracy of shoppers' product choices, the biases that exist in VR purchase data and how the fielding and design of studies affects the validity of VR data.

Area 1: Accuracy of shoppers' purchase choices

Previous studies have tended to presume that VR simulations capture shoppers' real purchase choices because real and VR market share estimates are highly correlated (e.g. *r* => .80). The only other studies to investigate this question, relied on participants self assessments of how realistically they behaved, a metric that is likely to be inaccurate (Martin and Morich 2011, Scamell-Katz 2012). However, the biases apparent in prior VR studies suggest that as many as 30% of participants choose items than they do not usually buy in VR shopping simulations.

RQ1: How accurately are consumers' real purchase choices captured in VR?

Additionally, previous studies have identified a number of biases in VR data including a tendency for shoppers select more premium grocery products than they usually buy (Higaki 2004, Needel 2013, van Herpen, van den Broek et al. 2016), to pay more attention to point of sales displays (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Harwood and Jones 2011, Burke and Leykin 2014) and to select more products than they usually buy on a single shopping trip (Burke, Harlam et al. 1992, Cohen and Gadd 1996, Kleinschmit, Reid et al. 2012).

Being able to replicate the real shares of products is an important test of the validity of VR data and is important to marketing managers aiming to use VR to predict how products will perform in real stores (Burke, Harlam et al. 1992, Mueller Loose and Lockshin 2013).

This research will investigate:

RQ2: Are VR purchase data for some brands, pack sizes and types of products consistently biased? How large are these biases in VR studies for different product categories and retail environments?

Little is known about the prevalence or magnitude of biases in VR studies for different products and retail contexts as previous VR studies did not investigate how the branding or pricing of products affected the biases observed.

Area 2: Can VR be used for shelf layout research?

Marketers and market researchers are also eager to use VR to collect quantitative estimates for new shelf layouts and determine the value of different shelf positions in real stores, but research establishing the validity of VR for this type of research is limited.

RQ6: Do VR simulations displayed on normal computer screens provide results consistent with what happens in real stores when shelf layouts are changed?

This will extend prior research demonstrating that some areas of the shelf differ in their popularity in VR simulations and real physical stores (Massara, Porcheddu et al. 2014, van Herpen, van den Broek et al. 2016) and determine if VR shopping simulations can provide realistic sales estimates for shelf layout changes not yet implemented in stores. Discussions with industry users found that VR shelf layout studies are typically displayed on normal computer screens, even when fielded in specialised market research facilities, an approach that has not been validated for this type of research. Previous validation studies used wide-screen and head mounted displays (Porcheddu and Venturi 2011, van Herpen, van den Broek et al. 2016).

Discussions with industry users identified that changes to category order, shelf layout and shelf position are often tested with VR but are rarely implemented in real stores as retailers are reluctant to allow changes to the arrangement of products. This question could be addressed using results from sales experiments conducted in real stores as benchmarks (e.g. Drèze, Hoch et al. 1994, van Nierop, Fok et al. 2008), if VR studies testing similar changes can be obtained from industry partners.

Area 3: How should VR shopping simulations be conducted?

The literature review identified three pertinent gaps in knowledge:

- Prior studies have not investigated the biases noted in past studies, despite researchers active in the field pointing out that the biases in VR data should be carefully examined to develop the accuracy of the method (Campo, Gijsbrechts et al. 1999).
- No studies, published or unpublished, have compared the accuracy of VR purchase data collected in-person in specialised research facilities with VR data collected online using panel members' home computers. This is a significant gap given the popularity of online panels for research (GreenBook 2015) and the many advantages of online fielding, including the ability to access young, tertiary educated respondents who are more likely to provide realistic VR data (Waterlander, Jiang et al. 2015). While the mode of research has been shown to *not* influence the results of cognition experiments in psychology (Klein, 2014), it is still widely believed to affect the validity of results in marketing and consumer research (Gosling, Vazire et al. 2004, Walker, Pettit et al. 2009).
- Early validation studies concluded that three or four practice trips were necessary to routinise behaviour before collecting VR shopping data (Burke, Harlam et al. 1992, Burke 1996, Campo, Gijsbrechts et al. 1999). However, since then consumers' familiarity with computer-mediated 3D environments has increased, in part due to consumer use of 3D visualisation to tools such as Google Earth. Furthermore, more recent studies have failed to demonstrate that "warm up" trips are necessary to improve the realism of shoppers' VR brand, product or pack size choices (Needel 2013, van Herpen, van den Broek et al. 2016). The only recent study to this factor found that less than 3% of shoppers' purchase choices differed across the 12 weeks simulated (Campo, Gijsbrechts et al. 1999), a very small difference given the costs associated with having participants simulate up to 12 weeks of purchases.

The following questions will address these gaps in knowledge:

RQ3: Why do biases for premium brands and private labels biases occur? To what extent are they a consequence of how VR studies are conducted and the technology utilised?

RQ4: Are multiple trips necessary to collect accurate purchase choice data?

RQ5: Does online fielding to pre-recruited research panels affect the generalisability of VR data to physical retail environments and is VR data collected online less accurate than VR data collected in specialised research facilities?

Discussions with industry users identified that there may be scope to address these questions by utilising data from VR studies commissioned by different CPG companies as many use similar VR equipment, recruitment approaches and fielding protocols. These questions cannot be addressed through meta-analysis of data in the public domain as prior studies did not report enough information about the brands, prices and products studied to enable additional (or even comparable) investigations to be undertaken. While controlled experiments are typically used to establish *why* biases occur, structured hypothesis testing could be used. For example, one of the explanations posited for the under-performance of generic brands and private labels products in VR studies is that shoppers are reluctant to select socially undesirable brands when observed by researchers (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Higaki 2004, Needel 2013). This hypothesis could be tested by comparing results from studies which differ in their observation/supervision.

CHAPTER 5

QUANTITATIVE DATA SOURCES & RESEARCH APPROACH

Overview

The interviews conducted with marketing managers and researchers to establish what biases, if any, exist in VR data, identified that studies are typically executed in similar ways across companies, fielding modes and projects. This provided scope to use data from different companies and research providers to establish the extent, magnitude and cause(s) of biases in virtual data.

This chapter discusses the benefits of this approach, outlines how suitable data sources were identified and obtained, and discusses how analyses were structured to address the research questions outlined in Chapter 4.

This thesis utilises data from eight pre-existing virtual studies that were commissioned to research packaging, merchandising and shelf layout issues. Purchase data from these studies is benchmarked to sales data from real stores and purchase estimates from participants to address RQ1 and RQ2. Suitable benchmarks were obtained from Nielsen's Retail Measurement (RMS) database, a database that records real purchases of 2.6 million products in 35,000 retail outlets across the USA using point-of-sale technology (Nielsen 2009). This validation approach can be used as all of the VR studies selected for analysis used product assortments, shelf layouts and prices from real stores and aimed to provide nationally representative data for normal shopping trips. Similar approaches have been used in previous studies (see Chapter 2 for further discussion).

To address RQ3 data from a variety of sources including real online shopping trips, VR exercises and VR gaze-tracking is utilised. To address RQ4 and RQ5 data from VR studies collected for the same products but using different approaches is compared. For RQ6 a slightly different approach is used; data from a virtual study that examined whether rearranging products on shelf would increase brand and category sales is compared to findings from published studies that tested similar changes in real stores (e.g. Drèze, Hoch et al. 1994). This is necessary because the changes tested in VR studies are rarely implemented in real stores (see Chapter 3 for further discussion).

Natural experiment based on industry data

Typically, researchers are interested in establishing how the conditions of a study affect the accuracy of data employ randomised controlled experiments. These are purposefully

designed to ensure that the *only* thing that varies between groups, are the factors being tested. While it may have been possible to design a series of experiments where subjects were randomly assigned to different conditions (i.e. shop a VR simulation vs a real store, shop a VR simulation in person vs online, hypothetical choice vs spend real money) to address the research questions, doing so would have been difficult. Significant funding would have been required to undertake new VR studies with sufficient samples to detect differences. Using existing studies in a natural experiment overcomes these limitations and enabled all of the questions outlined in Chapter 4 to be studied with robust samples and industry data sets.

Utilising studies designed for a variety of different purposes avoids the possibility that studies are unintentionally designed to produce confirmatory results (Holman, Head et al. 2015). It also enables multiple sets of data (MSoD) to be used for each research question. Using MSoD protects against findings that are limited in their generalisability and apply only to one retail context, study or category (Hubbard, Brodie et al. 1992, Easley, Madden et al. 2000, November 2004, Uncles and Wright 2004). MSoD are particularly useful for studying bias, a difference that occurs in all iterations of method (Groves and Lyberg 2010).

Natural experiments tend to employ one of two approaches. The first uses statistical procedures to eliminate the influence of extraneous variables (i.e. difference may influence results but are superfluous to the study) and to isolate the effects of confounding variables that may be responsible for the differences observed. The second approach uses careful design to limit the number of extraneous differences between conditions and eliminate, wherever possible, variables that may confound results (Sekhon and Titiunik 2012). The second approach produces more consistently trustworthy results (Dunning 2012) and is used in this thesis. Natural experiments are used extensively in advertising research to determine the impact of advertising exposures on consumers' purchase propensities but have not been utilised in VR research.

Step 1. Gaining access to industry data

The in-depth interviews used in Chapter 3 were also used to identify companies interested in contributing data for validation. Five companies agreed to share data from studies they had commissioned and together they provided access to 28 VR studies fielded between 2008 and 2012. The cost to conduct these studies is estimated to be in excess of US$1.2 million. None of these studies had been validated or reported on publicly, making them good targets for research.

Almost all of the studies provided were for packaged grocery products, e.g. paper towels, confectionery etc., due to the preponderance of these types of companies in research to identify users of VR (see Appendix 1, page 202). This does not pose a problem, as analysing data from *similar* product categories initially will provide a strong foundation for future

studies (Ehrenberg and Bound 1993, Lindsay and Ehrenberg 1993, Uncles and Wright 2004). Furthermore, focusing on packaged grocery categories means that concepts developed to explain how consumer behaviour varies across product categories (e.g. hedonic/utilitarian value) can be utilised (Sethuraman and Cole 1999, Wakefield and Inman 2003).

Step 2. Selecting studies for analysis and outlining data collection procedures

Data sets were selected prior to analysis to avoid selecting studies that were confirmatory or biased towards a particular result. All 28 data sets were reviewed and the most suitable were selected for each RQ. The objective was to:

(1) select studies that were capable of providing realistic data (i.e. had been conducted well,

(2) did not have any obvious sources of bias and,

(3) had been fielded with similar data collection protocols.

The criteria used to select studies for analysis are outlined on pages 71 through 79. Data from three companies was excluded as it did not meet these criteria. Differences between studies were limited to variables that were unlikely to compromise comparisons or the real-world validity of VR data based on prior research (see Chapter 2). This included small differences in the brands and number of products included in each study.

.

Criteria 1: Simulated environment representative of real purchase environment(s)

Past studies have established that it is not necessary to include 100% of the products available in market for simulated purchase data to be accurate (Campo, Gijsbrechts et al. 1999, Waterlander, Scarpa et al. 2011, Waterlander, Jiang et al. 2015, van Herpen, van den Broek et al. 2016) but the number of choices available affects decision making in real and VR environments (Campo, Gijsbrechts et al. 1999, Iyengar and Lepper 2000), so it was essential that the studies selected included a similar range and number of products as real stores. Generally grocery stores have larger product assortments than mass merchandise or drugs stores (Fox, Montgomery et al. 2004). The VR studies selected for analysis reflect this tendency (Table 5: *Products n*). VR studies tended to have larger product assortments when designed to replicate grocery rather than mass or pharmacy purchase environments.

In Table 5 the proportion of total category sales represented by the products included is reported for the each of the VR studies selected. This figure could be calculated accurate as all of the studies selected for validation simulated specific retail environments (e.g. a mass merchandise retailer like Target, or a grocery store like Kroger) for which category sales data is available.

Table 5: Product range for each VR simulation selected for analysis

VR Study	Product Cat'ry Simulated	Retail Enviro Simulated	Products n	Coverage NielsenRMS
6	Pantry staple	Grocery store	180	95%
3	Feminine care	Mass merchandiser	218	88%
7	Feminine care	Pharmacy store	214	88%
8	Disposable nappies	Mass merchandiser	16	87%
3	Feminine care	Pharmacy store	144	80%
1	Chewing Gum	Grocery store	56	80%
1	Confectionery	Grocery store	330	75%
4	Disposable nappies	Mass merchandiser	115	73%
4	Disposable nappies	Grocery store	123	71%
2	Pet care	Mass merchandiser	185	69%
5	Paper towels	Grocery	27	54%
5	Paper towels	Mass merchandiser	25	45%
		Average study	136	75%

On average the selected studies included 136 products, between two and four different pack sizes and covered 45% and 95% the market. None of the studies selected for validation included made-up brands in place of private labels or new products as these can disrupt real-VR comparisons (Bressoud 2013, Needel 2013).

Criteria 2: The arrangement of products mimicked the arrangement of products in real stores

Research has established that VR environments are enhanced by the meaningful arrangement of objects (Hoffman, Prothero et al. 1998). All of the studies selected based the arrangement of products on shelf layout plans from real stores (i.e. the planogram or schematic).

This criteria was necessary because the majority of purchase decisions in real stores are from familiar shelf layouts and product assortments (Park and Lessig 1981, Lingonblad 2015) as most grocery shopping is undertaken in stores with which consumers are highly familiar (Sommer and Aitkens 1982, East, Hammond et al. 2000, Sorensen 2016). Past validation studies have also tended to also use familiar shelf layouts (Burke 1996, Campo, Gijsbrechts et al. 1999, Soderman 2005) and there is some evidence from VR studies that when consumers are shown VR environments that do not look familiar they utilise visual information including signage and product displays to locate relevant product categories (Burke and Leykin 2014) rather than mental maps of the store's layout formed through past experience (Sommer and Aitkens 1982, Scamell-Katz 2012).

Participants assigned to shop new or unfamiliar shelf layouts were not included in analyses, nor were studies that *only* tested new shelf layouts that would be unfamiliar to shoppers.[4] Familiarity affects how consumers navigate the shelf, how long they take to make a decision and the cognitive processes in which they engage—unfamiliar purchase environments increase substantially the amount of cognitive effort, quantity of reading at shelf and time taken to make purchase decisions (Park and Lessig 1981, Russo and Leclerc 1994, Lingonblad 2015).

This criterion reduced substantially the sample sizes available for some VR studies. For example, Study 2, a study focused on the pet care aisle, had 50 participants shop a familiar shelf environment, as a control for the other 250 participants who shopped novel shelf environments. While a sample of 50 is not ideal, in real stores most shoppers are highly familiar with the shelf layout (Sommer and Aitkens 1982, East, Hammond et al. 2000, Sorensen 2016), so excluding the participants who shopped novel shelf environments is necessary to reduce number of extraneous differences between real and VR results. The importance of limiting the number of extraneous differences between conditions wherever possible is discussed on page 75. The average sample of purchases validated from each study was 495.

An alternative would have been to exclude shoppers who indicated in post shopping surveys that the arrangement and/or selection of products differed from the real store(s) at which they usually shop. This would have resulted in larger samples as typically more than 80% of participants indicated that the layout and assortment of products in the VR environment was familiar to them, even when they were exposed to a new test layout. However, this approach would have resulted in data from fewer studies and product categories being validated as the identifiers needed to match VR purchase data to survey data were not available for all projects, introduced a potential confound to the validation of data from real stores and relies on consumers being *aware* of shelf layout differences that may have influenced their behaviour.

Criteria 3: Studies simulated normal shopping trips

As budgets or fictitious scenarios can cause shoppers to deviate from their normal purchase choices (Sawyer, Worthing et al. 1979, Cohen and Gadd 1996), none of the studies selected for validation required shoppers to adhere to time limits or budgets, shop for a specific occasion or purchase particular products (e.g. imagine you have $15 to buy snacks for a home movie night). All emphasised that shoppers were to make a 'normal shopping' trip and use the 'purchase' button or command to select products they saw that they wanted to 'purchase'.

[4] Study 6 utilised on familiar shelf layouts and three unfamiliar so could be used in analyses.

To encourage consumers to make realistic purchase decisions, all of the studies selected for validation showed the real, non-sale prices of products. In instances where the non-sale prices of products varied across stores or markets, the most common price for urban centres was used and the price seen was not tailored to the research location

None of the studies selected for validation put products on sale temporarily or varied the prices of products in any way. Most studies also provided a running total of the cost of the products shoppers had selected, information that has been shown to encourage shoppers to substitute national brands for cheaper store brands (Van Ittersum, Wansink et al. 2013).

Information about how studies were conducted was obtained from questionnaires, reports and in-depth interviews conducted with marketers and research agencies (n=8). Interviews were necessary as not all of the information was documented in questionnaires. The instructions for each study are detailed in Appendix 4.

The data collection process was similar for all of the studies validated; shoppers were shown how to navigate the virtual simulation, taught how to select products in a practice session and then instructed to make a 'normal' shopping trip. For example:

> In a minute, we are going to ask you to make a shopping trip in a virtual reality store. In the interests of time, you are just going to shop for [product category] but please imagine it's a normal shopping trip. Please shop as you normally would, buying as much or as little as you would on a normal [retailer name] visit.

In some instances, shoppers also completed a *seek and find* task where they needed to find a specific product in the simulated environment as quickly as possible. As these were completed after their *normal* shopping trip, they were unlikely to have affected the *normal* trip data. Results from *seek and find* tasks are presented in Appendix 7 to show that shoppers were capable of navigating the virtual simulation and finding particular pack sizes. In some studies, attitudinal data about the retail environment, the product category or future purchase propensity was also collected after purchase exercises (see Appendix 5 for examples of the questions asked).

While there were slight differences in the wording of instructions across studies, they were sufficiently similar that results can be compared. This is established in Appendix 6. In studies where shoppers simulated more than one shopping trip (n = 2), participants were told to imagine that it was a new shopping trip at the same store. This is similar to the instructions used in other validation studies (e.g. Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Kleinschmit, Reid et al. 2012).

Criteria 4: Studies collected purchase data

All of the selected studies collected simulated purchase data that could be analysed quantitatively (see Appendix 3 for a list of metrics typically collected). Studies where only eye-tracking, purchase consideration or product/pack evaluation data was collected were included. This excluded two studies.

Criteria 5: Samples aimed to be representative of category buyers generally

The specific criteria used for each study are detailed in Appendix 4 and discussed in Appendix 6. Recruitment criteria were generally based on retailers' beliefs or marketers' segmentation schema. As discussed on page 251, some of the demographic and behavioural criteria imposed may not be necessary or advisable.

All of the selected studies recruited category buyers who were familiar with the retail environment simulated. Familiarity with the category and retailer was established using pre-recruitment screening questionnaires—shoppers were asked how frequently they purchase for our five product categories (i.e. fortnightly, once a month, etc.) to disguise the category of interest. The limitations of using recalled purchase data to determine the realism of shoppers' purchase choices is discussed in later in this chapter.

None of the studies selected for validation excluded buyers of particular brands or pack sizes and shoppers were not told which brand or company had commissioned the study. Studies that focused on particular segments of customers (e.g. first-time category buyers) were not selected. To aid validation, it would have been desirable to match the samples for each study to the purchase population for each category using a wide range of demographic and purchase characteristics. However, this was not possible as the identifiers needed to match virtual purchase data to survey data were not typically contained in project files, suggesting that VR purchase data is not typically weighted to the demographic and buying characteristics of the market. For studies where this is possible, the characteristics of the sample are compared to the buying population obtained from panel data.

When studies were fielded in-person, shoppers were recruited in ten or twelve selected shopping malls to ensure the data would be nationally representative. When studies were fielded online, large Internet panels were used. Past studies have demonstrated that samples obtained via online panels can differ in their age, income and ethnic profile to samples obtained through other means (Yeager, Krosnick et al. 2009, Baker, Blumberg et al. 2010), however all of the VR studies used quotas to ensure the samples would be representative whether collected online or in-person.

Recruitment statistics (i.e. incidence, response rates etc.) were not available but were likely to be low i.e. 10-20% for studies fielded in shopping malls and below 30% for online panel

studies. Researchers often have difficulty recruiting in shopping malls because of low traffic, high avoidance and high refusal rates but they continue to use shopping malls to recruit participants for consumer packaged goods (CPG) studies because of the similarities between shoppers and buyers of these products (Bush and Hair Jr 1985, Bock and Treiber 2003). Response rates for VR fielded online are lower than for surveys fielded online because more training and bandwidth is required for participation. Analyses on page 75 (Chapter 5) and page 136 (Chapter 6) demonstrate that the participants recruited were representative of category buyers.

Criteria 6: Studies had good visual realism

Past studies have found that the realism of the graphics has a significant impact on whether a virtual environment feels real to participants (Witmer and Singer 1998, Bowman and McMahan 2007), a finding that is consistent with the beliefs of marketing managers and researchers. All of the studies selected for analysis, including those fielded online had high quality graphics (see Figure 2 for examples) and good update rates to avoid the environment looking jerky when shoppers navigated through the environment. All were fielded using normal computer screens, an approach that succeeds in making products look highly realistic (Wright, Schnack et al. 2017).

They were also realistic in terms of what shoppers could do and how they could engage with the environment, an important consideration for VR (Bowman and McMahan 2007). In all of the studies selected, shoppers could navigate through the environment store, clicking on products to view close up and rotate products to read information on pack (see Figure 10). The ability for shoppers to interact and control their view of the environment is one of the key differences between virtual simulations and less realistic research methods using static images or projections (Witmer and Singer 1998, Soderman 2005, Tal and Wansink 2011, WARC 2016a).

Figure 10: Shoppers could control their view of the shelf and select products to view



Shoppers' view of the shelf was also tailored to match their real height in a number of studies. This meant that for taller shoppers the top shelves were at eye-level whereas for

shorter shoppers they appeared as though they were looking upwards (see Figure 29, page 246 for an illustration). A number of other visual and verbal cues were used to make shoppers feel as though they were actually shopping the virtual environment, some of which may be necessary for realistic VR data. For example, some studies had a basket or a shopping trolley visible on screen that moved as shoppers steered through the environment (see Figure 6 or Figure 10). A similar approach was used by other studies with a high level of visual realism (e.g. Bressoud 2013, Desmet, Bordenave et al. 2013). In real stores shoppers often approach products side rather than front-on (Underhill 1999, van Herpen, van den Broek et al. 2016).

Studies that did not meet these criteria were not selected. Figure 11 provides an example of one such study; shoppers' view of the shelf was not realistic because they always viewed the shelf front-on. While these types of simulations *may* be capable of providing realistic data, and may be suitable for some types of marketing reach, including high and low quality VR simulations would have introduced a confounding variable into this research.

Figure 11: Example of a less realistic virtual environment



In this study shoppers could only view the shelf front on. Studies of this type were not included.

Unfortunately, none of the available studies collected data about how realistic shoppers felt the VR was (i.e. the extent to which they actually felt present and immersed in the environment) although some did ask shoppers if the simulated environment looked like the store(s) they usually shop. In all of the studies selected, including those fielded online, the majority of participants (i.e. >95%) said that the virtual environment resembled the store(s) they usually shop.

Criteria 7: Products looked three-dimensional

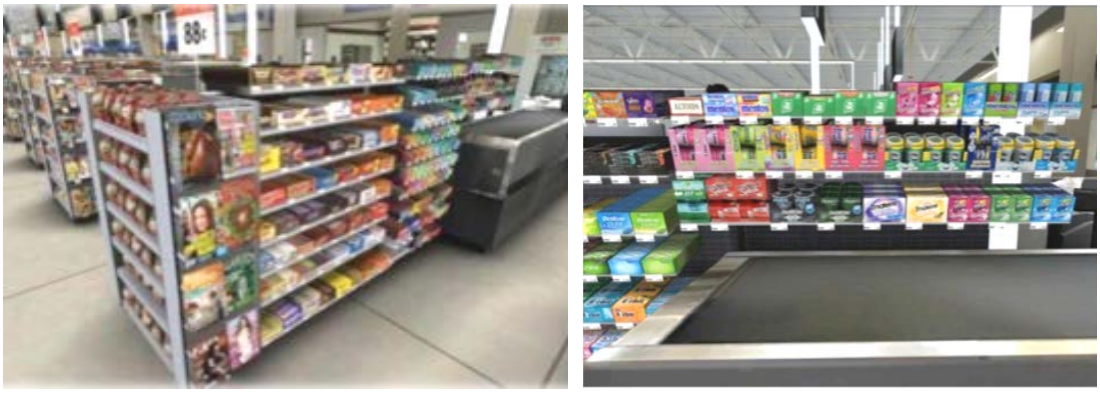
Past studies have established that consumers respond more realistically to pricing and packaging information in simulations where products look three-dimensional than in simulations where flat, two-dimensional images are used (Li, Daugherty et al. 2003,

Berneburg 2007, Kleinschmit, Reid et al. 2012). All of the studies selected used products that looked three-dimensional on shelf (see Figure 10 for example).

Criteria 8: Benchmarking data was available

All of the studies selected for validation simulated a specific retail environment for which real sales data is available. Studies that depicted environments for which sales estimates are not readily available (e.g. vending machines, purchases from seasonal displays, displays at checkout, dollar stores; *n* = 5 studies) were not selected. This was necessary as RQ1 and RQ2 required purchase data from real stores.

Figure 12: Examples of retail environments not validated



Criteria 9: Studies all came from a single market

A number of factors vary cross-culturally including the impact of social desirably (McCort and Malhotra 1993, Steenkamp, de Jong et al. 2010) and the pricing and acceptability of private label (PL) brands (Steenkamp, van Heerede et al. 2010, Nielsen 2014). While the impacts of these factors may be small (Klein, Ratliff et al. 2014) and the variance in the shares of private labels is larger across retailers and categories, than between countries (Nielsen 2014), the decision was made to use data from a single market to reduce the number of extraneous differences between data sets.

As the US was the only market for which multiple data sets were available (see page 69 for further discussion of the importance of these), so was selected for analysis. The data contributed included data from the UK, German, China, Australia. These were reviewed and showed many of the same biases but were not included in formal analyses and are not reported in this thesis.

Criteria 10: Data files could be accessed

Data from VR projects is typically highly detailed. Shoppers' interactions with products and positions in the VR environment are recorded on a second-by-second basis (see for example Appendix 3). This data is typically in proprietary databases and must be converted to CSV, XLS or SPSS format to be analysed by researchers outside the agency that collected them.

In some cases, research agencies had lost the staff or software necessary to access these databases retrospectively or had misplaced information needed to interpret the data extracted (e.g. which environment the participants saw, the list of products included, the .txt file for decoding stored data strings, response codes for survey data files, interaction data for VR studies). At a minimum, information about which products participants selected and environment they were assigned was required for a study to be included in analyses. This condition eliminated six projects.

Evaluations of project information also identified that determining the shelf layout of products from the layout images available for some studies would be difficult given the low resolution of images, as illustrated in Table 15, limiting somewhat the biases that could be investigated with the data sets available.

Figure 13: Illustrating poor resolution of shelf layout information available for some studies



Criteria 11: Eye-tracking data not collected in shelf layout study

For the data used to address RQ6, an additional criterion was applied to ensure that data would not be biased towards particular areas of the shelf. Calibrating eye-tracking equipment requires participants to spend a few seconds focusing particular areas of the screen before beginning their shopping trip. In the studies available, the VR environments used in shopping exercises were also used for calibration, thereby familiarising shoppers with the environment but also (potentially) pre-disposing shoppers to focus and select more items from the areas of the shelf they were instructed to focus on during calibration (Tatler

2007, Bindemann 2010).[5] To avoid this possibility only studies that did not collect eye-tracking data were used in analyses for RQ6.

In the studies validated a five-point calibration protocol in which participants focused on each point for 200 milliseconds was used. In academic studies, between 9 and 41 calibration points are typically used, an approach which would have provided slightly greater accuracy, particularly for environments comprised of many small packages (Tobii 2012, TobiiPro 2018). If shoppers were shown an unrelated environment or a blank screen during calibration, as is recommended (Duchowski 2007, Tobii Technology 2010), it may not have been necessary to exclude eye-tracking studies from analyses for RQ6. However, as mentioned in Criteria 10, determining the on shelf location of products from the layout images saved for some studies would have been difficult for a number of studies.

Gaze and click-tracking data, where participants click where their attention is focused, is used to shed light on investigations for RQ1, RQ2 and RQ3. The gaze tracking data utilised was collected using specialised eye-tracking equipment (e.g. Tobii X30, see Appendix 4 for further details) and not shoppers' webcams. Webcam eye-tracking is not accurate enough at this stage to ascertain whether shoppers are looking at product, pricing or background information (Sewell and Komogortsev 2010).

Step 3. Ensuring selected data sets were sufficient for the RQs

Eight data sets met the criteria outlined. These included commonly purchased grocery products including toilet paper, paper towels and confectionery and retail contexts including mass merchandise retailers (i.e. Target, Walmart), grocery stores and pharmacies (drug stores). The product categories and retail environments encompassed are detailed in Table 10. They are described in general terms (e.g. Study 8, a baby care study in a mass merchandise environment).

The data sets used to address each RQ are outlined in Table 10. At least two data sets are used to address each RQ. While a greater number of data sets would have been desirable (Ehrenberg 1995, Klein, Ratliff et al. 2014), why this was not possible is discussed below using RQ6 as an example.

Four of the studies that met the criteria outlined on page 71 to 80 tested shelf layout changes that could be benchmarked to results from experiments in real stores (i.e. Dreze, Hoch et al. 1995, Chung, Schmit et al. 2007, van Nierop, Fok et al. 2008). The effects these studies tested is summarised in Table 6.

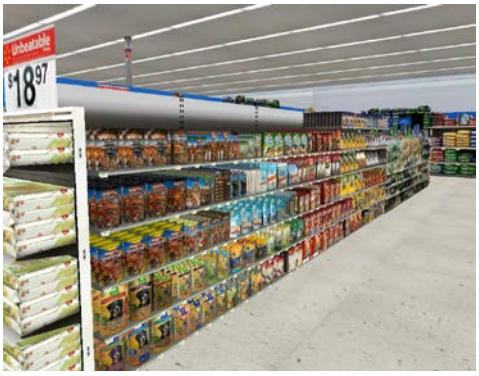
[5] This conclusion is somewhat controversial—some studies have posited that the tendency for participants to focus on the centre of screens is part of a natural predisposition to focus initially on the centre of the scene, a location which generally corresponds to the centre of the screen, and not a consequence of the calibration process for eye-tracking (Tseng, Carmi et al. 2009, Bindemann 2010). Either way, it was better to be safe.

Table 6: Changes tested in shelf layout studies available

Study	Fielded	Sample per change/layout	Organisation	Order in aisle	Shelf placement
2	InPerson	50 participants 131 purchases	No	Yes, but start position of participants not recorded	No
6	InPerson	72 participants 250 purchases	Yes	Yes	Yes, some brands
7	Online	300 participants 650 purchases	No	Yes, but start position of participants not recorded	Yes, some brands
8	Online	200 participants 372 purchases	No	Yes, but start position of participants not recorded	No

Table 6 summarises a number of issues with the studies available. Firstly, only one study, Study 6, tested the impact of internally organizing a category, a change that has been found to have a small but positive effect on category sales in real stores (Dreze, Hoch et al. 1995). Secondly, in all of the VR studies testing shelf layout changes, participants' starting positions were alternated as the direction from which shoppers approach the aisle will influence which products they see first (see Figure 14 as example). However, only one of the studies outlined in Table 6, Study 6, recorded the starting position of participants in the data files provided for analysis. This limited the number of studies for which order in aisle effects could be tested to one. While this was not ideal, whether can VR be used for shelf layout research was an area of sufficient concern to industry users (see Chapter 3), that it was deemed to be worth proceeding with investigations for RQ6 in spite of these limitations.

Figure 14: Example of how starting positions were varied in some studies, afffecting which products were seen first





The computing interfaces and scale of the virtual environment (i.e. proportion of the store visible) is outlined Table 7 and Table 8 for the full set of studies validated to demonstrate the comparability of the data sets included and range of conditions tested. The additional details for each study, including the instructions given to participants are outlined in Appendix 4.

Table 7: Scale of VR studies selected for analysis

Study	Fielded	Scope of VR environment
1	InPerson	330 chocolate, confectionery & gum products arranged in an aisle (approx. 9 meters); other product categories visible in the background
2	InPerson	184 canine snacks and treats arranged as though in an aisle (approx. 8 meters); other categories visible in the background
3	InPerson	200 feminine hygiene products arranged in an aisle approx. 9 meters long; not clear if other categories visible in the background as no photos available
4	InPerson	115 products arranged in as though in the aisle approx. 9 meters long; not clear if other categories visible in the background as no photos available
5	InPerson	Up to 45 different products arranged in as though in an aisle approx. 11 meters long; other categories visible in the background
6	InPerson	167 rice products arranged in an aisle (approx. 7 meters); other product categories visible but could not be shopped/selected
7	Online	200 feminine care products arranged in an aisle (approx. 7 meters); other product categories visible in background; store signage also visible
8	Online	Shop 16 pallets of products; arranged in aisle (approx. 15 meters long); entire store could be shopped and store signage visible in background

Table 8: Computing interfaces of VR studies selected for analysis

Study	Fielded	Computer interface for interacting with VR environment
1	InPerson	Normal computer mouse to select products on 20 to 28 inch computer screen
2	InPerson	Normal computer mouse to select products on 24 inch computer screen
3	InPerson	Joystick modelled on a shopping trolley, set below two 24 inch computer screens arranged to provide a 140° field of view
4	InPerson	Joystick modelled on a shopping trolley, set below two 24 inch computer screens arranged to provide a 140° field of view
5	InPerson	Joystick modelled on a shopping trolley, set below two 24 inch computer screens arranged to provide a 140° field of view
6	InPerson	Joystick modelled on a shopping trolley, set below two 24 inch computer screens arranged to provide a 140° field of view
7	Online	Consumers' home computer mice/trackpads and screens; many likely to be small laptop screens
8	Online	Consumers' home computer mice/trackpads and screens; many likely to be small laptop screens

Table 8 details how each study was fielded. When studies were fielded in person researchers verbally recruited participants in shopping malls, administered the survey questions, taught

participants how to navigate the virtual store and supervised participants. When studies were fielded online there was no personal contact between participants and researchers. Thus, the potential for social desirability (SD) bias to affect results varies across these modes (Colombotos 1969, Krysan, Schuman et al. 1994, Presser and Stinson 1998, Holbrook, Green et al. 2003, Duffy, Smith et al. 2005, Chang and Krosnick 2010) enabling RQ3 to be addressed using data from four studies (see Table 9). The likelihood of differences, if apparent in these studies, being statistically significant is discussed on page 98.

Ideally some of the VR studies validated would have included a questionnaire (e.g. Paulhus 1991) to substantiate that the likelihood of (SD) affecting VR purchase choices differed across modes; unfortunately this was not the case.

Table 9: Summary of studies used to address each RQ

Study No.	Product Catg.	RQ1 How accurately are real preferences captured	RQ2 What biases consistently present	RQ3 Likely causes of bias	RQ5 Do additional trips help	RQ4 Is online data any better	RQ6 VR suitable for shelf layout research
1	Confectionery		✓	✓	✓		
1	Gum		✓	✓			
2	Petcare		✓				
3	Femcare	✓	✓	✓		✓	
4	Babycare	✓	✓	✓	✓	✓	
5	P.Towels	✓	✓	✓			
6	Pantry staple	✓	✓				✓
7	Femcare					✓	✓
8	Babycare					✓	
Number of studies		4	7	5	2	4	1

Multiple studies are used to address each research question.

The statistical power of the studies available was assessed for RQ2, RQ5 and RQ6, research questions for which effect size benchmarks could be estimated from past studies. In past studies, the shares of PL brands were underestimated in by 50% or more (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013, van Herpen, van den Broek et al. 2016) and purchases of premium brands were overestimated by between 10-200% (*M* = 100%, *StD* = 80% Campo, Gijsbrechts et al. 1999, Higaki 2004, Needel 2013, van Herpen, van den Broek et al. 2016). In past studies, additional trips reduced purchase quantity biases by up to 35% (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999). Shelf layout changes have been shown to affect real sales by between 5% and 69%, depending on whether their focus is product organisation (M_{effect} = +5%, StD = 2.6%; Dreze, Hoch et al. 1995), locating products first in the aisle (M_{effect} = 17%, StD = 18%; Dreze, Hoch et al. 1995, Chung, Schmit et al. 2007, van Nierop, Fok et al. 2008) or featuring products at eye level (M_{effect} = +39%, StD = 30%; Hubbard 1969, Dreze, Hoch et al. 1995, van Nierop, Fok et al. 2008). However, sales effects are often larger in laboratory studies (Simonson and Winer 1992, Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Kahn and Wansink 2004, Porcheddu and Venturi 2011).

Each of the studies outlined in Table 11 recruited between 60 and 300 participants to shop the VR environment that was representative of real purchase environments (i.e. the control cell). As participants typically selected multiple products, data for a larger number of purchases is available, as outlined in Table 11. Participants who selected an unrealistically large number of products (e.g. eight chocolate bars on a single shopping trip) were excluded from calculations, and from subsequent analyses.[6]

To assess the power of the samples for addressing RQ2, the actual shares of premium and PL brands at the time of each VR was obtained from Nielsen's Retail Management (RMS) database, a database that records retail purchases in outlets using point of sale equipment. This type of data is routinely used by consumer goods marketers to determine brand performance (Neslin, Allenby et al. 1994, Dekimpe, Hanssens et al. 2005, Mueller Loose, Osidacz et al. 2010, Sharp and Bogomolova 2013) as it is enables differences in the shares of brands to be measured precisely (Hawkes and Piotrowski 2003). Effect size benchmarks from past studies were used to estimate the VR results and the significance of real and expected VR results determined with a single-sided, Chi-squared (X^2) test. For studies where participants were shown one of two retail environments (e.g. Study 3, pharmacy or mass environment), the sample size for each environment was used in statistical testing.

6 Tukey's box plot method was used to identify outliers. Typically, only two and three percent of shoppers were removed ($M_{outliers}$ = 2.7%, SD = 1.6%). While it would have been preferable to use individual-level panel data to derive cut-off points based on the real spread of the data, this type of data is expensive to obtain. Some of the researchers interviewed spoke about needing to "clean out the one or two percent why buy too much or shop to quickly", most of the data files provided had not been adequately cleaned. This suggests that more stringent (and objective) cleaning protocols are required in VR studies.

Table 11: Details of the VR studies selected for analysis

VR Study	Retail Enviro Simulated	Product Cat'ry Simulated	Sample n	Real PL	Real Premium
1	Grocery store	Confectionery	1 500	3%	10%
1	Grocery store	Chewing Gum	87	-	8%
2	Mass merchandiser	Pet care	131	20%	6%
3	Pharmacy & Mass	Feminine care	310	22%	15%
4	Grocery & Mass	Disposable nappies	680	23%	24%
5	Grocery & Mass	Paper towels	588	28%	31%
6	Grocery store	Pantry staple	697	18%	5%
7	Pharmacy store	Feminine care	638	14%	26%
8	Mass merchandiser	Disposable nappies	372	22%	27%
8 proj'	3 retail contexts	8 categories	5 003		

Summary of sample size test results

- PL brand biases (RQ2): Real/VR deviations likely to be highly significant ($p < .001$) for all studies.
- Premium brand biases (RQ2): Real/VR deviations likely to be highly significant ($p < .001$) for all categories except chewing gum (n = 83), if biases are consistent in magnitude with past studies (M = 100%, Campo, Gijsbrechts et al. 1999, Higaki 2004, Needel 2013, van Herpen, van den Broek et al. 2016). If deviations for premium brands are less pronounced, e.g. 40%, then results may not be significant for studies/categories with sample sizes or small premium brand shares (e.g. chewing gum, Study 2 and Study 3, see Table 11). If this occurs, it will be noted and results for multiple premium brands aggregated together may be reported to resolve this issue. An alternative would be to include in analyses shoppers who saw unfamiliar test layouts to increase sample sizes. However, this may obscure or distort effect sizes by introducing an uncontrolled factor and may reduce the face validity of real-VR comparisons (Sekhon and Titiunik 2012).
- Multiple trips (RQ5): Differences in the VR shares of PL brands on consumers' first and second simulated shopping trips are likely to be significant for Study 4 (Baby care, $p < .05$) but not for Study 1 (Confectionery, $p = .14$) due to the small share of PL brands in this category. However, differences in the shares of premium brands, a product group whose selection rates may be affected by an additional shopping trip (Toh, Lee et al. 2006), are likely to be highly significant ($p < .001$ to .01) for both product categories.
- Shelf layout effects (RQ6): Differences in the shares of products are likely to be significant at $p < .05$ where product shares are equal to or greater than 15% and if effect

sizes exceed or equal 20%. This is likely to occur for the location effects tested in Study 6 and Study 7, for 8 of the 31 of the brands included (26%). It may be necessary to adopt a lower level of significance e.g. $p < .10$ or $p < .15$ to determine if shelf layout changes affect products with smaller shares in VR studies or category organisation are evident (3-8% effect expected, Dreze, Hoch et al. 1995). Adopting a lower significance level does however increase the chances of conclusions being proving false.

- Impact of online fielding (RQ4): Market share results from studies fielded online and in person would need to differ by at 30% for results to have even a marginal level of significance ($p < .10$). An effect of this magnitude seems unlikely given that participants did not need to state verbally which brands they would select in studies fielded in person; only click on them in the VR environment. Past studies have found that the amount of personal interaction between researchers and participants influences the likelihood of socially desirable responding (Colombotos 1969, Krysan, Schuman et al. 1994, Presser and Stinson 1998, Holbrook, Green et al. 2003, Duffy, Smith et al. 2005, Chang and Krosnick 2010), so fielding mode is expected to have only a moderate effect on consumers' VR purchase choices. Confirmatory modelling will use data from a larger number of in person studies to determine if fielding mode is a significant predictor of biases in VR studies, increasing the chances that moderate differences, if apparent, will be statistically significant.

Collectively the studies outlined in Table 11 included 1633 different products and 153 different brands, ten of which were private label or store brands. This should be sufficient for between ten and one hundred variables to be assessed with confirmatory modelling (Stage 2, see pg 105) , depending on whether modelling is conducted at the brand or product level (Mendenhall and Sincich 1996). Market share results in Table 10 indicate that the studies which were available varied substantially in the shares of premium and PL brands and covered a range of simulated retail contexts including pharmacies (drug stores), mass retailers and grocery stores.

While the categories covered were not purposefully selected, all are product categories where the choice heuristics that shoppers rely on (i.e. brand, price and packaging) can be shown in a virtual environment. Additionally, some of these are products that have been validated previously (e.g. paper towels Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999), providing scope for comparison.

Step 4. Identifying appropriate validation data

Source 1: National purchase sales data

The most common approach for validating virtual data is comparing VR data to sales data from real stores. This approach has been used successfully in a number of past studies (e.g. Burke, Harlam et al. 1992, Burke 1996, Campo, Gijsbrechts et al. 1999, Higaki 2004, Tjostheim and Saether-Larse 2005, Needel 2013).

To address RQ2 sales data was obtained from Nielsen's Retail Measurement (RMS) database. Access to this type of data is typically expensive (i.e. US$2,000 per quarter per category). Data for the time periods, product categories and retail environments covered in the VR studies was obtained gratis from Nielsen.

Consumers were expected to draw on their most recent purchase experiences when selecting products, so data was obtained from the period immediately prior to the fielding of each VR study. This is the consistent with the assumptions of past studies. A 12-week period was used to ensure that seasonal trends in product sales were accounted for but each study had a sufficient sample of real purchases (e.g. 440,000 confectionery purchases in US grocery stores between October and December 2010) for results to be accurate at the brand and product (i.e. SKU) level. Nielsen data is collected at the individual product level and was manually adjusted to match assortment of products shown in the VR simulation.

Where the VR studies simulated a specific retail outlet (e.g. a Kroger grocery store) sales data for that retail outlet was obtained. This provided more precision in the shares of products, particularly private label products (see Table 12). In subsequent chapters results are reported by retail channel type (e.g. grocery, mass merchandiser etc.), to enable channel level patterns to be observed more clearly.

Table 12: Validation data at the store level

Feminine care product	Real Share	Virtual Share	
Walgreens	26%	11%	< Referred to as a 'Pharmacy' study
CVS	22%		
Kroger	19%		
Walmart	14%	10%	< Referred to as a 'Mass store' study
Target	13%		
Average	19%		

VR data is compared to benchmarks from the type of retail outlet simulated, not average data for all retail channels.

Using data from specific retail channels is particularly important for identifying biases for different pack sizes. In real stores, large and extra-large pack sizes are typically more popular in mass merchandise environments such as Target, Walmart and Costco than grocery stores or pharmacy stores across the categories studied. For example, across the categories studied, on average large and extra-large packs accounted for 51.6% of real sales in mass stores but 36% of sales or less in grocery and pharmacy stores, significantly different proportions (confidence intervals do not overlap at 95% C.L.).

Source 2: Survey data about shoppers' usual purchases

To address RQ1 data about the realism of shoppers' purchase choices, data regarding participants' normal purchase habits was obtained recruitment questionnaires for VR studies. The studies for which survey data is available is detailed in Table 7 (RQ1) and examples of the questions used are detailed in Appendix 5.

It would have been preferable to use real purchase records (e.g. receipts or diary panel data) marketing managers do not tend to collect this type of data in VR studies. Only one of the 28 managers interviewed mentioned using purchase records to determine what types of products shoppers usually buy.

While consumers are known to forward telescope (Sudman and Bradburn 1974) when self-reporting grocery purchases i.e. to think purchases were more recent than they really were and mistakenly include additional purchases, this is unlikely to affect substantially the accuracy of survey data for VR validation. Survey data for brand and product buying is reasonably accurate—the average difference between actual and recalled purchases is approximately 3% points at the brand level and data can be used to distinguish between heavy and light buyers (Nenycz-Thiel, Beal et al. 2012).

Source 3: Results from published experiments used for shelf layout study

Results from sales experiments conducted in real stores are used as benchmarks for the effects of shelf layout changes in real stores. Results from multiple studies are averaged to obtain robust benchmarks for the direction and magnitude of changes in real stores.

While it would have been preferable to use results from an in-store test of the changes tested virtually conducted during or after the VR test, this was not possible. The shelf layout changes tested in Study 6, the VR study selected for validation, were not implemented in real stores. This is not unusual—retailers are often very reluctant to allow research or shelf layout changes in real stores even when they will benefit brand and category sales (Treiber and Needel 2000, Kleinschmit, Reid et al. 2012). Indeed, their reluctance to do so is often a key motivation for research to be conducted using VR.

Some additional sources of validation data including purchase panel data for online shopping are used to shed light on particular biases. These are detailed in subsequent chapters.

Analysis approach

Stage 1: Descriptive analysis

This thesis begins by using simple descriptive analyses to identify patterns that are consistently present (i.e. repeated) using tables of figures and averages (Ehrenberg 2000).

Descriptive research is extremely valuable for knowledge development (Ehrenberg 1997, Ehrenberg 2000, Ehrenberg, Barnard et al. 2001, Wierenga 2012).

The similarity of shoppers' real and simulated product choices is measured in two ways. Firstly, at the individual level as the proportion of shoppers who select brands from the same brand, pack size, group or price tier in the virtual simulation as they reported purchasing usually in the pre-simulation shopping questionnaire. These are reported as repeat rate (RR) figures, a metric used in a number of past studies (Scriven and Ehrenberg 1995, Romaniuk and Dawes 2005, Jarvis, Rungie et al. 2006, Rungie and Laurent 2012). As consumers tend to overestimate the number of brands they have purchased recently, differences in consumers' real and VR purchase repertoires are likely to be underestimated (Nenycz-Thiel, Beal et al. 2012).

At the whole-of-market level, the similarity of shoppers' real and simulated choices is measured through the percentage difference in the real and simulated shares of the product, pack size, brand or product group, expressed as a percentage of real market share (e.g. 130% of real share or 60% of real share). The real and virtual shares of products are also reported (%) to provide context to percentage results. Percentage results are reported in all tables to facilitate comparisons between groups unequal in size. Statistical tests were conducted on count data. Less than 10 statistical tests were performed for each study, so *p* values without a Bonferroni correction for many comparisons, are reported throughout.

Descriptive analyses will examine the effect of variables individually (e.g. effect of brand type on biases, then pack size). While this approach can result in the effects of variables being under/overestimated or misattributed (Huberty and Morris 1989), modelling can overcome this limitation by considering simultaneously the effect of multiple variables. This will be undertaken in Stage 2 analyses.

The main focus of the descriptive analysis is identifying large or consistent differences. In the Empirical Generalisation (EmpGens) paradigm, patterns or regularities that repeat over a wide range of conditions (e.g., different product categories time periods, researchers, etc.), are known as 'empirical generalisations' and are valued as 'stubborn facts' (Bass 1995, Hubbard and Lindsay 2012).

How descriptive results are reported is demonstrated in Table 13 using data from one product category. Results for the ten largest brands are reported with brands ordered by size; Brands A and B are the top selling brands in this category in the retail context simulated. The results presented in Table 13 and throughout this thesis are unit rather than dollar sales. Validating unit sales (i.e. purchase choices not purchase value) is consistent with previous validation studies (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Desmet, Bordenave et al. 2013, Needel 2013) and ensures that the proportion of *purchases*

affected by biases is reported accurately. Validating to dollar sales would increase the biases for more expensive brands (i.e. premium products) and reduce the biases for cheaper brands.

Table 13: Example of how results are analysed and reported

Ordered by size	Real Share	Virtual Share	Virtual / Real
Brand A	13.9%	11.4%*	82%
Brand B	9.8%	14.7%*	150%
Brand C	8.4%	9.7%	115%
Brand D	4.6%	7.1%*	154%
Brand E	4.1%	2.5%*	61%
Brand F	3.6%	1.4%*	40%
Private Label	3.5%	1.3%*	37%
70+ other brands	…	…	…
Average	1.2%	1.2%	110%

* stat different p < .05

The share of each brand in real stores is compared to its share in the virtual study and the percentage difference is reported.

Analyses report the percentage differences between figures (i.e. Virtual/Real) rather that the absolute difference (i.e. that 12% is 2% *points* larger than 10%). This is a key departure from past studies, most of which used absolute deviations as their key metric (i.e. Nevin 1974, Burke, Harlam et al. 1992, Higaki 2004, Needel 2013).

The Mean Absolute Deviation is not a good measure of bias across product categories because it is *not* robust to differences in the number of products or brands shown. As Table 5 highlights the number of different products (i.e. SKUs) included in each simulation varied substantially across studies; some studies included more than 300 different products while others included less than 30. The number of products shown was dependent on what was appropriate for the product category and retail outlet simulated. As Table 14 shows, there is a clear relationship between the number of brands included and the MAD for each study—in studies with fewer brands the absolute size of deviations between real and virtual results are larger than in studies with more brands (i.e. MADs are larger in the baby care study than the confectionery study).

Table 14: Relationship between number of brands included and av. deviation figures

VR Study	Product Category	Brands *n*	Real/VR MAD	
4	Baby care	6	4.9% pts	
5	Paper towels	6	4.0% pts	} 4.7% pts
3	Feminine care	8	3.6% pts	
1	Gum	10	6.2% pts	
6	Pantry staple	11	3.4% pts	
2	Pet care	20	1.8% pts	} 2.1% pts
1	Confectionery	83	1.1% pts	

The average deviation figure is negatively related to the number of brands included making it unsuitable for cross-study comparisons.

A number of past studies also used the correlation (i.e. Pearson's *r*) between real and VR market shares as a key similarity metric (Burke, Harlam et al. 1992). However correlations can provide a false impression of the accuracy of results (for discussion see McFadden 1980, Mueller and Umberger 2009). For example, previous studies have found that even when the correlation between brands' real and VR market shares was high (i.e. $r = .80$), 40% of brands deviate from their real shares by more than 5 percentage points and 60% of brands deviate by a managerially significant quantity (Burke, Harlam et al. 1992).

Stage 2: Confirmatory modelling

Regression analysis will test whether the variables explored in descriptive analyses such as the price of products and type of brand, have significant effects on the size of VR-Real biases when the impacts of multiple variables are tested concurrently. This stage enables one of the key limitations of univariate descriptive analyses—omitted variables bias—to be addressed (Huberty and Morris 1989, Mauro 1990).

A significance level of $p < .10$ with no Bonferroni correction for the number of tests undertaken will be used to identify variables that may be significant predictors of VR bias. Data from multiple studies will be pooled to enable a large number of variables to be explored. As one of the objectives of this research is to identify variables that should be considered in, and addressed by, future studies, having a Type I error rate in excess of 10% is not of concern. Exact *p* values for each variable and test undertaken will be reported.

In analyses, product category will be modelled in as a variable of interest. In addition to determining if the product category simulated has a significant effect on the magnitude of biases, this will enable data from eight studies to be used to determine if fielding mode and the number of trips simulated are significant contributors to VR bias. This will increase substantially the number of data points available address the research questions about the impact of these variables (RQ4 and RQ5). The sample of brand share (n = 153) and product share (n =1633) estimates available are sufficiently large that this stage of analysis will have sufficient statistical power (e.g. > 80% with $\alpha < .05$) to detect even small effects (e.g. $f^2 < .10$ or .05).

The regression approach will be determined through exploratory analysis of the dependent variables. Given the relatively undeveloped state of VR research (Campo, Gijsbrechts et al. 1999) and the limited number of variables likely to be tested, it is unlikely that the model estimated will be accurate enough that it can be used to calibrate VR results.

Results are interpreted in the paradigm of significant sameness

The broad focus of this thesis is identifying what (if any) differences are consistently present in VR data across product categories, retail contexts and VR studies, an approach consistent with a number of past validation studies (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Higaki 2004, Needel 2013).

To do so, this thesis follows the paradigm and recommendations of *significant sameness* (Nelder 1986, Hubbard and Lindsay 2012, Uncles and Kwok 2013). For example, averages are used in descriptive analyses to highlight the similarity of results across studies, product categories and retail contexts but are not weighted to favour studies with larger samples. Aggregation is used to highlight systematic patterns (Bound and Ehrenberg 1989, Desmet, Bordenave et al. 2013) and reduce the number of small brands, for which sampling errors are expected to be large. This may exacerbate omitted variable biases (Huberty and Morris 1989, Mauro 1990) but is consistent with the approach used in previous validation studies (e.g. Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999) to highlight relevant variables. However, developing deep and robust generalisations, the core objective of this paradigm, typically requires many studies over a considerable period of time (Nelder 1986).

All analyses were conducted in Excel and decimals were maintained to minimise rounding errors, although results are only reported to two or three variable figures to highlight key patterns (Ehrenberg 2000). Modelling was undertaken using RegressIt™ a freely available plugin for Excel.

Limitations of the research approach

Limitation 1: As national data is used, benchmarks include stores with different layouts

Some of the stores from which benchmarking data was obtained would not have had have had the same layout, product assortment or pricing as the VR study. Prices vary nationally for grocery products and compliance with shelf layout plans is known to be a problem—audits of US grocery stores have found that between 40% and 70% of stores deviate in some way from the intended layout and/or assortment of products (Corsten and Gruen 2003), and these difference can have a large impact on the shares of products (Drèze, Hoch et al. 1994, Koelemeijer and Oppewal 1999, Berger, Draganska et al. 2007). Similarly, prior studies have established that on average the prices product vary by up 16% within retail supermarket chains (Competition Commission 2003; 2005).

However, there was no way to eliminate this source of bias without conducting a paired real store and VR study; an approach which would have limited substantially the number of categories and factors that could be studied. However, using nationally aggregated data should limit the impact of this factor. Furthermore, prior studies have established that there is more variation in product ranges, prices, promotions and consumer spending *across* retail contexts than *within* retail chains (i.e. more variation in between Walgreens and Walmart, than within each of these store's outlets; Competition Commission 2003; 2005). Data is reported at the chain/channel level and this limitation is not a major source of concern.

Limitation 2: Data was not restricted to purchases at full price

In real stores approximately 20% of grocery purchases are made on sale.[7] It was not possible to exclude these purchases when obtaining validation data. This introduces a source of difference between the real and the VR purchase data as all of the VR data was collected using the non-sale prices of products.

However, price promotions typically have a short-lived effect on the shares of products (Litvack, Calantone et al. 1985, Dawes 2004). Thus, taken from 12 weeks is likely to reflect the normal, 'non-sale' shares of products reasonably well. Readers concerned about this bias should focus on results for grocery stores and mass merchandisers as they have fewer promotions than US drug stores (pharmacies, Fox, Montgomery et al. 2004).

7 Promotional activity varies across retail outlets (e.g. 12% of sales in Walmart vs 43% in Target are on deal), but 20% is a robust average figure. Across the product categories studied the proportion of real purchases on-deal varied from 21% to 41% (M = 32% , SD = 7%), suggesting the prevalence of promotions in grocery and mass outlets has grown since 2004 (Fox and Montgomery et al. 18% and 14% of sales respectively) or the sub-set of categories studied have particularly high promotional intensities.

Limitation 3: Studies with multiple trips simulate only two shopping trips

Studies that have established the benefits of additional VR trips for purchase *quantity* metrics found that the largest improvements were evident between shoppers' first and second shopping trips, but that up to eight additional trips benefit results (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Garber, Burke et al. 2000). All of the VR studies available for analysis simulated just one or two normal shopping trips. While this should be sufficient to determine if an additional trip can improve purchase *choice* data, a greater number of shopping trips (i.e. four or six) would have been preferable.

Limitation 4: Samples may not be sufficient for differences to be statistically significant for all studies/RQs

As discussed on page 98, with the samples available, differences may not be statistically significant for all of the VR studies or effects tested. It may not, for example, be possible to detect a difference in the sales of organised and unorganised product assortments with the samples available (n=444 and 232), if effect sizes are consistent with past studies (e.g. 5%, Dreze, Hoch et al. 1995). However, as results are expected to be significant for the other two shelf effects investigated for RQ6, this is not a major limitation.

Of more concern is that the majority of the product categories for which data is available are highly fragmented with individual brands often garnering less than 10% of category sales. Consequently, for all but the largest brands, the number of VR purchases is estimated to be less than 35 for the majority of studies (Table 11), a number at which differences in real and VR shares, even if evident, are not likely to be significant. This will be resolved by aggregating results for similar brands in descriptive comparisons and investigating the impact of this variable with regression. Enlarging the VR samples to avoid this issue by including shoppers who saw experimental test layouts was considered but rejected due to the likelihood of this uncontrolled difference affecting results (Sekhon and Titiunik 2012).

Limitation 5: Samples for studies fielded online and in shopping malls differed in their demographic characteristics

One of the limitations of using pre-existing studies is that it is not possible to match precisely the samples used in comparisons. While similar recruitment criteria were used in all studies (see Appendix 4) the samples recruited did differ somewhat (see Appendix 4 and Table 16). These differences were considered in the interpretation of results using a similar approach to past studies (van Herpen, van den Broek et al. 2016). For example, the baby care study fielded in-person (Study 4: Disposable nappies) had more female participants than the baby care study fielded online (Study 8: Disposable nappies), a factor that may have increased the propensity for PL brands to be selected. However, the in-person study also had fewer University-educated participants, a factor that would have reduced the

propensity for PL brands to be selected by a similar amount (Dick, Jain et al. 1995, Ailawadi, Neslin et al. 2001, Baltas and Argouslidis 2007).

In future studies it would be desirable for samples to be more precisely matched, across fielding modes and to the buying population more generally to provide greater confidence in the validity of comparisons.

Limitation 6: All VR studies provided cash incentives

In all of the studies selected participants were provided cash incentives of between US$15 and US$30. Providing cash incentives is standard practice for commercial research projects (Göritz 2006). While providing cash incentives in advance of simulated trips can affect which products consumers, providing them at the end of a study, as these studies did, is less likely to affect purchase choices (Sawyer 1975).

Limitation 7: VR equipment not standard across studies

The studies selected for analysis, particularly those fielded in shopping malls and online, varied in terms of whether shoppers interacted with the virtual environment using a mouse, keyboard, touchscreen or joystick. How shoppers interact with a virtual environment, and the extent to which it feels natural or familiar can influence the extent to which shoppers feel that the environment is real (Witmer and Singer 1998). If navigating a virtual simulation requires learning new commands, the environment may feel artificial until commands become engrained. This does not pose a major problem for the VR data analysed as all of the studies gave shoppers a chance to familiarise themselves with the VR controls before undertaking shopping trips.

Limitation 8: Only data from single category simulations is validated

The number of categories shopped has been shown to affect the number of items participants select and how long they spend shopping (see for example Table 4, Chapter 2), it may also affect what products participants choose. To ensure that analysing data from VR studies where only one or a small number of product categories were shopped did not limit the generalisability of the findings, a study where participants shopped up to eleven product categories is discussed in Appendix 6. The same biases were evident in this study as in the studies reported in Chapter 6 and Chapter 7, indicating that this factor does not limit the generalisability of this thesis' findings.

Limitation 9: Virtual simulations do not reproduce reality perfectly

All of the VR studies selected were studies that looked reasonably accurate and that marketing managers believed could provide accurate data. However, they are by no means perfect replicas of the shopping experience—sounds, smells, tactile sensations and social

interactions were all absent. For this reason, there are expected to be some differences between real and VR results.

Chapter Summary

Compiling a natural experiment from industry data provides scope to determine what biases are consistently present in virtual data across a range of product categories and retail environments and to examine how methodological decisions (e.g. fielding mode, number of trips simulated) affect the validity of VR data. The realism of VR data can be determined retrospectively using tailored sales and survey data.

As it was not possible to retrospectively alter the execution of studies, studies were carefully selected to avoid confounding variables. All of the studies selected are comparable in their visual realism (or better) than studies presented in the literature (e.g. Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Higaki 2004, Tjostheim and Saether-Larse 2005, Waterlander, Jiang et al. 2015, van Herpen, van den Broek et al. 2016) and none were likely to bias towards particular brands, pack sizes of or products because of how they were conducted. In summary the studies selected for validation all:

- Included an appropriate range of products, brands and different size packs for the retail outlet simulated; studies with made-up store brands were not used.
- Used images of real products.
- Restricted shopping to a single product category or area of the store.
- Instructed participants to shop as they 'normally would'.
- Used representative samples of category buyers who were familiar with the retail environment and products simulated.
- Were all fielded in the USA in shopping malls or to online panels.
- Represented a range of product categories, including categories which have not been validated before (chewing gum, feminine hygiene and disposable nappies), and simulated a variety of retail contexts, two of which have not been studied before.

While there are a number of benefits of the research approach, including having multiple sets of data with which to address the research questions and using high-quality, industry relevant data, there are also limitations, most of which arise from using VR data fielded by other researchers in post-hoc analyses and the samples of studies.

CHAPTER 6

STUDY 1: BIASES FOR BRANDS

Overview

Study 1 focuses on how the tendency of consumers to select premium brands in virtual shopping simulations affects the shares of cheaper and more expensive brands, and explores whether these biases are more likely to be result of novelty, social desirability, or sample biases.

Overview of theory

VR validation studies have found that purchases of more expensive brands are typically overestimated in VR studies (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013, van Herpen, van den Broek et al. 2016). Past validation studies have typically described brands as 'more' or 'less' expensive without quantifying the extent of their price differences. In marketing, the term 'premium' is commonly used to describe products that are more expensive than the majority of frequently purchased brands (Nielsen 2016).

Research has established that consumers select premium grocery brands for a range of reasons including:

- superior real or perceived product quality, taste or performance (Rao and Monroe 1996, Sethuraman 2003, Allsopp 2005, Anselmsson, Johansson et al. 2007, Nielsen 2016)
- superior product design (Breetz 2014)
- for the ethical, natural, local or green credentials of the product (Anselmsson, Johansson et al. 2007, Nielsen 2016)
- for their emotional value to impart feelings of success, confidence and self-esteem (Allsopp 2005, Nielsen 2016)
- to signal status, particularly when purchasing products for shared or public consumption (Allsopp 2005, Nielsen 2016).

Research has also found that distinctive packaging and regular advertising reinforce the ability of grocery brands to obtain and sustain price premiums (Anselmsson, Johansson et al. 2007, Steenkamp, van Heerede et al. 2010, Breetz 2014).

The applicability of these drivers is likely to vary across grocery categories. For example, for purely functional products such as paper towels, feminine hygiene products and baby nappies, consumers may be more likely to purchase premium products to access superior product performance or reassurance (Nielsen 2016). In such categories, the real or emotional cost of malfunction may be considered too high to risk using inferior products (Allsopp 2005, Steenkamp, van Heerede et al. 2010, Nielsen 2016). In contrast, in hedonic

product categories, emotional motivations such 'treating oneself' and 'feeling good' are likely to be the most relevant drivers of demand for premium products (Allsopp 2005, Nielsen 2016).

In previous studies, purchases of premium brands may have been overestimated because consumers could enjoy the emotional benefits that premium brand purchases bestow e.g. feelings of indulgence and self-worth (Allsopp 2005, Nielsen 2016) without having to pay for the products they choose. Numerous studies in marketing and economics have shown that that consumers overestimate their willingness to pay and select more premium alternatives when they purchase choices do not impact consumers' own budgets (Sawyer, Worthing et al. 1979, Lusk and Schroeder 2004, Hensher 2010). As discussed in Chapter 5, none of the VR studies examined in this thesis or reported in the literature, required participants to spend their own money when selecting VR products.

Past studies have found that consumers are more likely purchase premium products in categories where their level of category involvement is high (Allsopp 2005). While no validation studies examined this factor, it is possible that consumers who are willing to participate in a VR study for a grocery category are more highly involved than the majority of shoppers. While all of the VR studies examined offered monetary incentives and disguised the product category of interest during recruitment, it is possible that these measures were insufficient to overcome this source of bias.

Prior validation studies typically attributed biases for generic, store and private label products to the effects of social desirability (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013, van Herpen, van den Broek et al. 2016). Social desirability (SD) bias is one of the most pervasive sources of bias in research (King and Bruner 2000, Steenkamp, de Jong et al. 2010). It is motivated by participants modifying their behaviour/responses to conform with social norms (Steenkamp, de Jong et al. 2010), create a good impression with the reseracher (King and Bruner 2000) or consumers wanting to be 'good' or 'helpful' participants by giving researchers the answer they think the resaercher wants (Nichols and Maner 2008). Historically, generic and PL products were marketed as cheap alternatives to national brands and were often percieved to be of poor quality (Bellizzi, Krueckeberg et al. 1981, Croft 1998, Dunne and Narasimhan 1999), so it is concievable these products were avoided in early VR studies for reasons of SD. Social desirablity may also explain why purchases of premium national brands were overestimated in a number of VR studies (e.g. Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013).

Experiments have found that the amount of personal interaction between researchers and participants strongly influences the likelihood of socially desirable responding occurring, and that in-person studies are generally more affected by SD bias than studies fielded online, via mail or over the phone (Colombotos 1969, Krysan, Schuman et al. 1994, Presser

and Stinson 1998, Holbrook, Green et al. 2003, Duffy, Smith et al. 2005, Chang and Krosnick 2010). Many of the VR studies for which brand biases were reported were fielded in person (e.g. Burke, Harlam et al. 1992, Cohen and Gadd, 1996, Campo, Gijsbrechts et al. 1999, van Herpen 2014) although biases have also been reported in studies fielded online (Needel 2013).

The novelty of VR shopping simulations may also affect the incidence of SD responding. Studies of consumer diary panels have found that some consumers over-report purchases of premium brands and under-report purchases of cheaper brands when first participating in purchase panels but that over time as participation becomes less novel, reporting becomes more accurate (Toh, Lee et al. 2006). The same may be true in VR if results are affected substantially by SD responding.

However, some researchers have posited that fast moving consumer goods (FMCG) such as groceries are unlikely to be affected by socially desirable responding (Baumgartner and Steenkamp 2001, Toh, Lee et al. 2006), particularly when respondents are from the US and UK, markets where SD response tendencies are low (Steenkamp, de Jong et al. 2010). Furthermore, SD response tendencies may not affect the selection of PL products in contemporary VR studies where the PL brands are high quality alternatives to the national brands (Nenycz-Thiel and Romaniuk 2011, Nenycz-Thiel and Romaniuk 2014).

The investigations used to explore these possibilities and address the research questions outlined in Chapter 4 are now discussed.

Structured investigations

RQ1.1: Consistency of shoppers' real and simulated price tier choices

If social desirability and other experimental factors do not influence the purchase choices of shoppers in VR studies, then there will be a high level of consistency between shoppers real and VR product choices. However, given the biases apparent in past studies (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013), up to 30% of participants are expected to choose products from price tiers they do not typically purchase.

To address this question the proportion of shoppers who select a product from the same price tier as they normally purchase will be calculated and the choices of different groups compared.

RQ2.1: Is VR purchase data for premium grocery brands always biased?

This investigation seeks to establish if VR purchase data for a range of grocery categories and retail environments is always biased towards the over-selection of premium brands.

How the price premiums of brands were calculated is outlined in the section titled: *Key variables*.

This investigation addresses RQ2 and lays the foundation for additional analyses in this chapter and in Chapter 8.

RQ2.2: Do brands with larger price premiums deviate more from their real market shares than brands with smaller price premiums?

This investigation seeks to understand if brands that retail at a different price points are equally affected by purchase biases in VR studies by comparing if biases vary for brands with large and small price premiums relative to other brands in the category.

RQ2.3: Do up-scaling biases appear to be related to the type of product category?

Previous research has established that price premiums are larger in categories where consumers derive pleasure or enjoyment from consumption i.e. hedonic categories such as chocolate, than in product categories that are more utilitarian in nature (Sethuraman and Cole 1999, Wakefield and Inman 2003). If RQ2.2 is supported and brand with large price premiums deviate more from market share benchmarks than brands with small price premiums, then it is likely that VR purchase data will be more biased in hedonic product categories than in utilitarian product categories.

RQ2.4: Do cheap private label brands deviate more from their real market shares than more moderately priced private label brands?

Contemporary private labels retail at a variety of price points (Kumar and Steenkamp 2007) and use a range of branding strategies (Amrouche, Rhouma et al. 2014, Nenycz-Thiel and Romaniuk 2016).

It is expected that premium PL brands that are comparable in price, quality and/or brand image to well-known national brands will have VR purchase shares similar to their real market shares as recent studies have shown these brands are perceived similarly to national brands and have low rates of avoidance (Nenycz-Thiel and Romaniuk 2011, Nenycz-Thiel and Romaniuk 2016). In contrast, PL brands that are substantially cheaper than their NB competitors are expected to suffer more from SD bias if it is apparent.

This investigation also addresses RQ2 by investigating a variety of PL types including store brands, PLs marketed as generic, commodity products and PL brands marketed under standalone brands names (e.g. Kirkland, Equate). PL products marketed as store brands expected to be more affected by SD bias if it is apparent than PL products marketed as standalone brands (Hurley, Ouzts et al. 2012, ter Braak, Geyskens et al. 2014).

RQ3.1: Are purchase biases a consequence of sampling biases?

If VR samples overepresent demographic groups that select more premium brands in real stores, such as young and high income households (Dick, Jain et al. 1995, Ailawadi, Neslin et al. 2001, Nielsen 2016), or underrepresent demographic groups that select more private label brands (Allsopp 2005, Nielsen 2016), then VR purchase data may exibit biases towards premium and PL brands. Sample biases may occur due to the recruiment criteria specified for VR studies (Leister, Tjøstheim et al. 2007), recruitment approaches used (Bush and Hair Jr 1985, Burke, Harlam et al. 1992, Bruwer, Haydam et al. 1996, Baim, Galin et al. 2009) and/or use of computers in VR studies excluding older shoppers (Yeager, Krosnick et al. 2009, Baker, Blumberg et al. 2010).

A number of demographic and buying behaviours will be examined including:

- age (Dick, Jain et al. 1995, Ailawadi, Neslin et al. 2001, Nielsen 2016),
- household size (Voleti and Sethuraman 2015),
- ethnicity (Palumbo and Herbig 2000) and
- household income (Ailawadi, Neslin et al. 2001),
- category buying frequency (Allsopp 2005).

These are factors that have been shown to affect the propensity of shoppers to purchase premium and PL grocery brands in real stores.

RQ5.1a: Are biases for cheap private label brands less apparent in simulations fielded online?

RQ5.1b: Are biases for premium brands more pronounced in simulations fielded in-person?

The assumption that social desirability is responsible for the biases evident in past studies for premium and private label brands is tested using data from simulations fielded online and in-person. If biases for premium and PL brands in VR studies for FMCG products are a consequence, partially or wholly, of social desirably, then relative to their real shares, the shares of premium brands will be *larger* in studies fielded in-person than in studies fielded online, and the shares of PL brands will be *smaller* in studies fielded in-person than in studies fielded online.

This investigation also addresses RQ5: Is VR data collected online through pre-recruited panels is less generalisable to real purchase situations than VR data collected in specialised research facilities?

RQ4.1: Are biases less apparent in subsequent shopping trips?

If SD responding affects the accuracy of VR purchase data consumers are expected to select more cheap brands and fewer premium brands on their second trip through a VR

environment than on their first, as familiarisation has been shown to reduce the effects of SD responding (Toh, Lee et al. 2006).

This investigation also addresses RQ4: Are multiple shopping trips necessary to collect accurate VR data?

Data sources & method: Overview of pertinent details

This chapter uses simple univariate descriptive analyses to highlight differences in consumers' selection of premium, generic and private label brands and understand what factors should be incorporated into further analysis of the biases in VR studies. It also uses descriptive analyses to explore why differences may be evident and how the fielding of studies may influence the magnitude of biases.

Purchase choice from eight VR studies that met the ten criteria outlined in Chapter 5 is presented is this chapter. Typically data for the subset of brands for which biases should be most pronounced, is presented in this chapter. The characteristics of these studies, and their suitability for analysis, are discussed below.

Virtual shopping data

The studies included in this stage of analyses:

- Included at least one real private label brand and one premium brand. In total 146 brands were included; ten were private label brands. All were cheaper than the average national brand in the category. Some of these PLs were brands marketed as cheap alternatives to national brands (e.g. Kroger's Great Value), whilst some were comparable in price/quality to national brands on offer (e.g. Walmart's Ol' Roy, Costco's Kirkland).
- Included a representative selection of products (SKUs) for the retail environment simulated.
- Located premium and PL products at a range of shelf locations. For example, in the paper towel studies, PL products were located in the centre of the shelf horizontally and vertically; in the baby care studies they were at the start of the aisle on shelves at a variety of heights; in the pet care study they were near the end of the aisle on shelves that would be at eye-level.
- Aimed to recruit samples that were broadly representative of category buyers.
- The in person and online studies compared were similar in regards to the retail environments simulated and variety of brands available. The prices of brands did however differ between studies, necessitating the consideration of price and mode in modelling in Chapter 8.

- One study, the baby care study fielded in person, did however use a joystick designed to mimic the motion of pushing a shopping trolley. All of the other studies used a standard computer mouse. However, this differences is more likely to affect how quickly consumers navigated the VR environments than to affect consumers' purchase choices (Roupé, Bosch-Sijtsema et al. 2014).

Real world benchmarks: Retailer scanner data

Data from Nielsen's Retail Scan database is used to establish the extent of biases for premium and private label brands. As the popularity of private label brands differs substantially across retailers (see Chapter 5), VR data is benchmarked to sales data from the same retail brand/channel as was simulated.

To provide confidence that the VR samples were not biased, data from the Nielsen Homescan panel data, a panel of grocery buyers commonly used by marketers to understand the characteristics and behaviours of buyers (Economic Research Service 2008) is utilised.

Published studies

Preexisting studies in the literature are used to categorise products into (a) those high in hedonic value (e.g. chocolate, chewing gum and pet treats) and (b) products that are high in utilitarian value (e.g. paper towels, feminine hygiene and disposable nappies, Wakefield and Inman 2003, Sethuraman and Cole 1999).

Hedonic–utilitarian classifications of products were drawn from three studies, each of which used slightly different survey measurement approaches[8] but provided consistent results regarding whether product categories were *more* hedonic/utilitarian in the value they provide to consumers. Additionally, information about the size of price premiums and brand shares in hedonic product categories (Sethuraman and Cole 1999, Steenkamp, van Heerede et al. 2010) was used to confirm the relevance of prior studies to the VR studies examined and to categorise pet treats. Comparative results for hedonic and utilitarian product categories are reported in Table 22 and Table 23; price premiums are larger in hedonic categories than utilitarian, as is the scope for bias due to the real shares of premium and PL brands across the categories studied.

While it would have been preferable to have question participants about the hedonic–utilitarian value of the category/categories they shopped[9], this was not possible as the data was sourced from industry studies already conducted. An alternative would have been to

8 Wakefield and Inman (2003) is superior to Sethuraman and Cole (1999) for discriminating between products (e.g., is chocolate more hedonic than chewing gum, but the approaches provide consistent categorisations of the dominant value for product categories.

9 To avoid influencing shoppers' behaviour, questioning would have had to have been undertaken after the shoppers made their VR product choices.

survey a representative sample of US consumers and generalise results to the samples analysed. However, as this thesis is primarily exploratory—aiming to explore how consistent biases are across categories and identify factors that *may* influence the size of biases so they could be examined more thoroughly in future studies—this was not felt to be necessary.

Key variables

The following variables were used to investigate the extent of biases and their causes in this chapter:

Price Premium (%): the percentage difference between the price of a brand and the average price of all NBs included in the study. This is similar to the price index method used by Steenkamp et al. (2010) and Eales (2014) and provides a consistent method with which to estimate the magnitude of premiums across categories. PL brands were not included when calculating average prices as this would have classified some mid-tier NBs as premium.

$$\text{Price premium (\%)} = \frac{\text{Average price of Brand A} - \text{Average price of all brands in VR}}{\text{Average price of NBs brands in the VR study}}$$

In industry studies 'premium' often refers to products that are 20% more expensive than the average brand or product (Stilinovic 2016). However excluding products that are 5 or 10% more expensive, would have reduced the number of categories for which results could be reported as price premiums are typically small in utilitarian product categories (Steenkamp, van Heerede et al. 2010). Moreover, systematic differences are expected to be evident in biases for products with large and small price premiums.

The relative prices of PL brands are reported using the same formula as outlined above. This provide insight how the r*elative* pricing of PL and store brands affects their selection in VR studies.

Real Share (%): the proportion of total category sales (units chosen) for each brand. Why unit sales were used is discussed on page 91 of Chapter 4.

Virtual Share (%): the proportion of simulated purchases for each brand (unit sales).

Virtual: Real the difference between the share of a brand in a virtual simulation and its market share in retailer scanner data, expressed as a percentage of its real market share. This makes differences in shares easily relatable to differences in prices (e.g. brands at a 20% price premium were over-purchased by 25% on average). In some tables this variable is referred to as *Bias* or *V/R.*

Statistical tests

Chi-squared tests (X^2) are used throughout this chapter to identify significant differences between real and VR premium and private label purchase shares. The expected number of purchases was calculated using purchase benchmarks from real stores and tests were generally conducted with one degree of freedom (i.e. purchased target brand or did not). Where the less than five shoppers selected or were expected to select a brand, an adjusted version of Fisher's exact conditional test was used as it aids in estimating the likelihood of results being significant. Barnard's unconditional tests were also performed as spot checks and did not provide substantially different results. Chi-squared tests are also used to test for differences between sample and buying populations, to compare selection rates on shoppers' first and second VR shopping trips and results for studies that depicted different retail environments.

Overlapping confidence intervals (C.I.) are used to determine if the magnitude of biases was different across studies, product groups and different types of product categories. T-tests are used for continuous variables. Statistical tests were performed using www.graphpad.com.

Results: Consistency of premium brand bias across studies

Table 15 reports the real and simulated purchase shares of premium grocery brands in six VR studies that were fielded in person.

Table 15: Purchase shares of premium grocery brands in VR studies

Virtual Study	Environment simulated	Product category	Real Share	Virtual Share	Virtual / Real
6	Grocery store n=253	Pantry staple	13.2%	17.5%	133%
2	Mass retailer n=131	Pet care	39.7%	51.0%*	128%
5	Grocery store n=359	Paper towels	31.1%	38.4%*	123%
3	Mass retailer n=155	Feminine care	45.2%	55.4%*	123%
1	Grocery store n=1500	Confectionery	51.3%	62.3%*	121%
4	Grocery store n=359	Disposable nappies	58.6%	65.1%*	111%
5	Mass retailer n=229	Paper towels	38.8%	42.2%	109%
3	Pharmacy n=155	Feminine care	52.7%	56.6%	107%
	Average grocery study not weighted, total n=2442		38.6%	45.8%*	122%⁺
	Average mass study n=515 purchases		41.2%	49.5%*	120%⁺
	Average pharmacy study n=155 purchases		52.7%	56.6%	107%

*Statistically different to real share given sample, $p < .05$ ⁺Statistically different to bias for pharmacy stores, $p < .05$

In virtual shopping simulations the purchase shares of premium brands are consistently overestimated; the size of the bias ranges from 7% to 33%.

In Table 15, X^2 tests were used to compare observed to expected counts of premium brand purchases. Differences were highly significant ($p < .01$) for the majority of studies. Statistical results for each study are reported in Table 60 of Appendix 7. As discussed, *Premium* refers to all brands that were more expensive than the average NB brand included in the study.

Across the six studies validated, biases for premium brands ranged from 7 to 33% greater than the products' real shares (Table 15: *VR/Real)*. This is consistent with the extent of biases reported in past studies; Higaki (2004) for example reported that purchases of premium beers were overestimated by 35% and Campo et al. (1999) found that the non-promotional shares of high-quality jams were overestimated by 29%.

Interestingly, the shares of premium brands were not significantly different in the different retail environments tested in the paper towel and feminine care studies reported in Table 15 (Paper towels, X^2 (1, n = 591) = .77, p = .38; Feminine care, X^2 (1, n = 310) = .01). This is not entirely surprising as the real shares of these products were also similar (e.g. expected share of premium paper towels in mass environment 38.8% vs. 31.1% in grocery environment).

Regression analysis will, if possible, investigate if the magnitude of biases varied across retail outlets when multiple variables, including price, are considered simultaneously. As discussed in Chapter 5, focusing on a single variable can provide misleading results, so results in this chapter should be considered preliminary.

Results: Causes and extent of upscaling biases

RQ3.1: Biases are not likely to be a result of the samples recruited for VR studies

Table 16 compares the characteristics of one of the VR samples for which results were reported in Table 15 (Study 1), to the real buying population for the category.

Table 16: Validation of the sample for VR Study 1, Confectionery

	VR Sample n=400	**Buying Pop'n** n = 28,698	**Statistical Results**
Aged 18-34 yrs	47%	17%	Significantly diff. distribution
35-44 yrs	25%	17%	X^2(5, n=400) = 317, p < .001
45-54 yrs	10%	30%	
55-64 yrs	9%	21%	
65+ yrs	9%	15%	
HH 1 person	13%	17%	Significantly diff. distribution
2 people	37%	35%	X^2(4, n=400) =22, p < .001
3-4 people	42%	34%	
5+ people	8%	14%	
HH Income US$0-$20,000	11%	10%	Significantly diff. distribution
US$20-$39,000	32%	23%	X^2(5, n=400) =89.1 p < .001
US$40-$69,000	38%	26%	
US$70-$99,000	16%	27%	
US$100,000 +	3%	14%	
White	75%	70%	Significantly diff. distribution
Black	11%	10%	X^2(5, n=400) =16.4 p < .05
Hispanic/Latino	10%	11%	
Asian	2%	2%	
Other	2%	7%	

A number of statistically significant differences were evident between the VR sample and the buying population including age, household size and income.

The results in Table 16 demonstrate that the VR study overrepresented younger buyers under the age of 35, a group predisposed to selecting premium grocery brands (Dick, Jain et

al. 1995, Ailawadi, Neslin et al. 2001, Nielsen 2016). This may have been due to the demographic profile of the shopping malls in which the VR studies were conducted or the recruitment tactics of research agency responsible for fielding this study (Bush and Hair Jr 1985, Hornik and Ellis 1988).

When reweighted according to the buying population, the overall purchase share of premium brands was 61.5%, a proportion that is not statistically or managerially different to the unweighted result reported in Table 15 (VR = 62.3%, confidence interval extends from 59.9% to 64.8% at a 95% C.L.) This was because the VR purchase share of premium confectionery brands did not vary significantly across five age groups (X^2 (4, n = 400) = 1, p = .90, share between 59 and 66% for all age groups, see Table 61 for specific proportions).

It was not possible to reweight the sample to resolve the other biases observed in Table 16 due to issues with the data files provided. The sample did however also overrepresent middle income households, bias that *ceteris paribus* should have predisposed the sample towards an underrepresenting premium brand purchases (Ailawadi, Neslin et al. 2001).

To further validate that sample biases were unlikely to be responsible for the biases observed, a second study is examined in Table 17. This study asked shoppers how frequently they purchase the category, which brands they had purchased in the last month and a range of other demographic and behavioural questions. Unfortunately, it was not possible compare some of this data to the buying population as the response categories used in the VR study differed from those used by Nielsen to record buyers' characteristics. It was however possible to examine some characteristics of relevance to the selection of premium products in real stores including household size and category involvement (Allsopp 2005, Voleti and Sethuraman 2015). These are reported in Table 17.

Table 17: Validation of the sample for VR Study 5, Paper towels

	VR Sample n=233	**Buying Pop'n** n = 27,427	**Statistical Results**
HH is 1-2 people	41%	42%	Distribution not sig. different
3-4 people	48%	44%	X^2 (2, n = 233) = 2.24, p > .30
5+ people	11%	14%	
M category purchases /mth	1.93 (*StD* 1.44)	1.99 (*StD* N/A)	Not sig. different, 95% CI of VR 1.7 – 2.1

The VR sample was representative of the buying population for the characteristics that could be validated.

The results in Table 17 demonstrate that this VR sample was representative of the buying population in regards to household size (X^2 (2, n = 233) = 2.24, p > .30) and category buying rate (C.I. of VR result overlaps at 95% C.L.).

Other demographic characteristics could not be validated but in Table 30 the VR purchase choices of different demographic and behavioural groups are reported to determine if groups that have higher rates of premium brand purchasing in real stores (Dick, Jain et al. 1995, Ailawadi, Neslin et al. 2001, Voleti and Sethuraman 2015), also had higher rates of selection in VR. In Table 18 results for three price tiers of brands are presented: Cheap, MidPriced and Premium. How brands were categorised is outlined in Table 19.

Table 18: VR purchase choices of different demographic groups – price tiers

Paper towel study	Cheap	MidPriced	Premium	Statistical results VR compared to Real
Real Share	28%	35%	37%	
Aged 18-29 yrs	13%	52%	35%	X^2 (2) n = 103, = 17.8, p < .001
30-39 yrs	10%	45%	45%	X^2 (2) n = 108, = 17.4, p < .001
40-49 yrs	15%	55%	30%	X^2 (2) n = 133, = 24.7, p < .001
50-59 yrs	6%	52%	42%	X^2 (2) n = 82, = 21.6, p < .001
HH is 1-2 people	11%	56%	34%	X^2 (2) n = 203, = 47.3, p < .001
3-4 people	11%	45%	44%	X^2 (2) n = 172, = 24.7, p < .001
5+ people	10%	60%	30%	X^2 (2) n = 51, = 16.5, p < .001
HH Income < $40,000	10%	55%	35%	X^2 (2) n = 181, = 42.5, p < .001
$40-$74,999	11%	44%	44%	X^2 (2) n = 204, = 28.2, p < .001
$75,000 +	20%	50%	30%	X^2 (2) n = 41, = 4.8, p = .09
Purch. catg. Weekly	16%	61%	23%	X^2 (2) n = 75, = 22.8, p < .001
Fortnightly	12%	51%	38%	X^2 (2) n = 181, = 28.9, p < .001
Monthly	9%	47%	44%	X^2 (2) n = 134, = 24.6, p < .001
Every 2-3 mths	10%	58%	33%	X^2 (2) n = 30, = 7.6, p = .02

The VR choices of all groups differed significantly from real purchase shares.

Contrary to expectations, the distribution of Premium, Midpriced and Cheap brand purchasing did not differ substantially across income (X^2 (2, n = 421) = 8.3 p = .08), age (X^2 (2, n = 427) = 9.1 p = .16), household size (X^2 (2, n = 428) = 6.8 p = .15) or buying frequency groups (X^2 (2, n = 422) = 1.4 p = .11). Furthermore, the X^2 statistical results in Table 30 demonstrate that the VR purchase choices of all demographic and behavioural groups deviated substantially from purchase benchmarks for the overall market. This indicates that all demographic groups contributed towards the biases for premium and PL brands in the study.

RQ1.1: Consumers who typically purchase cheap and mid-tier brands upscale their purchases

In Table 19 the proportion of shoppers who selected a product from the same price tier in the VR study as they claimed to have purchase in the last month in the VR recruitment survey is reported. Data from the same study as was reported in Table 17 and Table 18 is reported. In this study participants not explicitly instructed to purchase the same brand or products as they usually buy, only to 'shop as they normally would'. Results are presented cautiously as it is possible that SD bias affected the reporting of (real) premium and PL brand purchasing. Indeed, when results were benchmarked there was found to be a significant difference in the proportion of the sample that indicated they had purchased cheap, mid-priced and premium brands recently and the penetration of these brands (X^2 (2, n = 233) = 19.4, p < .01). Comparisons found that fewer shoppers than expected reported purchases of cheap brands (22% vs. 30%) and mid-tier brands (32% vs. 35%) and more shoppers than expected reported purchases premium brands (61% vs. 42%) in the sample for the VR study.

Table 19: Individual level consistencies in real-virtual brand choices

Brands purchased in the last month	**Sample** n=233	**Repeat Rate** Selected in VR
Cheap < $5.50 per pack	22%	43%
Mid-priced $5.50 to $7.00 per pack	32%	44%
Premium > $7.00 per pack	61%	53%
Total sample	115%	49%

Less than half of the consumers who claimed to have purchased mid-tier or cheap brands in the last month selected brands from these price tiers in VR.

The *Repeat Rate* results and statistics in Table 19 demonstrate that the level of VR purchase consistency was higher for shoppers who reported purchasing premium brands in real stores than shoppers who reported purchasing cheap (X^2 (2, n = 142) = 5.6, p = .02) or mid-priced brands (X^2 (2, n = 142) = 4.5 p = .04). Overall, only 49% of shoppers selected products from the same price tier as they reported purchasing in the last month in VR, fewer than was expected based on prior studies (Tjostheim and Haugland 2005, Jarvis, Rungie et al. 2006).

Additional analysis was undertaken to determine which if any demographic groups were more likely to deviate from heir claimed brand purchases. Repeat rates were found to be particularly low for private label products, a product reported with *Cheap* brands in Table 19.

Only 20% of the consumers who reported buying private labels recently (n = 15) selected a private label product when shopping the virtual simulation. It is possible that the repeat rate was low for this group as shoppers were asked to in the recruitment questionnaire if they had purchased store brands in any outlets in the last month but in the VR study only the store brands available in the outlet simulated were shown. This was a store brand bearing the retailer's name but prior studies have noted that consumers are not good at imagining a single store brand to be a representative of all the store brands they might buy (Needel 2013).

Repeat rates were also significantly lower for consumers responsible for buying less than 75% of their household's paper products (34% *Repeat Rate* vs 55% for main HH shoppers; X^2 (2, n = 239), = 11.78, p = 0.003). However, this group constituted just 10% of the sample so could not be responsible for all of the bias observed. As shopping responsibilities correlate with gender—in most grocery categories women outnumber men as the primary household shoppers (Nielsen 2014)—*Real/VR* repeat rates were somewhat lower for male participants (*RR* = 41% vs 59% for women, $p < .10$).

RQ2.2: Deviations are larger for more expensive premium brands than less expensive brands

In Table 20, results for two tiers of brands are presented:

Brands with small price premiums of between one and 19% more expensive than the average brand in the study (n=19, $M_{premium}$ = 110%, StD = 6%),

Brands with large price premiums of between 20 and 60% more than the average brand in the study (n=13, $M_{premium}$ = 145%, StD = 15%).

Table 20: Results for brands with large and small price premiums

	Small price premiums (1 – 19%)				Large price premiums (20 - 60%)			
	Brands *n*	Real Share	Virtual Share	Virtual Real	Brands n	Real Share	Virtual Share	Virtual Real
Study 1	11	10.7%	12.4%*	116%	9	40.6%	49.9%*	123%
Study 2	5	35.1%	44.4%*	125%	3	4.6%	6.6%$^{+}$	143%
Study 3	3	37.4%	38.1%	102%	1	15.3%	18.5%$^{+}$	139%
Average	-	27.7%	31.6%*	114%	-	20.2%	25.0%*	124%

* Statistically significant at $p < .05$; $^{+}$ $p < .10$

The results in Table 20 demonstrate that VR purchase shares of brands with small price premiums were overestimated by 14%, whereas the VR shares of brands with large price premiums were overestimated by a greater amount on average, 24% (confidence intervals do not overlap at a 95% C.L.).

That *more* expensive premium brands were *more* compromised by upscaling biases than relatively cheaper brands, is consistent with prior studies—Higaki (2004) for example found that the VR purchase shares of beers with large price premiums were consistently (i.e. always) overestimated while the shares of beers with modest price premiums were typically, but not always, overestimated by a lesser amount. This pattern is observed in the brand level results reported in Table 21, although a number of differences are not statistically significant when shares are unaggregated due to the samples sizes for Study 2 and Study 3.

The correlation between the price premiums and *VR/R* biases reported in Table 21 (n = 25) is moderate ($r = .52$, $p < .01$), suggesting that further investigation of price as an explanatory factor is warranted.

Table 21: Biases for premium brands in virtual simulations: Brand level results

	Price Premium	Real Share	Virtual Share	Virtual Real
Study 1 All premium brands		**56.0%**	**62.3%***	**111%**
Brand AK	+ 60%	1.1%	2.7%*	245%
Brand B	+ 55%	9.8%	14.7%*	150%
Brand A	+ 35%	13.9%	11.4%*	82%
Brand G	+ 29%	2.8%	3.3%	118%
Brand C	+ 27%	8.4%	9.7%	115%
Brand D	+ 20%	4.6%	7.1%*	154%
Brand U	+ 20%	1.1%	1.5%	136%
Brand O	+ 18%	1.5%	1.9%	127%
Brand H	+ 15%	2.8%	2.7%	98%
Brand AJ	+ 7%	0.6%	1.1%	183%
Brand G	+ 3%	1.8%	0.9%	50%
Brand E	+ 2%	3.6%	1.4%*	40%
	+ 8 other brands			
Study 2 All premium brands		**39.7%**	**51.0%**	**128%**
Brand H	+ 54%	3.6%	5.5%	153%
Brand O	+ 34%	0.3%	0.5%	167%
Brand N	+ 28%	0.7%	0.6%	86%
Brand D	+ 19%	5.3%	5.5%	105%
Brand C	+ 14%	13.4%	22.6%	169%
Brand A	+ 12%	15.4%	20.5%	133%
Brand B	+ 6%	3.5%	1.2%	116%
Brand M	+ 6%	1.1%	1.2%	109%
Study 3 All premium brands		**52.7%**	**56.6%**	**107%**
Brand B	+ 36%	15.3%	18.5%	121%
Brand D	+ 13%	9.0%	9.0%	132%
Brand A	+ 12%	26.5%	26.5%	93%
Brand F	+ 10%	1.9%	2.6%	118%

* Statistically significant at $p < .05$; $^{+}p < .10$

The VR shares of all of the brands with price premiums in excess of 19% exceeded their real shares; the VR shares of most but not all of the brands with price premiums of less than 20% exceeded their real shares.

RQ2.3: Hedonic product categories are more affected by biases for premium brands than utilitarian product categories

In Table 22 and Table 23, VR results for the most premium brand included in each study, the brands for which biases should be most pronounced, are benchmarked against their real

market shares for three hedonic and three utilitarian product categories. Maximum price premiums for each category are also reported. These are consistent with past studies such as Wakefield and Inman (2003) which found price premiums are larger in hedonic product categories.

Table 22: Results for premium brands in hedonic product categories

Most expensive product		Price Premium	Real Share	Virtual Share	V.Share Diff'nce	Virtual/ Real
Chocolate	Brand AK	160%	1.1%	2.7%*	+ 1.6%	245%
Chewing gum	Brand E	127%	8.3%	25.3%*	+ 17.0%	305%
Pet treats	Brand B	136%	15.3%	18.5%	+ 3.2%	121%
Total for hedonic catg.		140%	24.7%	46.5%	+ 21.8%	188%

*statistically different from real share, $p < .05$

In the hedonic product categories, VR sales estimates for premium brands deviate substantially from real benchmarks; VR share of Brand AK, substantially larger than expected ($X^2(1, n = 1500)$ 14.44, $p < .001$), VR share of Brand E substantially larger than expected ($X^2(1, n = 87)$ 13.70, $p < .05$).

The results in Table 22 demonstrate that in hedonic product categories, VR sales estimates for premium brands deviate significantly from real purchase estimates and that the magnitude of biases is substantial—88% on average across the three categories examined (StD = 93%).

Table 23: Results for premium brands in utilitarian product categories

Most expensive product		Price Premium	Real Share	Virtual Share	V.Share Diff'nce	Bias Virtual / Real
Feminine care	Brand C	129%	15%	19%	+ 3.8%	125%
Paper Towels	Brand A	124%	31%	38%*	+ 7.0%	123%
Baby nappies	Brand B	121%	24%	34%*	+ 10.5%	144%
Total for utilitarian catg. n = 1718		125%	75%	88%	+ 4.3%	131%
Total for hedonic catg. n = 743		140%	25%	47%	+ 21.8%	188%

*statistically different from real share at $p < .05$

In the utilitarian product categories, VR sales estimates for premium brands deviate from real benchmarks.

The results in Table 22 demonstrate that in utilitarian product categories, VR sales estimates for premium brands deviate significantly from real purchase estimates (Brand A, $X^2(1)$, $n = 229$) 4.75, $p < .05$, Brand B, $X^2(1)$, $n = 330$) 16.0, $p < .001$ but that biases are typically smaller—25% on average across the three categories examined. This average does not overlap with the C.I. of the average bias for hedonic product categories, 37% to 44% bias. This is consistent with expectations and indicates that modelling should consider the category type in addition to the size of price premiums.

RQ2.4: Price appears to influence the extent to which private label brands are under-represented

Validation results for private label brands that retail at large discounts relative to their competitors are reported in Table 24. Validation results for private label brands that are more comparable in price to national brands are reported in Table 25. In these tables, whether each brand is marketed as a store brand with the retailers' brand name (e.g. Kroger's Great Value), a standalone brand with a separate brand name (e.g. Kirkland, Equate) or a generic product with no brand name, is also reported. Pack images were shown in the VR study; consumers did not need to recall the names of PL brands to select them as simulated purchases.

Table 24: Results for large-discount PL brands in virtual studies

Private label brand in:		Branding	Price vs NBs	Real Share	Virtual Share	Virtual / Real
Mass retailer n= 131	**Pet care**	Generic	55%	4%	1%	19%
Grocery store n=1500	**Confectionery**	Store brand	60%	3%	1%**	33%
Mass retailer n=350	**Baby nappies**	Standalone brand	63%	35%	9%**	26%
Grocery store n=179	**Pantry staple**	Store brand	67%	18%	4%**	22%
Average large-discount PL n=2160			61%	15%	4%**	25%

** Statistically significant at $p < .01$given sample size

Sales of large-discount private labels are underestimated by up to 81% in VR simulations.

In Table 24 differences between the real and VR shares of PL brands were highly significant (p <.01) for the majority of studies (confectionery, nappies and rice). The difference between the real and VR share of the generic pet care product in Table 24 was not significant, in part to the small sample size for this study (n = 131 purchases). Additional statistics for the tests undertaken for these brands can be found in Table 62, Appendix 7.

Table 25: Results for small-discount PL brands in virtual studies

Private label brand in:		Branding	Price vs NBs	Real Share	Virtual Share	Virtual / Real
Mass retailer n=155 purchases	**Fem. care**	Standalone brand	95%	14%	10%$^{+}$	71%
Pharmacy n=155	**Fem. care**	Store brand	82%	22%	11%**	50%
Grocery store n=359	**Paper towels**	Store brand	81%	28%	15%**	54%
Mass retailer n=131	**Pet care**	Standalone brand	78%	16%	3%**	20%
Grocery store n=330	**Baby nappies**	Store brand	76%	23%	10%**	43%
Mass retailer n=229	**Paper towel**	Store brand	74%	25%	7%**	25%
Average smaller discount PL n=1359			81%	21%	9%**	45%
Average large discount PL n=2160			61%	15%	4%**	25%

** Statistically different to real at $p < .01$; *$p < .05$ $^{+}p < .10$

Sales of smaller discount PL products are underestimated by smaller amounts on average than sales of large discount PL products; the confidence intervals of results do not overlap.

In Table 25, differences between the real and VR shares of PL brands were highly significant (p <.01) for the majority of studies (feminine care, paper towels, pet care, nappies etc.). Additional statistics for the tests undertaken be found in Table 62, Appendix 7.

The *Average* results in Table 25 suggest there may be a systematic pattern between the prices of PL brands and the extent to which purchases are underestimated in VR studies. On average large-discount private labels brands had just 25% of their real share in VR studies while small-discount private label brands had 45% of their real share, a statistically larger proportion when the confidence intervals for these results were compared (95% C.L, 23% to 27%).

Although not tested statistically, in Table 24 and Table 25 there does not appear to be a relationship between how PL brands are marketed—generic, standalone or store brand—and the magnitude of the bias. For example, in Table 25, sales of the feminine care product marketed under its own brand name were underestimated by less than a third (29%) whereas the sales of the pet care product marketed under its own brand name in the same retailer were underestimated by 80%. On average, private label products marketed under their own brand names (n = 3) garnered 38% of their real share (StD = .13) and private label products marketed as store brands (n = 5) garnered 39% of their real shares (StD = .28, confidence intervals likely to overlap). Analyses in Chapter 8 will investigate if how a PL

brand is marketed affects the magnitude of biases when considered as a factor independently of price.

RQ5.1: The scale of biases is similar in studies fielded online and in-person

In Table 26 and Table 27 studies that were fielded online and in-person are compared to determine if the mode of fielding may affect the accuracy of VR purchase data. In Table 26 and Table 27 the Virtual/Real metrics are most comparable across studies as the Real Shares of some brands grew in the two-year interval between the fielding of the in person and the online studies. The statistical significance of Real/VR differences is indicated in Table 24 and Table 25 and additional statistics are reported in Table 63 of Appendix 7.

Table 26: Results for PL brands in simulations fielded in person and online

Results for brand when:		Fielded In-Person			Fielded Online		
		Real Share	Virtual Share	Virtual / Real	Real Share	Virtual Share	Virtual / Real
PL Brand	Fem. care	22%	11%**	50%	26%	13%**	53%
PL Brand	Baby care	23%	9%**	30%	27%	17%**	63%
Average		23%	10%	44%	27%	15%	57%
Average	Real - Virtual		- 12.5% pts			- 11.5% pts	-17%

**statistically different to real share $p < .01$

Relative to their real shares at the time of each study, private label brands are *less* underrepresented when virtual studies are fielded online.

The results in Table 24 indicate that in VR studies fielded online the shares of PL brands are substantially underestimated compared to their real market shares (Feminine care study, X^2 (1, n=330) = 62.26, $p < .01$, Baby care study, X^2 (1, n=200) = 17.1, $p < .01$).

PL shares were larger in the VR studies fielded online than the studies fielded in person (i.e. 57% of real share vs. 44%, C.I. do not overlap at 95% C.L.). This finding suggests that social desirability bias, a bias that is influenced by the mode of fielding (Colombotos 1969, Krysan, Schuman et al. 1994, Presser and Stinson 1998, Holbrook, Green et al. 2003, Duffy, Smith et al. 2005, Chang and Krosnick 2010), may be responsible for some of the biases observed in VR studies. It was hypothesised that if biases for premium and PL brands in VR studies were a consequence of SD bias, then relative to their real shares, the shares of premium brands would also be larger in studies fielded in-person than in studies fielded online. However, the results in Table 27 are not consistent with this finding.

Table 27: Results for premium brands in simulations fielded in person and online

Results for:	Fielded In-Person			Fielded Online		
	Real Share	Virtual Share	Virtual / Real	Real Share	Virtual Share	Virtual / Real
Brand B Feminine care	15.3%	18.5%	N.S. diff	14.4%	18.1%	N.S. diff
Brand A Baby care	29.4%	31.6%$^{+}$	107%	22.4%	28.4%$^{+}$	127%
Average	22.4%	25.0%	112%	18.4%	23.3%	126%
Average Real - Virtual		+ 2.7% pts			+ 4.9% pts	+12.5%

*statistically different to real share $^{+}p < .10$

The shares of premium brands are sometimes overrepresented in virtual studies fielded online and in person; when present, the bias is slightly larger when simulations are fielded online.

Aggregated results from two product categories suggest that that the VR shares of premium brands are larger in VR studies fielded online than in studies fielded in person (i.e. +26% vs. +12%, confidence intervals of results do not overlap at 95% C.L.). This is the opposite of what would be expected if social desirability was contributing to biases in VR studies.

However, differences between the real and VR shares of brands were only marginally significant ($p < .10$) for one of the categories examined (Baby care) when results from each study were considered independently (In person, X^2 (1, n=340) = 2.7, $p < .10$; Online, X^2 (1, n=200) = 3.5, $p < .10$). Differences in the real and VR shares of the feminine care study compared in Table 27 were not significant for either the in person study (X^2 (1, n=155) = 1.1, $p > .10$) or the online study (X^2 (1, n=330) = 0.66, $p > .10$). However, samples also had insufficient statistical power to be confident there was not a difference in the shares of Brand B across modes (<10% power for $\alpha = .05$). In the in person baby care study reported in Table 26 and Table 27, participants shopped either a grocery store (n=359 purchases) or a mass store environment (n=340 purchases). In Table 26 and Table 27, only results from the mass environment are reported as the online study also simulated a mass retailer. Utilising the samples from both environments was considered as Table 15 suggested that the retail environment simulated may not affect VR purchase choices substantially. However, doing so did not improve statistical power of comparisons (e.g. 18% power for $\alpha = .05$). Multivariate modelling in Chapter 8 enable the impact of fielding mode on the magnitude of VR biases to be determined with more statistical power.

RQ4.1: Additional trips have a marginal impact on purchase biases

In Table 28 results from VR studies where participants made two experimental shopping trips through the same environment are reported. The results of tests undertaken to determine if purchase shares were statistically different on shoppers first and second trips are also reported. A broader range of brands are reported than in Table 27 as only marginal differences were found in this table when only the most premium brand in each category was considered. The samples for both studies had sufficient statistical power (>85% at *α* <.05*)* to detect correctly differences in the purchase shares of premium and mid-tier brands, presuming the difference was consistent in magnitude with past studies (i.e. <20%, Campo, Gijsbrechts et al. 1999, Edwards and Gangadharbatla 2001, Kleinschmit, Reid et al. 2012).

Table 28: Shares of price tiers on consumers' first and second trips

	Brands	Real Share	Virtual Share 1st trip	Virtual Share 2nd trip	Statistical Results
Baby care study	**Premium**	59%	65.1%	63.5%	Distribution not sig. different
	Mid-tier	18%	28.0%	26.7%	X^2 (n = 700) = 1.58, p = .45
	PL	23%	6.90%	9.80%	
Confectionery	**Premium**	51%	62.3%	61.0%	Distribution not sig. different
	Mid-tier	46%	36.4%	37.4%	X^2 (n = 3000) = .78, p = .67
	PL	3%	1.30%	1.60%	
Average impact				2% pts	

The shares of brands from different price tiers where not significantly different on consumers' first and second VR trips.

The results in Table 28 demonstrate that the proportions of premium, private label and mid-tier brand purchases were not significantly different on shoppers' first and second VR trips.

Additionally, the results in Table 28 demonstrate that simulating a second shopping trip did not overcome biases towards the selection of premium and PL brands in VR studies. In both studies the distribution of premium, private label and mid-tier brand purchases deviated

significantly from real sales benchmarks (Baby care study: $X^2(2)$, n = 350) 43.2, $p < .001$, Confectionery study $X^2(2)$, $n = 1500$) 63.3, $p < .001$).

While a small decrease in the purchase shares of Premium brands was apparent in both studies on shoppers' second trips (Table 28: Baby care 3% reduction in share from 65.1% to 63.5%, Confectionery 2% reduction in share from 62.3% to 61.0%), this difference was not statistically significant. Whether additional shopping trips improve the accuracy of VR shopping data will be further explored in Chapter 8.

Chapter Summary

This chapter demonstrates that the shares of premium brands were consistently overestimated in VR studies, even when samples were weighted to represent the buying population. As discussed in Chapter 5, all of these studies displayed realistic prices and packages and instructed regular category buyers to 'shop as they normally would'. Biases for premium brands were substantial, especially in hedonic product categories where price premiums were in excess of 25% for some brands.

Analyses confirmed that purchases of contemporary PL brands were consistently underestimated in VR studies irrespective of whether products were marketed as generic/store brands or independent brands comparable in price to NBs.

Additional analysis of demographics from two studies found that buying groups that typically favour premium brands in real stores were not more likely than other groups to select premium brands in VR studies. Additionally, the VR shares of premium and mid-priced were found to exceed real shares across almost all demographic and buying groups validated. Moreover, less than half of all participants (49%) select a brand from the same price tier as they claim to typically purchase in real stores, a lower rate of consistency than observed in past studies (Tjostheim and Haugland 2005).

Overall the results reported in this chapter suggest that the research setting (i.e. hypothetical VR) and not the samples recruited were responsible for the biases observed. Moreover, that differences were not found across demographic groups suggests that the economic concerns that inform shoppers' purchase choices in real stores may not influence shoppers' VR purchase choices in hypothetical studies, a finding that would be consistent with prior studies (Sawyer, Worthing et al. 1979, Campo, Gijsbrechts et al. 1999, Lusk and Schroeder 2004, Hensher 2010).

Descriptive comparisons for private label and premium brands suggested there may be a systematic relationship between the prices of products and the extent to which purchase shares are under or overestimated in VR studies. This may be because selecting premium brands and avoiding cheaper brands, including PL brands, in hypothetical experiments

provides consumers with emotional benefits (e.g. feelings of self esteem, positive self image etc., Sawyer 1975). It is also possible that consumers selected more premium brands because the packages of these products attracted more attention (Anselmsson, Johansson et al. 2007, Steenkamp, van Heerede et al. 2010, Breetz 2014) and the higher prices of these products did not deter shoppers because the VR experiments did not have any purchase consequences (Sawyer, Worthing et al. 1979, Lusk and Schroeder 2004, Hensher 2010).

The nature of the relationship between price and bias will be further analysed in Chapter 8 where variables will be considered simultaneously for their impact(s) on VR bias.

Other potential explanations will be explored in Chapter 7 as there was not strong evidence for social desirability bias affecting results. While the results in this chapter indicated that shoppers were less averse to selecting PL brands in VR simulations fielded online than in simulations fielded in person, online fielding and simulating additional shopping trips did not improve substantially or eliminate biases towards the selection of premium and PL brands in hypothetical VR studies. Analyses in Chapter 8 will also further investigate whether the mode of fielding and number of trips simulated are statistically relevant factors in predicting the size of VR biases across a range of brands.

CHAPTER 7

STUDY 2: BIASES IN PACK CHOICES

Overview

Study 2 focuses on how the tendency of consumers to select larger packs (e.g. bundle packs, family sizes and share packs) in virtual shopping simulations affects the accuracy of virtual data in different product categories and retailers and explores whether these biases may be a result of *(1)* the samples recruited, (*2)* the size of the screens on which virtual simulations are displayed in different studies, *(3)* the absence of physical cues to confirm pack sizes or (*4)* the hypothetical nature of VR purchases.

Overview of theory

As discussed in Chapter 2 there is scant evidence regarding the ability of VR simulations to collect accurate data about consumers' pack size preferences (Burke, Harlam et al. 1992, Desmet, Bordenave et al. 2013).

In real purchase situations, concerns about handling, storage and perishability typically limit the market for large and extra large pack sizes (Granger and Billson 1972) and consumers typically buy larger packs to take advantage of price promotions or low unit prices (Mitchell, Lennard et al. 2003, Hendel and Nevo 2006, González-Benito, Loyola-Galván et al. 2010, Dawes 2012). When consumers do purchase larger packs, they tend to purchase fewer, if any, additional units from the category (Granger and Billson 1972), extend the time until their next purchase of the category (Gupta 1991, Fok, Paap et al. 2012), and in some categories, increase their rate of consumption (Chandon and Wansink 2002, Garber, Hyatt et al. 2009). However, past studies have found that inventory cues do not moderate consumers' purchase rates in VR studies (Campo, Gijsbrechts et al. 1999).

In real purchase situations, some households exhibit preferences for large pack sizes (Granger and Billson 1972, Blattberg, Buesing et al. 1978, Singh, Goodhardt et al. 2001, Rungie and Laurent 2012). Typically these are households with three or more people, high incomes and who travel to grocery stores by car (Griffith, Leibtag et al. 2009). VR studies may be prone to over inclusion such shoppers due to their use of intercept and online recruitment (Bruwer, Haydam et al. 1996, Baim, Galin et al. 2009).

Observational and experimental research indicates that consumers rely predominately on visual information (pack size and shape) to determine package sizes, rather than written information displayed on packs or shelves (Wansink 1996, Mitchell, Lennard et al. 2003, Garber, Hyatt et al. 2009, Ordabayeva and Chandon 2013, Objective Digital 2014).

Furthermore, when making pack size assessments, consumers often make judgments rapidly, use incomplete information (Wansink and van Ittersum 2003) and suffer from a number of perceptual biases, perceiving for example tall slim packs to be larger in size/capacity than shorter, wider packs (Wansink 1996, Garber, Hyatt et al. 2009, Ordabayeva and Chandon 2013). In VR experiments where consumers are likely to underestimate the size and distance to objects (Interrante, Ries et al. 2008) and cannot confirm the visual assessments by handling products, errors in shoppers' pack size choices are likely to be apparent. These hypotheses are explored through a number of structured investigations.

Structured investigations

RQ1.2: Consistency of shoppers' real and simulated pack size choices

If VR fielding does not influence the purchase choices of shoppers, then there will be a high level of consistency between shoppers' VR product choices and the pack sizes they claim to typically purchase in VR recruitment questionnaires.

To address this question, the level of consistency between shoppers real and simulated pack size choices will be quantified and the choices of different groups compared to determine if some groups of consumers choose products more accurately in VR simulations.

While using panel or diary data would have been preferable to address this question as consumer recall sometimes flawed (Tourangeau and Rasinski 1988, Ludwichowska, Romaniuk et al. 2017), this was not possible.

RQ2.5: Consistency of bias for large pack sizes

This investigation seeks to establish if VR purchase data is consistently biased towards larger pack sizes.

To highlight the biases evident, the purchase shares of large and extra-large pack sizes will be aggregated. While the dichotomisation and aggregation of continuous variables such as pack weight/unit count is not recommended (Irwin and McClelland 2003, Fitzsimons 2008), it is used in this chapter to aid in the identification of patterns that are consistently present across categories, an approach consistent with the empirical generalisations paradigm (Bound and Ehrenberg 1989, Ehrenberg 2000).

How large packs will be identified is outlined in the section titled: *Virtual shopping data*.

RQ3.2: Are sample biases responsible for the pack size biases observed?

The possibility that buyers of large packs were over-sampled in VR studies, is tested with data from a virtual study where consumers stated what pack size they typically purchase

prior to shopping the VR simulation. This could only be investigated with data from one study as most of the studies available for validation did not record what pack sizes shoppers typically purchase.

RQ3.3: Are the characteristics of virtual simulations responsible for the biases evident?

When interviewed, some industry researchers suggested that the scaling-down of products to fit onto computer screens and the lack of physical (touch) sensations may be responsible for consumers inadvertently selecting larger pack sizes in virtual simulations. To test this hypothesis, data from virtual shopping simulations is compared to data from (real) online shopping trips consumers where consumers also relied on scaled-down images of packs to make purchase selections and data that can provide insight into shoppers' attention to pack sizes, utilised.

RQ5.2: Is data from virtual simulations fielded online is more affected by biases for large packs?

In addition to addressing RQ5 regarding the accuracy of VR data collected online, this question investigates whether biases in data collected in VR are likely to be consequence of how VR studies are conducted and the technology utilised (RQ3). When simulations are fielded in specialised research facilities, large computer screens are typically used (i.e. 34 to 55 inch screens) whereas consumers home computer screens are used in studies fielded online.

Small objects benefit more from increases in display size than objects which are already relatively large (Orquin, Scholderer et al. 2012), so it is possible that the shares of small pack sizes are less underrepresented in simulations fielded in person with larger screens than in studies fielded online. The size of the screens used in each study examined is listed in Table 8.

RQ4.2: Are multiple trips necessary to collect accurate pack size data?

In real stores there is an inverse relationship between the number of items selected from a product category and pack size—when larger packs are purchased, fewer additional units are typically selected (Granger and Billson 1972).

An increase in the shares of large packs on participants' second VR trip would suggest that participants were aware of the pack sizes they choose and purposefully select larger pack sizes to compensate for selecting fewer items on their second trip, the pattern typically observed in VR studies (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999). An increase in purchase shares of large packs of between five to 15% is expected given that large packs are typically twice the size of small packs in these categories. In addition to addressing RQ4, are multiple trips necessary to collect accurate purchase data, this

investigation provides evidence to support RQ3, to what extent are biases in VR data a consequence of how VR studies are conducted.

RQ3.4: Could visibility and the absence of real consequences explain pack size biases?

A possible explanation for why purchases of large pack sizes are overestimated in VR is that large pack sizes are often more visible than regular pack sizes because of their greater surface area (see Figure 15). A number of prior studies have demonstrated that object size is one of the factors that drives unconscious visual attention in both physical and computerised settings (Janiszewski 1998, Wedel and Pieters 2000, Folkes and Matta 2004, Chandon, Hutchinson et al. 2009, Orquin, Scholderer et al. 2012, Clement, Kristensen et al. 2013, Peschel and Orquin 2013).

While large packs also attract more attention in real stores than small packs (Folkes and Matta 2004, Clement, Kristensen et al. 2013, Peschel and Orquin 2013), in real stores shoppers are deterred from selecting these pack sizes because of storage, handling and wastage concerns (Granger and Billson 1972). As already noted, these concerns may not influence shoppers in hypothetical VR simulations (Campo, Gijsbrechts et al. 1999, Lusk and Schroeder 2004). Similarly, prior research in real and experimental stores has established that colour affects which products shoppers notice when browsing the shelf but product requirements and taste preferences outweigh these effects when shoppers select products to purchase (Garber, Burke et al. 2000, Chandon, Hutchinson et al. 2009, Atalay, Bodur et al. 2012, Burke and Leykin 2014).

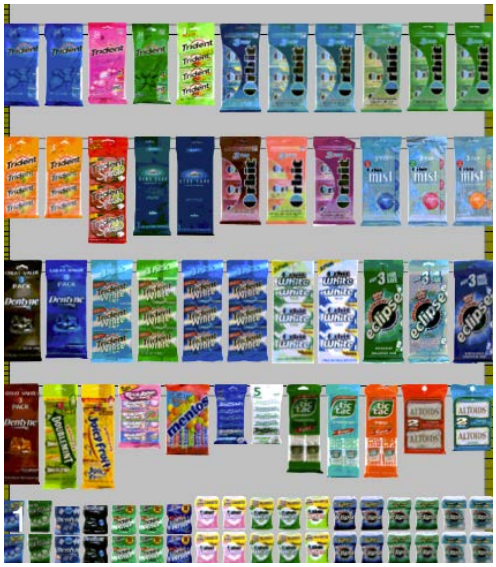
This question can be investigated descriptively by comparing results from product categories such as baby care, where packs containing more items have a larger surface area on shelf than small packs, to product categories such as chewing gum where packs containing more items/units have relatively smaller packs due to differences in pack format. Examples of these differences are shown in Figure 15 and Figure 16. In Figure 16, packages of gum that contain a large number of units (e.g. 120 pellets) are sold in bottles that likely have less on shelf visibility than multi-packs of gum pellets/sticks. Visual attention data will also be examined to address this question.

Figure 15: Example of a study where large packs were more visible than small packs



↑ Large pack = larger surface area

Figure 16: Example of a study where smaller multi-packs packs were more visible than large packs



N.B. Images from planning document used; VR more realistic looking

← Multi-packs of pellets = packs have large surface area

← Large count bottles = smaller surface area

Other factors that have been shown to increase attention to products including colour salience and pack dimensions/orientation (Van Der Lans, Pieters et al. 2008, Mormann, Navalpakkam et al. 2012, Clement, Kristensen et al. 2013, Clement, Aastrup et al. 2015) will also be investigated to determine their likely relevance to modelling.

Data sources & method: Overview of pertinent details

As with Chapter 6, this chapter uses simple univariate analyses to highlight differences in consumers' selection of different pack sizes in VR studies and real grocery purchase situations. Purchase choice data from VR studies that met the criteria outlined in Chapter 5 is presented is this chapter; all were designed to provide data that was generalisable to real stores. Study 8 (fielded online) was not used as it only included one pack size.

Virtual shopping data

The details of the VR studies presented in this chapter that are relevant to this stage of analysis are:

- A variety of different pack sizes were shown in the studies validated in this chapter. A 'regular' size was typically included, along with a large 'family' size, bundle or value pack and an extra-large 'mega' or 'party size' pack. None of these studies included small 'single serve' packs as these are not typically stocked in the main aisle of grocery stores or mass retailers in the US. Pack sizes were grouped based on the quantity of product contained in each pack (i.e. the unit count, weight or volume), generally into three categories. While the dichotomisation of continuous variables such as pack weight/unit

count is not recommended (Fitzsimons 2008) it is used to assist with the identification of patterns for further exploration.

- The studies presented in this chapter placed large pack sizes in a range of locations. In some studies, large pack sizes were mainly located on the upper and lower shelves (confectionery); in others they were spread across all shelves (feminine care) and in some they were mainly located on middle and lower shelves (paper towels). This is likely to control for any effect shelf location may have on biases, although crudely.
- Data from a range of product categories that varied in the pack dimensions that were prototypical to the category (Underwood 2003, Mocanu, Lockshin et al. 2012) was available.
- None of the studies excluded buyers of particular pack sizes and all aimed to have samples that broadly represented category buyers. The extent to which this was achieved is examined in section RQ3.2 of Results: Causes of upsizing biases.
- None of the studies had price promotions for specific products or pack sizes.
- Consumers could confirm the size of packs by viewing product information up close and rotating 3D images of packs. The proportion of shoppers that utilised this feature is reported in Table 34 to provide insight into results.
- The online and in-person studies compared were similar in terms of the products included, scale and sophistication of the VR simulation and fielding protocols. They did however differ in the hardware used to display the VR environment as outlined in Table 8.

Real benchmarks: Retailer scanner data

Estimates for the popularity of different pack sizes in real stores came from Nielsen's Retail measurement database. Precise benchmarks were obtained by aggregating the real shares of the products included in each VR study into pack sizes.

Unfortunately, it was not possible to obtain purchase estimates for different demographic groups from the panel data available. This would have strengthened the comparisons made in Table 30. Instead the purchase choices of different demographic groups are compared to the market overall.

Online purchasing: Consumer panel data

Nielsen panel data is used estimate the accuracy of large pack sizes on real online shopping trips. This was necessary because transaction data for online retailers could not be obtained.

Members of Nielsen's consumer panel, Homescan, report the details of all household grocery purchases using handheld scanners. Previous studies have established that this

data is fairly accurate (i.e. 90% of purchases are scanned correctly[10] (Zhen, Taylor et al. 2009, Einav, Leibtag et al. 2010). Only two percent of household panelists reported purchasing grocery products online in the periods analysed, a similar proportion the rest of the US population (2%, Eckersley 2011). Nielsen provided this data gratis for a small selection of product categories.

Visual attention data

Data that can provide insight into the factors that shoppers considered is also presented in this chapter:

- Study 2 and Study 4 collected eye-tracking data using a desk-mounted Tobii eye-tracking system. The system recorded eye fixations and saccades between points of interest for the duration of their shopping trip with a high degree of precision (Tobii 2012). This data was processed by the research agency responsible for fielding the study and the number of fixations on each element of the VR simulation (e.g. price, packaging etc.) reported. Further analysis of this data was not possible as the necessary data files were not available.
- Study 1 collected information about what products and areas of the screen shoppers clicked while shopping the VR simulation. This data is used as a proxy for visual attention data as shoppers' interactions generally follow their gaze path in computerised studies (Chandon, Hutchinson et al. 2009, Kleinschmit, Reid et al. 2012). Gaze tracking data was not collected in this study.

Key variables

The metrics used in this chapter are:

Real Share (%): the share of purchases of each pack size. Data is drawn from the retail outlet simulated (e.g. Walgreens for Pharmacy, Walmart or Target for mass merchandiser studies etc.) over a 12 week period prior to the VR study. For the online shopping data it was necessary to use a longer purchase period (12 months) due to the small sample of purchases reported.

Virtual Share (%): the proportion of simulated purchases for each pack size in the virtual study reported.

Virtual: Real the difference between the share of a pack in the virtual simulation and actual (real) share, expressed as a percentage of real market share.

Repeat Rate: the proportion of consumers who selected the pack size they claimed to buy most frequently in the virtual simulation. This measures the consistency between consumers' real and simulated choices.

10 10% of products are not scanned, primarily because shoppers consume these products before returning home. Out-of-home consumption is not likely to have compromised the accuracy of estimates for categories such as pet care (see Table 33).

Fixations: fixations are the pauses between eye movements (saccades) during which the eye gathers information about what is being seen. The total number of fixations on pack size information and proportion of fixations devoted to this element are reported in this chapter. In the VR studies for which data is reported, fixations of 50 milliseconds (1/200th of a second) on elements of the VR environment were identified by research agencies using visual attention analysis software (i.e. Tobii Studio).

Statistical tests

Chi-squared (X^2) tests are used to determine whether there is a significant difference between observed and expected number of large pack size purchases. Tests were generally conducted with two degrees of freedom. Where the less than five shoppers selected or were expected to select a pack size, an adjusted version of Fisher's exact conditional test was used; elsewhere unadjusted chi-squared tests were used.

T-tests are used to compare continuous purchase statistics derived from different groups and purchase conditions, for example the number of items selected by shoppers who purchased regular vs. larger packs in the VR experiment. Confidence intervals are used to determine if bias results from different groups/conditions (e.g. different fielding modes) overlap.

Results: Consistency of large pack bias across studies

In Table 29 the shares of larger pack sizes are compared to sales estimates from real (physical) stores. Chi-squared statistics for each study and retail outlet are reported in Table 64 of Appendix 7.

Table 29: Purchase shares of larger pack sizes in real and virtual studies

Virtual Study	Environment	Product category	Real Share	Virtual Share	Virtual / Real
1	Grocery store n=1083	Bagged confectionery	20.3%	34.9%**	173%
3	Pharmacy n=155 purch'	Feminine care	35.9%	51.5%**	143%
2	Mass retailer n=131 purch'	Pet care	28.4%	39.3%*	138%
4	Grocery store n=330 purch	Disposable nappies	41.9%	57.6%**	137%
6	Grocery store n=253 purch	Rice	21.5%	28.2%*	131%
5	Grocery store n=359 purch	Paper towels	55.6%	70.2%**	126%
3	Mass retailer n=155 purch'	Feminine care	45.3%	55.6%**	123%
4	Mass retailer n=330 purch'	Disposable nappies+	71.7%	79.5%**	120%
5	Mass retailer n=229 purch'	Paper towels	60.9%	70.9%**	116%
	Average pharmacy study n=310 purchases		35.9%	51.5%*	143%
	Average grocery study not weighted, total n=2025		34.8%	47.7%*	137%
	Average mass study n=845 purchases		51.6%	61.3%*	124%

+ All cells, not just control group. 1st trip only * $p < .05$, ** $p < .01$

Virtual shares of larger packs were overestimated by between 16 and 73%; differences were statistically valid in all studies.

In Table 29 the shares of large pack sizes were consistently overestimated compared to sales estimates from real (physical) stores (M_{bias} 34%, StD = 17%, differences significant i.e. $p < .05$ for all studies).

The results in Table 29 demonstrate, the virtual shares of large and extra large packs were *similar* in different retail environments (*Paper towels* V_{share} = 70.9% mass retailer, V_{share} = 70.2% grocery store, not significantly different $p > .10$; *Feminine care* V_{share} = 55.6% mass retailer, V_{share} = 51.5% Pharmacy, not significantly different $p > .10$). This is consistent with the results in Chapter 6, the shares of premium brands were *consistent* in VR studies for

different retail environments. Data permitting, whether retail outlet is a factor that influences the size of biases in VR studies will be explored with modelling in Chapter 8.

Behaviour of shoppers purchasing large and small packs

In real stores, when consumers purchase larger packs they tend to purchase fewer if any, additional units from the category (Granger and Billson 1972). The same tendency is evident in VR data. For example, in Study 1, shoppers who only selected larger bags of confectionery selected on average 1.6 items from the product category (StD= 1.5), a significantly fewer than shoppers who chose regular sized packs (M_{items} = 3.8, StD = 1.89, t (201) = 8.8, p = .001).

Results: Potential causes of pack size biases

RQ3.2: Sample biases are not likely to be responsible for the pack size biases observed

To validate that the VR purchases choices of particular demographic groups recruited for the study were not responsible for the biases towards large packs, the VR purchase choices of different demographic and behavioural groups are reported in Table 30 to determine if groups that have higher rates of large pack purchasing in real stores, also had higher rates of selection in VR two VR studies.

Table 30: VR purchase choices of different demographic groups

Baby care study	Regular packs	Large packs	X-Large packs	Statistical results VR compared to Real
Real Share	28%	64%	8%	
Aged 18-28 yrs	29%	31%	40%	X^2 (2) n = 180, = 436.6, p < .01
29-39 yrs	29%	33%	37%	X^2 (2) n = 165, = 348.5, p < .01
40-50 yrs	30%	36%	33%	X^2 (2) n = 30, = 48.3, p < .01
HH 1 person	29%	31%	40%	Not tested; very small sample
2 people	29%	33%	37%	X^2 (2) n = 53, = 121.3, p < .01
3-4 people	30%	36%	33%	X^2 (2) n = 270, = 630.8, p < .01
5+ people	29%	31%	40%	X^2 (2) n = 49, = 111.3, p < .01
Employed full time	31%	31%	38%	X^2 (2) n = 166, = 229.9, p < .01
Employed part time	30%	32%	39%	X^2 (2) n = 42, = 63.7, p < .01
Not employed	31%	31%	39%	X^2 (2) n = 117, = 164.6 p < .01

The VR choices of all demographic groups differed significantly from the distribution of results for the overall market, indicating there were differences in the VR purchase choices of some demographic groups.

Paper towel study VR	Regular packs	Large packs	X-Large packs	Statistical results VR compared to Real
Real Share	42%	14%	44%	
Aged 18-29 yrs	27%	4%	69%	X^2 (2) n = 63, = 15.1, p < .01
29-39 yrs	30%	7%	64%	X^2 (2) n = 61, = 10.0, p < .01
39-49 yrs	33%	10%	57%	X^2 (2) n = 64, = 4.2, p > .10 N.S.
49-59 yrs	31%	8%	61%	X^2 (2) n = 46, = 5.6, p < .10
HH 2 people	37%	8%	55%	X^2 (2) n = 38, = 1.9, p > .10 N.S.
3 people	21%	6%	73%	X^2 (2) n = 69, = 23.2, p < .01
4 people	20%	9%	70%	X^2 (2) n = 73, = 23.2, p < .01
5+ people	37%	7%	56%	X^2 (2) n = 52, = 3.2, p > .10 N.S.
Employed full time	31%	8%	62%	X^2 (2) n = 113, = 13.3, p < .10
Employed part time	33%	7%	60%	X^2 (2) n = 36, = 4.6, p < .01
Not employed	29%	7%	65%	X^2 (2) n = 87, = 13.7, p < .01
Income <$20,000 p.a.	23%	7%	70%	X^2 (2) n = 24, = 4.2, p > .10 N.S
$20-49,000 p.a.	35%	7%	58%	X^2 (2) n = 116, = 10.3, p < .01
$50-75,000 p.a.	30%	9%	62%	X^2 (2) n = 70, = 8.4, p < .05
$75,000+ p.a.	13%	17%	70%	X^2 (2) n = 24, = 8.5, p < .05

The VR choices of most demographic groups differed significantly from the distribution of pack size amongst the overall market, indicating there were differences in the VR pack size choices of some demographic groups.

Contrary to expectations (Griffith, Leibtag et al. 2009), the shares of pack sizes did not differ significantly between age, income, household size or employment groups. (Age; Baby care study X^2 p = .89, Paper towel study Income: p = .30; Household size: Paper towel study

p = .26, Baby care study p =.23; Employment: Paper towel study p = .30, Baby care study p =.83, additional statistics reported in Appendix 7). This is consistent with the findings of comparisons undertaken in Chapter 6.

Additionally, the distribution of pack size purchases deviated significantly from real purchase benchmarks for most demographic groups. For example, in the paper towel study, the pack size choices of shoppers aged between 18 and 29, 29 and 39 and 45 and 59 differed significantly from real benchmarks, as did the choices of shoppers in all employment groups, and households earning over $20,000 per annum. In Table 30, across most demographic groups the largest purchase share was for large pack sizes.

To further substantiate that sample biases were unlikely to be responsible for the biases observed in Table 29, one of the samples validated is benchmarked against the population of buyers it was designed to represent in Table 31.

Table 31: Validation of the sample for VR Study for baby nappies

	VR Sample n=250	Buying Pop'n	Statistical Results
Usually buy Regular pack size[11]	28%	30%	Not significantly diff. distribution (X^2 (2) n = 250, = 2.27 p > .10)
Large pack size	64%	60%	
X-Large pack size	8%	11%	
Aged 18-28 yrs	48%	51%	Significantly diff. distribution X^2 (2) n = 250, = 7.4, p < .05
29-39 yrs	44%	37%	
40-50 yrs	8%	12%	
HH 1 person	1%	1%	Significantly diff. distribution X^2 (3) n = 250, = 16.4, p < .01
2 people	14%	11%	
3-4 people	72%	64%	
5+ people	13%	24%	
Employed full time	51%	34%	Significantly diff. distribution X^2 (2) n = 250, = 31.9 p < .01
Employed part time	13%	20%	
Not employed	36%	46%	

The proportion of the sample claiming to regularly purchase regular, large and extra large packs sizes did not differ significantly to the shares of these pack sizes in real sales data.

The results in Table 31 demonstrate that the sample recruited for the VR baby care study did not differ significantly from the real benchmarks in their preferences for different pack sizes (X^2 (2) n = 250, = 2.27 p > .10). Small biases were however apparent in the age, household and employment profiles of the sample recruited. However, as outlined in Table 30, the shares of large and extra large packs exceeded the real shares of these pack sizes across all demographic groups.

[11] In Table 31 results for small packs are not reported as they are responsible for less than 5% of total category sales and are not typically available in mass retailers, the outlet simulated.

RQ3.2: Consumers who typically buy small packs tend to upsize their purchases in VR

In Table 32 the proportion of shoppers who selected a product from the same size tier in the VR study as they indicated they typically purchase in real stores (Repeat Rate) is reported. Data from the same study as was reported in Table 31 is reported. In Table 32 the results of X^2 tests conducted to determine if some groups were statistically more likely than to overall sample to deviate from their typical pack size choice, are reported.

Table 32: Claimed purchasing of different pack sizes VR purchases

	Survey n=250	Repeat Rate in VR	Statistical Results
Regular 23-45 nappies per pack, approx. $8.50 to $11 per pack	29.8%	44%	Significantly diff. to overall RR X^2 (2, n = 74) = 13.3 p < .01
Large 46-73 nappies per pack, approx. $18 to $22 per pack	59.5%	77%	Significantly diff. to overall RR X^2 (2, n = 149) = 11.6 p < .01
X-Large 74-144 nappies per pack, approx. $29 to $32 per pack	10.7%	61%	*Not sig. different* to overall RR X^2 (2 n = 27) = 0.16 p > .10
Total sample	100.0%	64%	
Total larger packs	71.2%	78%	Significantly diff. to overall RR X^2 (2, n = 175), = 15.2 p < .01

More buyers of large packs selected the pack size they typically purchase than buyers of small packs, this is likely due to large packs being the most popular pack size in the VR study (57.6% of purchases).

The *Repeat Rate* results and statistics in Table 32 demonstrate that only 64% of shoppers selected a product from the same pack sizes as they typically purchased. Given that consumers were asked which pack sizes they normally purchase immediately prior to shopping the virtual simulation it is surprising that repeat rates are not higher as questioning likely primed some participants to select the same pack sizes (Morwitz, Johnson et al. 1993, Fitzsimmons and Morwitz 1996, Malhotra, Shaw et al. 1996, Chandon, Morwitz et al. 2005).

The *Repeat Rate* results and statistics in Table 32 demonstrate that the level of Real/VR purchase consistency was higher for shoppers who typically purchase larger pack sizes, the most popular pack size in this category, than for the sample overall. Moreover, the *Repeat Rate* results for shoppers who purchased regular or large pack sizes are statistically different (Table 32: C.I.s for RR_{Reg} 44% does not overlap with C.I.s for RR_{XL} 71% or RR_{Lg} 61%, 95% C.L.).

Differences in repeat rates for demographic and behavioural groups were also explored. Repeat rates were not statistically or significantly lower for any demographic or shopper groups across the variables examined (age, household size and income, education level and employment status). Repeat rates were however lower for the small group of participants

(3.6%) who indicated in the post-shopping interview that they had selected something new in the virtual simulation than for the VR sample overall (*RR* = 25%, significantly different to overall result, X^2 (2, n = 9), = 8.0 $p < .01$).

Interestingly, in the post-shopping interview for this study the majority of participants (96.6%) claimed to have selected the pack size they would normally buy. This suggests that either, participants failed to pay attention to the size of the products they selected, or they believed they had selected smaller pack sizes. The investigations on page 138 to 141 test these competing explanations. This result does however suggest that consumer's self reports of the realism of their behaviour, a measure used in past studies (Tjostheim and Haugland 2005) are not to be trusted.

RQ3.3: The 'virtual' characteristic of VR shopping is not responsible for the biases evident

On real online shopping trips consumers must rely on written and visual information to make product selections and cannot handle products prior to purchasing to verify their size/weight. Table 33 compares the proportion of large and extra-large packs purchased on real online shopping trips, real physical stores and VR stores. The results of statistical comparisons of online and VR shares are also presented.

Table 33: Purchase shares of larger pack sizes in-store, online and in VR

Study	Aggregated share of Lg & XLarge packs		In-store	Online	Virtual	Online v. Virtual Statistical Results
1	Grocery store n=1083	Bagged confect	20.3%	22.2%	34.9%	X^2 (2)=77.4, p = .001
2	Mass retailer n=131	Pet care	28.4%	34.1%	39.3%	X^2 (2)=1.5, p =.28 N.S.
5	Grocery store n=359	Paper towels	55.6%	55.1%	70.2%	X^2 (2)=38.8, p = .001
5	Mass retailer n=229	Paper towels	60.9%	55.1%	70.9%	X^2 (2)=27.3, p = .001
	Average study / category		41.3%	41.6%	53.8%	

Large pack sizes are more popular in virtual simulations than on real online shopping trips.

The results in Table 33 demonstrate that the VR purchase shares of larger pack sizes are significantly larger than benchmarks from online shopping data for three of the four studies compared. For example, in the simulation for confectionery, 34.9% of the bagged confectionery products selected were packs larger than 20 ounces (560g), a size that constitutes 22.2% of purchases online, a significantly larger proportion (X^2 (2, n = 1083) = 77.4, $p < .01$). The only VR study for which online/VR purchase shares were not significantly different, pet care had a very small sample size (n=131).

RQ3.4: Consumers pay very little attention to pack size information in VR studies

Eye-tracking data from virtual simulations suggests that only a few participants actively consider the size/weight of products they chose. Consumers' gaze patterns were tracked in the simulations for disposable nappies and feminine care products and less than three percent of the fixations recorded focused on the sizing information on packs (*n*=667 consumers, 47,437 fixations). It is possible that attention to pack size information (e.g. ounces, units contained etc.) also quite low in real stores with the majority of visual attention being devoted to colour, brand name and pictorial information on the front of product packaging (Scamell-Katz 2012, Orquin, Bagger et al. 2013, Hendrickson and Ailawadi 2014).

In real stores consumers rely predominately on visual information i.e. package dimensions and shape to identify pack sizes (Wansink 1996, Mitchell, Lennard et al. 2003, Garber, Hyatt et al. 2009, Ordabayeva and Chandon 2013, Objective Digital 2014). However, in VR studies, pertinent size cues are often obscured by the positioning of products, as Figure 18 demonstrates. In such instances, the only way consumers can visually confirm the size of products is to access the three-dimensional models of products built into many VR simulations. These allow shoppers to see the back, side and rear pack dimensions of the products. The proportion of shoppers who used this feature is reported in Table 34 for a number of the studies validated.

Figure 17: How pack and product size is conveyed in VR



Table 34: Proportion of consumers who rotate a product they selected

Retail environment	Product category	Rotate
Mass retailer n=155	Feminine care	7.1%
Grocery store n=1651	Paper towels	7.5%
Grocery store n=1083	Chocolate	8.4%
Pharmacy n=155	Feminine care	9.1%
Mass retailer n=330	Disposable nappies	12.1%
Grocery store n=330	Disposable nappies	14.1%
	Average	9.7%

Less than 10% of participants rotated products to determine their dimensions, suggesting most shoppers are uninterested in confirming the size of the products they selected.

The results in Table 34 demonstrate that very few consumers utilised the three-dimensional information contained in VR studies for any purpose including verifying the size of the products. This suggests that consumers were not overly concerned with the size of packs in these hypothetical VR simulations.

RQ5.2: VR data collected online is not more affected by pack size biases

The aggregated shares of regular packs in two VR studies, one fielded online and one fielded in-person in a specialised research facility with large screens, are compared in Table 35. The online study used consumers home computers including laptops, to display the VR simulation. The in-person study used two 24-inch computer screens arrayed side by side to provide a wide view of the VR store. Regular packs, the smallest pack size shown in this study, were the size whose purchase shares was hypothesised to be most affected by the difference in display size (Orquin, Scholderer et al. 2012).

Table 35: Results for pack sizes in simulations fielded with different screen sizes

Feminine care	Fielded In-Person			Fielded Online		
	Real Share	Virtual Share	Virtual / Real	Real Share	Virtual Share	Virtual / Real
Regular packs	64.1%	48.5%	76%	65.5%	50.9%	78%
Large & XL packs	35.9%	51.5%	143%	34.5%	49.1%	142%

The shares of regular and larger pack sizes are not significantly different in VR studies fielded online and in person with different sized screens.

The results in Table 35 indicate that the shares of regular and larger pack sizes were similar in VR studies fielded online and in specialised research facilities (X^2 (2) = 0.751, not significantly different, p = .39). Additionally, VR/Real bias for large packs, did not differ significantly between the in-person and online studies (confidence intervals overlap at a 95% C.L.). However, the samples for these studies did not have sufficient statistical power to be confident an effect was not evident (<10% power , α = .05).

RQ4.2: Additional trips have a no impact on pack sizes biases

In Table 36 the shares of larger pack sizes on consumers' first and second trips are compared for two VR studies. An increase in the shares of large packs on participants' second VR trip would suggest that participants purposefully selected more larger pack sizes on their second shopping trip to compensate for selecting fewer items.

Table 36: Shares of larger pack sizes in consumers' first and second trips

		Real Share	Virtual Share 1st trip	Virtual Share 2nd trip	X^2 test of difference 1st vs 2nd trip	Virtual / Real 1st trip	Virtual / Real 2nd trip
Baby nappies	**Large + XL**	41.9%	57.6%	59.9%	p >.10 N.S.	139%	144%
Confectionery	**Large + XL**	20.3%	34.9%	35.1%	p >.10 N.S.	172%	173%
Average study		31.1%	46.3%	47.5%		156%	159%

Shares of larger pack sizes were not significantly smaller on participants' second simulated shopping trips.

The results in Table 36 demonstrate that the shares of larger pack sizes were not significantly larger on consumers second VR shopping trips in either of the categories validated (Confectionery: X^2 (2, n = 1500) = 0.02, p = .89, Baby nappies: X^2 (2, n = 700) = 0.48, p = .49). Moreover, simulating an additional trip did not have any effect on the magnitude of the bias observed for large pack sizes (C.I.s overlap at 95% C.L).

RQ3.4: Pack visibility may explain biases for larger packs

In Table 37 market share results for three product categories where packs containing a large quantity of products had a larger visible surface area than smaller packs are compared to a product category, Chewing gum, where the opposite was true. In this category, the largest pack size, bottles 150 pellets of gum, had less visible surface area than multi-stick packs of gum. In all the other categories listed, packs that contained the largest quantity of products were more visible on shelf. The differences in the surface area of different pack sizes across categories were illustrated in Figure 15 and Figure 16.

Table 37: Purchase shares of larger pack sizes across categories

Study	Product category	Real Share	Virtual Share	Virtual / Real	Statistical Results
1	Chewing gum	16.9%	11.1%	66%	X^2 (1, n= 87) = 2.85, p < .10
5	Paper towels	55.6%	70.2%	126%	X^2 (1, n=359) = 36.0, p < .01
4	Disposable nappies	41.9%	57.6%	137%	X^2 (1, n=330) = 33.5, p < .01
1	Bagged confectionery	20.3%	34.9%	173%	X^2 (1, n=1083) = 101, p < .01

Fewer consumers selected large pack sizes of gum in the VR study than in real stores; the opposite is true for all the other categories.

The results in Table 37 suggest that the visibility of different pack sizes may influence whether the purchases of large pack sizes are overrepresented in VR studies. *Chewing gum* was the only category in which purchases of large packs were under rather than over estimated (*Chewing gum*, Real/VR difference marginally significant, p < .10; Real/VR difference highly significant p < .01 for other studies).

That pack visibility may explain pack size biases is consistent with the hypothesis outlined on page 127. A number of prior studies have demonstrated that object size is one of the factors that drives unconscious visual attention in both physical and computerised settings (Janiszewski 1998, Wedel and Pieters 2000, Folkes and Matta 2004, Chandon, Hutchinson et al. 2009, Orquin, Scholderer et al. 2012, Clement, Kristensen et al. 2013, Peschel and Orquin 2013). Pack orientation and colour/salience have also been shown to drive unconscious visual attention in past studies (Van Der Lans, Pieters et al. 2008, Mormann, Navalpakkam et al. 2012, Clement, Kristensen et al. 2013, Clement, Aastrup et al. 2015). Whether these factors are likely to impact VR purchase biases is explored to identify other factors that could be included in modelling in Chapter 8.

In Figure 18, how much attention participants paid to different pack sizes in Study 1 is visualized using a heat map. This heat map was prepared by the research agency responsible for fielding Study 1. This study collected data about what products and areas of

the screen participants clicked while shopping the VR shelf, information that is likely to be indicative of what caught shoppers' attention in the VR simulation (Chandon, Hutchinson et al. 2009, Kleinschmit, Reid et al. 2012). In Figure 18, packs that were clicked on by many participants are highlighted with warm colours (i.e. orange, yellow) and packs that were clicked on by few shoppers are highlighted with cool colours (i.e. green and blue). The planogram from the study is shown alongside the heatmap to aid in the interpretation of patterns.

Previous studies have established that packs which are tall and slim are more likely to be noticed than shorter, wider packages (Wansink and van Ittersum 2003, Clement, Kristensen et al. 2013) and objects that are orientated differently to others on the shelf attract more attention than objects which are uniform (Itti, Rees et al. 2005). The heat map results in Figure 18, suggest that this was the case in the VR study—the extra-large confectionery packs on the top shelf received more attention than the regular and large packs. Purchase results from this study confirmed that packs which were tall and slim had larger than expected purchase shares—23.3% of purchases vs. 9.1% of purchases, a statistically significant difference, X^2 (2, n = 1083) = 121.0 $p < .01$.

The same pattern was evident in purchase results from Study 6, a VR study for a staple food product. In Study 6, purchases of products with low, wide packaging were substantially underestimated (*Real* = 15.8%, *VR* = 7.3%, significantly difference, (X^2 (2, n = 253) = 14.3 $p < .001$). In contrast, purchases of products with tall, narrow packaging, the norm for this category, were substantially overestimated (*Real* = 15.8%, *VR* = 7.3%, significantly difference, (X^2 (2, n = 253) = 13.1 $p < .001$). Images of this study are shown in Chapter 9.

Analyses on page 146 explore another factor that could be incorporated into modelling in Chapter 8, pack colour uniqueness.

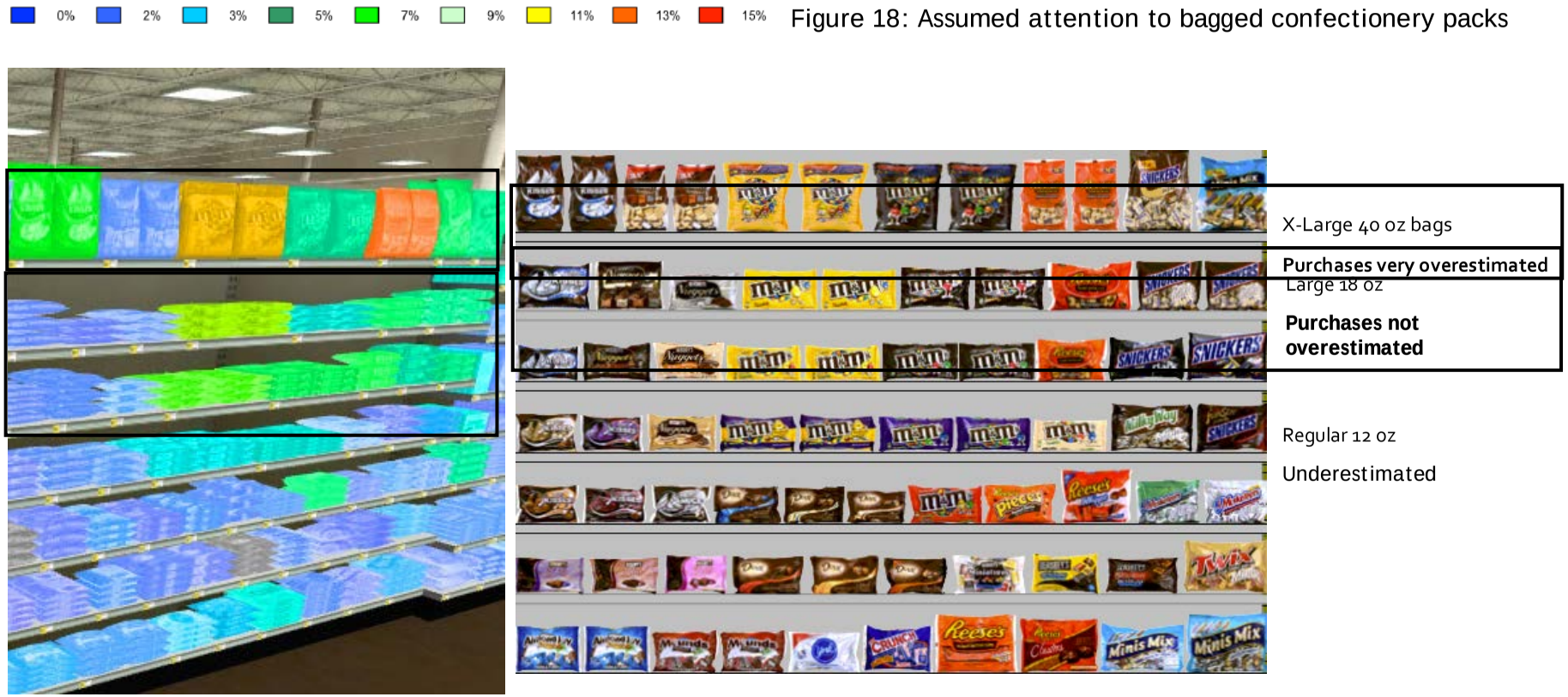


Figure 18: Assumed attention to bagged confectionery packs

Tall, slim packs (top shelf) are likely to have received more attention based on the pattern of interactions with VR products. Packs that received more attention were more over purchased. Even when all packs were stacked upright, as in the image on right, extra-large packs were over purchased relative to large and regular packs (27.3% simulated share vs 17.8% and 54.7%).

The heat map results in Figure 18 suggest that products with brightly coloured packaging that stood out from other products on the shelf (e.g. Peanut M&Ms, Reese's), received more attention than products with predominately brown packaging (e.g. Chocolate M&Ms, Hershey's Kisses), a colour commonly used in the category. Past studies have found that the uniqueness of a packaging's colour is a major determinant of its visual salience on shelf (Van Der Lans, Pieters et al. 2008). The uniqueness of these pack colours was quantified by uploading an image of the VR environment to a colour analysis application.[12] This application determined the precise hue of each pixel and estimated what proportion of the environment's total pixels were comprised of different hues using a k-means algorithm to cluster colours with similar hues (Krzywinski, 2006). Through trial and error, eight hue clusters was found to be optimal for identifying hues used in a small number of pixels in this environment including orange, yellow and green. Algorithms with fewer than eight clusters pushed green hues into other clusters, likely increasing the sum of squares within clusters.

The proportion of the environment's total pixels represented by each colour-cluster is reported in Figure 19. The Delta E (ΔE) statistics reported are not relevant to the analysis but were included on standard outputs for the image analysis tool. They indicate the closeness the of central colour for each cluster to a standardised colour in the CIELAB system (Krzywinski, 2006). This is a system used by illustrators and industrial designers to mathematically compare colours. The colour location map in Figure 19 was used to identify packs that featured unique colours.

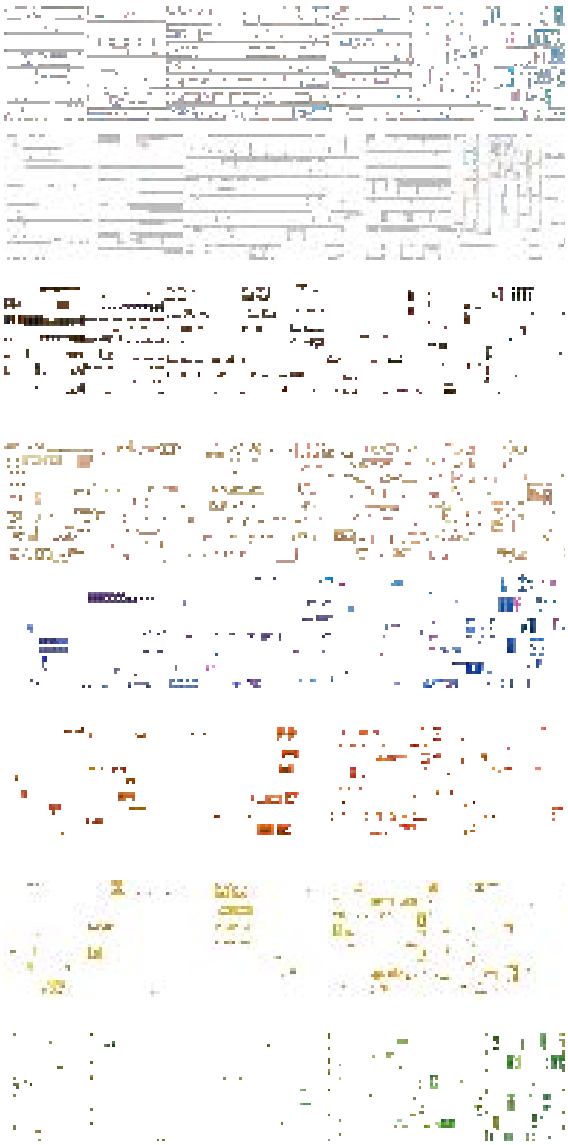
[12] Image Color Summarizer 0.76 © 2006-2018 Martin Krzywinski |http://mkweb.bcgsc.ca/color-summarizer. While a variety of colour analysis applications are available on the web, Image Color Summarizer is easy to use, provides detailed descriptive statistics and allows the user to specify the number clusters and pixel rate used in analysis.

Figure 19: Results of colour composition analysis for VR confectionery environment

Proportion of environment pixels that were:

35.21%	186,186,186 grey ΔE=0.8
20.27%	139,134,133 suva grey ΔE=1.1
13.02%	89,69,55 brown derby ΔE=1.4
10.68%	188,158,130 nude brown ΔE=3.1
6.33%	102,102,153 dark blue gray ΔE=3.0
5.78%	177,89,47 fiery orange ΔE=1.9
4.70%	198,168,75 roti ΔE=4.8
4.01%	122,148,97 highland ΔE=3.9

Visualisation of locations:



Brown was present in 24% of number of pixels of the VR environment as it is a colour commonly used in chocolate packaging; orange, yellow and green were less common.

The results in Figure 19 demonstrate that green, orange and yellow were the least common hues in the VR simulation for confectionery, each comprising less than 5% of the environment's total pixels. Brown and beige were among the most common hues, collectively representing 24% of the environment's total pixels. Past studies have found that the uniqueness/distinctiveness of a packaging's colour is a major determinant of its visual salience on shelf (Garber, Burke et al. 2000, Van Der Lans, Pieters et al. 2008, Orquin, Scholderer et al. 2012).

The impact of pack colour on shoppers' VR purchase choices is explored in Table 38 for bagged confectionery products. The products compared in this table had similar shelf positions, prices and size ranges in the VR simulation. The statistical results in Table 38 demonstrate that the VR shares

of chocolate bags were consistently different to their real shares. The *p*-values reported in Table 38 range between $p < .01$ and .05.

Table 38: The influence of visual salience on purchase choices

Bag confections	Pack Colours include:	Real Share	Virtual Share	Virtual/ Real	Statistical results
Reese's	Orange	11.8%	22.8%	193%	X^2 (2, n=1083) = 74.2, $p < .01$
Peanut M&Ms	Yellow	16.4%	20.6%	126%	X^2 (2, n=1083) = 11.4, $p < .01$
3 Musketeers Mint	Green	0.1%	0.3%	300%	X^2 (2, n=1083) = 40.5, $p < .01$
Hershey's Kisses	Brown	13.5%	9.7%	72%	X^2 (2, n=1083) = 17.7, $p < .01$
Milk M&Ms	Brown	14.0%	9.6%	69%	X^2 (2, n=1083) = 24.5, $p < .01$
Hershey's Nugg	Brown	6.7%	5.3%	79%	X^2 (2, n=1083) = 4.7, $p < .05$
Snickers	Brown	7.9%	12.4%	157%	X^2 (2, n=1083) = 19.6, $p < .01$
3 Muskateers	Silver	0.3%	0.8%	266%	X^2 (2, n=1083) = 93.9, $p < .01$

Products with unusual pack colours typically have larger virtual purchase shares than expected while products with brown packaging typically have smaller VR shares than expected; although some exceptions are evident i.e. Snickers.

The shares of products with pack colours that were unusual in the context of the VR simulation were generally overestimated in the VR study while the shares of products with predominately brown packaging, were generally underestimated. For example, in Table 38, Peanut M&Ms were significantly selected by significantly more VR shoppers than Milk Chocolate M&Ms in the VR study (V_{share} 20.6% vs 9.6%, X^2 (2, n= 1083) = 51.0 $p < .01$). This suggests that colour uniqueness is a factor that should be considered in Chapter 8.

However, the results in Table 38 indicate that the visual distinctiveness of products is unlikely to be the only factor that affects the VR shares of products. In Table 38, the VR share of Snickers, a brand with brown and gold packaging, is significantly overestimated (X^2 (2, n=1083) = 19.6, $p < .01$). At the time this study was conducted, Snickers had a very successful advertising campaign, '*You're not you when you're hungry'*, on air (Kennedy and McColl 2012). This may have influenced the number of shoppers who selected Snickers products in the VR study; an interesting finding that may suggest, if it can be confirmed through further research, that VR can be used to estimate the in-store effects of out of store advertising.

Chapter Summary

The results presented in this chapter demonstrated that there was a bias towards the selection of larger pack sizes in VR studies conducted for a range of grocery product categories. These were particularly pronounced in studies for retail outlets such as grocery and pharmacy stores. Data permitting, whether retail outlet is a factor that influences the size of biases in VR studies will be explored in Chapter 8.

The results presented in this chapter also demonstrate that the VR samples did not overrepresent shoppers who typically purchase large or extra-large packs and that shoppers with a variety of demographic characteristics deviated from their typical pack size choices in VR studies. While comparing the (single) pack size that shoppers claim to purchase most commonly to real sales data to confirm that the VR samples did not overrepresent shoppers who typically purchase large or extra-large packs was not ideal as shoppers typically purchase a range of pack sizes (Kennedy and Singh 2002, González-Benito, Loyola-Galván et al. 2010, Dawes 2012), this could not be avoided.

That the that the lack of physical (touch) sensations may have been responsible larger pack sizes being over-selected was explored with data from real online shopping trips, and not supported—the purchase shares of larger pack sizes were significantly larger in VR studies than in real online shopping data.

Analyses also explored, and rejected, the possibility that additional shopping trips improve the realism of VR shopping data—the shares of larger pack sizes were not significantly larger on consumers' first and second VR shopping trips. Indeed, the visual attention and interaction results presented in this chapter demonstrated that most participants were uninterested in determining the size of the products they selected. This may have been because participants' purchase choices did not have any real consequences or because the majority of participants (i.e. 96.6% in study validated) thought they had selected the pack size they normally buy.

The absence of significant variation in shoppers pack size choices across household and income groups in Table 30 suggests that economic concerns and consumption rates fail to inform shoppers' VR purchases, a finding consistent with past studies (Campo, Gijsbrechts et al. 1999) and results in Chapter 6 regarding consumers' brand choices.

The explanation with the most evidence to support its acceptance is that in VR studies without real purchase consequences, consumers browse the VR shelf, selecting whatever products catch their eye, irrespective of whether they would select these products in real stores. Larger packs, packs with visually salient colours and certain orientations/shapes, attract more attention in real and computerised settings (Janiszewski 1998, Wedel and Pieters 2000, Folkes and Matta 2004, Chandon, Hutchinson et al. 2009, Orquin, Scholderer et al. 2012, Clement, Kristensen et al. 2013, Peschel and Orquin 2013). Analyses suggested that some of these variables should, if possible, be

incorporated into modelling to predict the size of biases in VR data. Only one of these factors, pack format, was mentioned in prior analyses for VR data (Higaki 2004).

Analyses also demonstrated that VR purchase data collected online via pre-recruited research panels is similar in terms its level of pack size bias to data collected in specialised VR research facilities. That the fielding mode and infrastructure does not affect the overall level of bias in VR purchase data will be investigated with regression analysis in Chapter 8.

CHAPTER 8

STAGE 2: CONFIRMATORY MODELLING

Overview

This chapter investigates if the factors identified in Chapters 6 and 7 as potential causes of biases are significant predictors of the size of biases in VR studies when the effects of these variables and product category are considered simultaneously. It also validates findings from Chapters 6 and 7 regarding the effects of fielding mode and number of trips simulated on the size of biases in VR studies.

Introduction & overview of theory

The univariate analyses presented in previous chapters identified a number of variables that may impact purchase biases in VR studies including price premiums, product category type, pack size, pack orientation and pack colour. The key objective of this chapter is to determine which variables have significant impacts on the size of VR biases when tested concurrently.

In this chapter the following variables are examined:

Type of brand: Prior validation studies concluded that the shares of national brands were typically overestimated in VR studies while the shares of generic and private label brands were sometimes underestimated (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013, van Herpen, van den Broek et al. 2016). Analyses in Chapter 6 demonstrated that the shares of private label brands were always underestimated, irrespective of whether they were marketed as generic, store or standalone private label brands.

Relative price: Analyses in Chapter 6 suggested that there may be a systematic relationship between the prices of products relative to others in the category (i.e. *Price premium*) and purchase share biases in VR studies. This conclusion is consistent with a prior study that demonstrated the accuracy of VR market share estimates for new beer brands could be improved by 30% by calibrating results based on product's real prices (Higaki 2004).

Product category: The reporting of results by product category in Chapters 6 and 7, suggested that there were differences in the magnitude of VR biases across the six product categories studied, although this was not tested directly. Past studies also found that the accuracy of VR purchase estimates varied across grocery products categories (Burke, Harlam et al. 1992, Needel 2013, Waterlander, Jiang et al. 2015).

Type of product category: Descriptive comparisons in Chapter 6 highlighted a substantial difference in the size of biases for premium brands in studies for functional products such as baby diapers, compared to the size of biases in hedonic products such as chocolate and chewing gum.

Whether this difference was a consequence of the product category types or differences in the magnitude of price premium across these categories could not be determined with univariate analyses.

Package size: Descriptive comparisons in Chapter 7 found that the shares of larger pack size products were overestimated in VR studies for a range of product categories. Prior VR validation studies have concluded that consumers fail to consider factors that typically deter the purchasing of large packs in real stores, such as consumption rates and storage capacity (Granger and Billson 1972, Campo, Gijsbrechts et al. 1999). Additionally, object size has also been shown to positively influence attention to products in many prior studies (Janiszewski 1998, Wedel and Pieters 2000, Folkes and Matta 2004, Chandon, Hutchinson et al. 2009, Orquin, Scholderer et al. 2012, Clement, Kristensen et al. 2013, Peschel and Orquin 2013). Whether this is a factor that influences purchase biases when the impact of other variables is controlled will be tested.

Package orientation: Descriptive comparisons of products in two categories suggested that purchases of products with low, wide packages (i.e. a horizontal pack orientation) are often underestimated in VR studies while purchases of products with tall, narrow packages are typically overestimated in VR studies. Prior research has established that pack orientation derives unconscious visual attention to products (Wansink and van Ittersum 2003, Clement, Kristensen et al. 2013).

Unique colour: Descriptive analysis in Chapter 7 identified that purchases of products with pack colours that were unusual (i.e. used in less than 5% of the environment's total pixels) were generally overestimated in VR, while purchases of products with more commonly used colours were often underestimated in VR. Past studies have found that the uniqueness/distinctiveness of a pack's colour is a major determinant of its visual salience on shelf but that pack colour does not typically influence shoppers' purchase choices (Garber, Burke et al. 2000, Chandon, Hutchinson et al. 2009, Orquin, Scholderer et al. 2012, Scammel-Katz 2012). Colour may have a stronger effect on consumers' purchase choices in VR settings.

Retail environment simulated: Univariate analyses in Chapter 6 and 7 indicated that the VR shares of products were similar in VR studies that simulated different types of retail environments e.g. pharmacy store, grocery store etc.

Number of trips simulated: However, descriptive comparisons in Chapter 6 and 7 found that additional trips did not affect the biases evident for large packs, PL brands or premium brands. In these comparisons, VR results for the same product categories were compared, limiting the number of studies which could be used in comparisons to four. Modelling will enable data from a larger number of VR studies to be compared to confirm this finding.

Fielding mode: Analyses in Chapter 6 and 7 found that there was not a substantial difference in the magnitude of VR for biases premium brands, private labels or large pack sizes when results from VR

studies fielded in person and online were compared. As with the number of trips simulated, only VR results for the same product categories were compared. Modelling will enable data from a wider range of studies to be compared and the effects of other variables including product category and price to be accounted for in comparisons. This is prudent as some of the brands compared differed in their level of price premiums as was noted in Chapter 6.

Size of the brand: Throughout Chapter 6 and 7, percentage bias results were reported (i.e. *VR/Real*) for all brands, irrespective of their real market shares. As pointed out by Burke et al. (1992), even small deviations can be large when reported in percentage terms for small brands.

In this chapter, the significance of these factors in predicting VR biases will be explored using multivariate regression modelling.

Description of data

Virtual shopping data

The characteristics of the data sets relevant to this stage of analysis are:

- The same data sets as were used in descriptive analyses in Chapter 6 and 7 are used in this chapter.
- As outlined in Chapter 5, all of the data sets included a representative selection of national brands and at least one private label product.
- Only data from VR trips where participants were instructed to 'shop as they normally would' was utilised. Two of the studies, Study 1 and 4 simulated more than one 'normal' shopping trip.
- Two of the studies, Study 7 and 8, were fielded online using participants' home computers; the remainder were recruited and administered by trained researchers in shopping mall research facilities.

Key variables

Variables were coded as follows:

- *Type of brand* was coded as a dichotomous variable with the reference category of zero representing nationally available brands and one representing Private label brands.
- To aid in the interpretation of model coefficients, the price premium variable used in Chapter 6 was transformed. Brands were assigned positive or negative integer (e.g. +20, -20% etc.) based on their price relative to the average price of the NBs in each VR study.
- Product category was coded into six dichotomous variables, *PCatg1* to *PCatg6,* to represent the seven product categories from which data is analysed. These are listed in Table 39.
- Product category type was coded into a dichotomous variable, *Hedonic_catg.*, with values assigned on the basis of hedonic/utilitarian values reported in prior studies (Wakefield and Inman 2003, Sethuraman and Cole 1999). Why prior studies were used is justified in Chapter 6. This variable was used to calculate an interaction term, *Price:Hedonic,* to determine if the nature

of the product category increased the effect of *Relative Price* on *VR Bias.*[13] Rice, a grocery product whose hedonic/utilitarian value in the US has not been established by prior research, was assigned a value of zero when calculating the interaction variable, *Price:Hedonic.*

- The type of retail environment simulated was coded into two variables, *Grocery_store* and *Mass_store,* to cover the three retail environments from which data was analysed: grocery stores, pharmacy stores and mass merchandise environments.
- The product weight/unit count of each product included in the VR study e.g. three rolls, 10 ounces etc. was determined. To enable categories that varied in their units to be compared, the 'regular' pack size for each category was assigned a value of 1 (see Table 39). What constituted a 'regular' pack size for each category was determined via pack size classifications in industry data. Values for other packs were then computed based on their size relative to the regular pack size e.g. two rolls = .66 of 'regular' pack size, six rolls = 2, to provide a metric that was consistent across categories.

[13] In additive regression models, the main effect of any variables tested for interaction effects, are typically included. However it is acceptable to violate this principle if at least one of the variables on which the interaction term is based, e.g. *Relative price,* is included and the limitations of this decision are understood or addressed by the researcher i.e. variable is mean centred or used only to determine the significance of variables, not their effects (Nelder 1999).

Table 39: Pack size and price values across (product) categories

Category	Coded as:		n data points	Size of 'Regular' pack	Range of Pack Sizes in VR
Chocolate	*PCatg.1* =1	*Hedonic.Catg.* = 1	26	12 oz	25 to 187% of 'regular'
Chewing gum	*PCatg.2* =1	*Hedonic.Catg.* = 1	16	45 items	11 to 133%
Petcare	*PCatg.5* =1	*Hedonic.Catg.* = 1	29	24 oz	26 to 333% of 'regular'
Feminine	*PCatg.3* =1	*Hedonic.Catg.* = 0	18	15 oz	41 to 161% of 'regular'
Rice	*PCatg.6* =1	N/A, excluded	18	8 oz	69 to 377% of 'regular'
Paper towels	*PCatg.1-6* =0	*Hedonic.Catg.* = 0	20	3 rolls	67 to 266% of 'regular'
Baby nappies	*PCatg.4* =1	*Hedonic.Catg.* = 0	22	44 units	24 to 199% of 'regular'
Baby nappies	*PCatg.4* =1	*Hedonic.Catg.* = 0	3	44 units	199% of 'regular'
Feminine	*PCatg.3* =1	*Hedonic.Catg.* = 0	30	28 units	36 to 170% of 'regular'
		Hedonic.Catg. = 1	71	-	25 to 333% of 'regular' sizes
		Hedonic.Catg. = 0	94	-	24 to 266% of 'regular'

- Products with unique package colours were identified using the same k-means hue clustering approach described in Chapter 7. Packages that featured a colour present in 5% or less of the environment's total pixels were assigned a value of one; all other packages were assigned a value of zero. While a more nuanced approach could have been adopted, simple dichotomous coding was felt to be sufficient for exploratory analysis of this variable. The unique colours for each product category/study are reported in Appendix 7.
- *Pack orientation* was coded by measuring the vertical and horizontal dimensions of each product and assigning packages with a vertical measurement greater than their horizonal, a value of one and all other packages, with a value of zero. While coding *Pack orientation* as a continuous variable based on the vertical to horizontal ratio of product dimensions would have provided a more accurate estimate of the impact of this variable, simple dichotomous coding was felt to be sufficient for exploratory analysis to determine if this was significant predictor of VR bias across categories. In the categories studied, packs assigned a value of one typically had a vertical dimension 25% greater than the horizontal dimension.

- The number of normal shopping trips simulated in each study was coded with zero representing the reference category of one and one representing an additional shopping trip.
- The fielding mode of each study was coded with zero representing in person fielding, the default position for VR studies published prior to 2010, and one representing fielding via online panel.
- The real unit share of each product was obtained from Nielsen's Retail Scan database and recorded in percentage points e.g. 3.75 for a share of 3.75%. A number of products had market shares of or close to 0% at two decimal places. To avoid having to undertake a log transformation of this variable, the purchase shares of products with the same brand name and pack size were aggregated and *VR Bias* calculated. This necessitated the assigning of *Unique colour* and *Pack shape* variables to groups of products. The implications of this decision are discussed on page 157.
- The dependent variable (DV), *VR Bias,* was calculated for each product group using real and VR market share data and transformed into a continuous variable ranging from -100 to +300. The value assigned represented the proportional difference between products' VR shares and real shares as a positive or negative integer. For example, a value of -75 indicated the VR share of the product group was 75% smaller than its real market share; a value of +75 indicated the product group's VR share was 75% greater than its real market share.

Descriptive statistics for each variable are reported in Table 67, Appendix 7.

Exploratory modelling

To determine if linear regression was a suitable, the relationship between each independent variable (IV) and the dependent variable, *VR Bias,* was investigated. This was done using correlations (Persons' *r*), simple linear regression and scatter plots. Analysis was undertaken for each product category to provide a foundational understanding of the data and the extent to which data derived from multiple categories would describe the patterns evident in the data.

Investigation of linear relationships

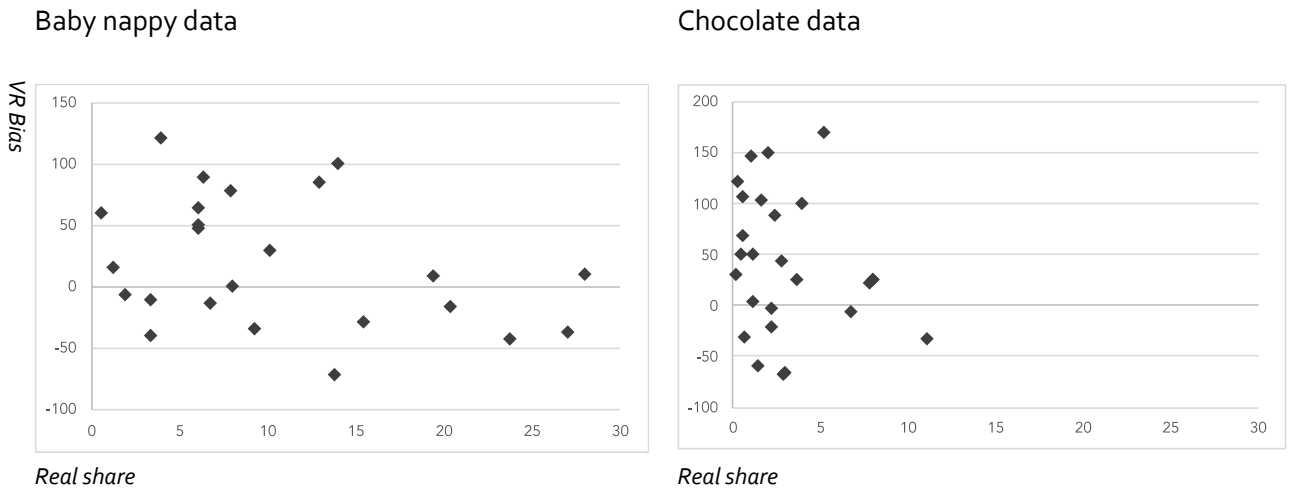
Analyses presented in Chapter 6 presumed that there was a relationship between *Relative price* and *VR Bias,* but did not test this presumption directly. Correlation analysis confirmed the presence of a positive linear relationship between *Relative price* and *VR Bias* in all seven product categories. Across the categories examined, the strength of the relationship ranged from weak (r = .16 for chocolate data) to strong (i.e. r = .84 for pet care data), suggesting this variable is likely to have the largest single effect on *VR Bias* when the impacts of multiple variables are quantified. *Relative price* had a significant effect on *VR Bias* in a simple regression using data from all product categories (p < .001).

Positive linear relationships (e.g. r = .12 to .41) were evident between *Relative pack size* and the DV for the majority of the product categories analysed, suggesting that the pack size variable coded provided a reasonably proxy for pack visibility. However, negative relationships were evident between *Relative pack size* and *VR Bias,* in two of the product categories analysed: rice (r = - .06) and

chewing gum (r = - .44). In the rice study, large packs typically laid flat on the shelf while small and regular packs stood in boxes on shelf which had more visible surface area. Similarly, in the chewing gum study, the largest pack size by unit count/weight were bottles of gum that had less visible surface area than multi-packs of gum. These differences will be considered in analyses for this variable.

Negative linear relationships were evident between *Brand type* and *VR Bias* in all product categories. Across product categories, the association between *Real share* and *VR bias* ranged from very weak (i.e. r = - .05) to strong (r = - .68), suggesting it may be a significant factor in predicting *VR bias* when data from multiple categories is aggregated. Scatter plots for *Real market share* and *VR bias* are shown in Figure 20 for two product categories. They demonstrate the pattern evident across all product categories: biases are larger for product groups with small market shares than for product groups with larger real shares. In simple regressions of these variables against *VR Bias,* both variables were significant predictors of *VR Bias* (*Brand type* and *VR Bias*, p < .001, *Real share* p < .002).

Figure 20: Scatter plots for *VR bias* and *Real share* biases



Product groups with small market shares in real stores showed more variance (i.e. vertical spread) in the dependent variable, *VR Bias,* than product groups with larger real shares.

Linear modelling of *Unique [pack] colour and Pack orientation* against *VR Bias,* found weak to moderate positive relationships between *Unique colour* and *Pack orientation* and *VR Bias* in the chocolate, feminine care and nappy data and weak to moderate negative relationships in the paper towel and chewing gum data. No relationship ($r < .05$) between these variables and *VR Bias* was evident in the rice and pet care data, suggesting these variables may not be meaningful for all product categories. These were product categories where only a few brands deviated from the colours and pack shapes typical for the product category. This is notable as prior studies have found that objects that are orientated differently to others on the shelf attract more attention irrespective of their shape (Itti, Rees et al. 2005), so stronger negative correlations were expected. The variance evident in the strength and directionality of results across categories will be considered in the interpretation of results. In simple regressions of these variables against *VR Bias,* only *Unique colour* was significant ($p = .01$).

To confirm that the aggregation of data into brand-size groups had not reduced the correlations evident between *Unique colour* and *Pack orientation* and *VR Bias,* correlation coefficients (r) from aggregated and disaggregated data were compared. Results for four product categories are presented in Table 40. To determine if aggregation affected results significantly, the *r* values were transformed into Fisher's *z* values and p-values calculated (two-tailed test of significance) Aggregation was not expected to affect results as in most of the categories analysed, product variants marketed under the same brand name had similar pack colours and orientations, so the aggregation of data into product variant groups would not affect the proportion of data points assigned each value. Additionally, there was a good theoretical basis for aggregating results for different product variants. Unique colours were expected to attract visual attention to an area of the shelf and this may benefit sales of products in close proximity. In the studies analysed, this was typically product variants of the same size. However, observational research is recommended to

confirm this assumption as it is also possible consumers' eyes saccade between visually interesting products and do not notice other products.

Table 40: Impact of aggregation on correlation coefficients

	Aggregated data *r*	Disaggregated data *r*	*Impact*	
	Pack orientation & VR Bias			
Petcare	-.02	-.06	N.S. different given sample sizes $p = .43$	
Rice	.05	.33	N.S. different	$p = .29$
Chocolate	.28	.15	N.S. different	$p = .29$
Feminine	.25	-.01	N.S. different	$p = .15$
	Unique Colour & VR Bias			
Petcare	.02	.06	N.S. different given sample sizes $p = .43$	
Rice	-.05	-.27	N.S. different	$p = .20$
Chocolate	.59	.06	Significantly higher when aggregated $p > .01$	
Feminine	.30	.03	Significantly higher when aggregated $p = .10$	

Aggregating data into brand-size groups did not significantly affect the correlation between the pack shape and colour variables and VR bias in most cases; aggregated data will be used in modelling.

The results in Table 40 indicate that aggregating results for product variants of the same size was unlikely to be responsible for the low correlations evident between *Unique colour* and *Pack orientation* and *VR Bias.* In fact, aggregation significantly increased the correlation between *Unique colour* and *VR Bias* in the chocolate category ($r = .59$ vs. .06, $p = .01$) likely by reducing the impact of *Real share* and reducing noise in the data. These impacts will be considered when interpreting results.

Investigations of collinearity and interactions

Correlation values for all variables are reported in Table 68, Appendix 7. A number of variables were found to be moderately or strongly correlated with each other ($r > .43$ to .75) including the type of retail environment simulated (*Grocery_store* or M*ass_store*), fielding mode and the number of trips simulated.

The collinearity of these variables was a result of the small number of variable combinations covered by the eight studies coded for analysis. For example, the majority of the studies that simulated multiple trips, used grocery store environments and none of the studies that were fielded online simulated a grocery store environment. As establishing the impact of fielding mode and the number of trips simulated were key objectives of this research (RQ4 and RQ5), the retail

environment variables were dropped from further analysis after confirming that the Variance Inflation Factors (VIF) of these variables were large (e.g. >5) when they were included in a regression model.

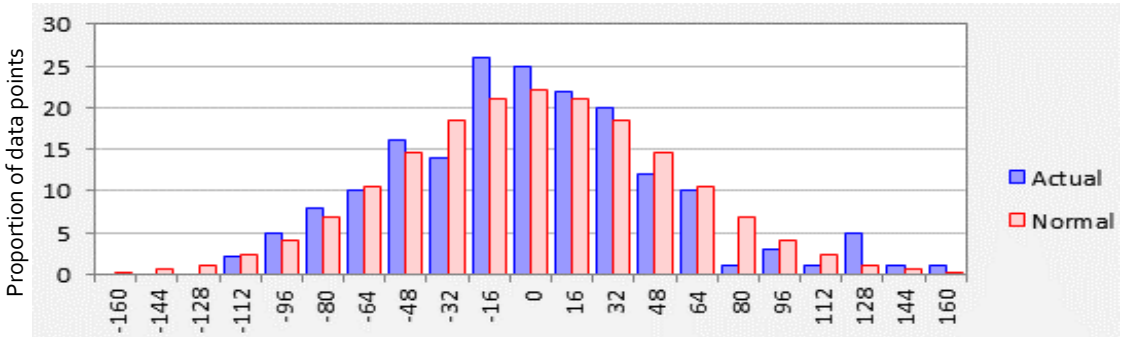
Unsurprisingly, the interaction variable *Hedonic:Price* was strongly correlated with the main pricing variable, *Relative price* (i.e. r = .74). Means centering was ineffective in reducing the correlation between *Relative price* and *Hedonic:Price.* However, analysis found that the inclusion of the interaction variable did not significantly inflate the variance of the model or provide misleading results. This was confirmed by calculating a model that did not include the interaction term; results for the remaining variables did not differ.

Additionally, the number of shopping trips simulated was found to be moderately correlated (e.g. r = .57 to .31) with the product category variables *PCatg.1-5* and to inflate the variance of parameters when included in models due to the small number of studies in the data set that simulated more than one normal shopping trip. This will be considered in analyses.

Investigations of normality and homoscedasticity

A linear regression equation was estimated to ensure that the residuals of the remaining variables were normally distributed and homoscedastic. In this equation, all variables except *Hedonic:Price* were tested for main effects. The probability that the residuals followed a normal distribution was found to be low (e.g. p = .01) using the Anderson-Darling test. The residuals were examined for normality and homoscedacity using plots presented in Figure 21, Figure 22 and Figure 23.

Figure 21: Distribution of model residuals: actual and normal distributions

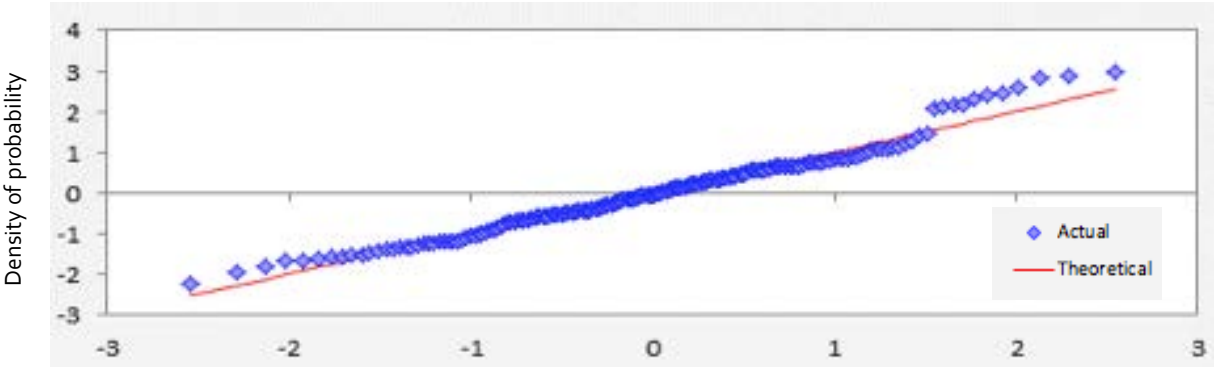


Size of residual (Percentage points, *VR Bias*)

The distribution of prediction errors was slightly positively skewed with a heavy tail of overprediction errors.

The histogram in Figure 21 demonstrates that the actual residuals of the data deviated slightly from a normal distribution at the tails of the distribution. This was particularly apparent when the residuals were transformed and plotted against a normal probability function (Figure 22).

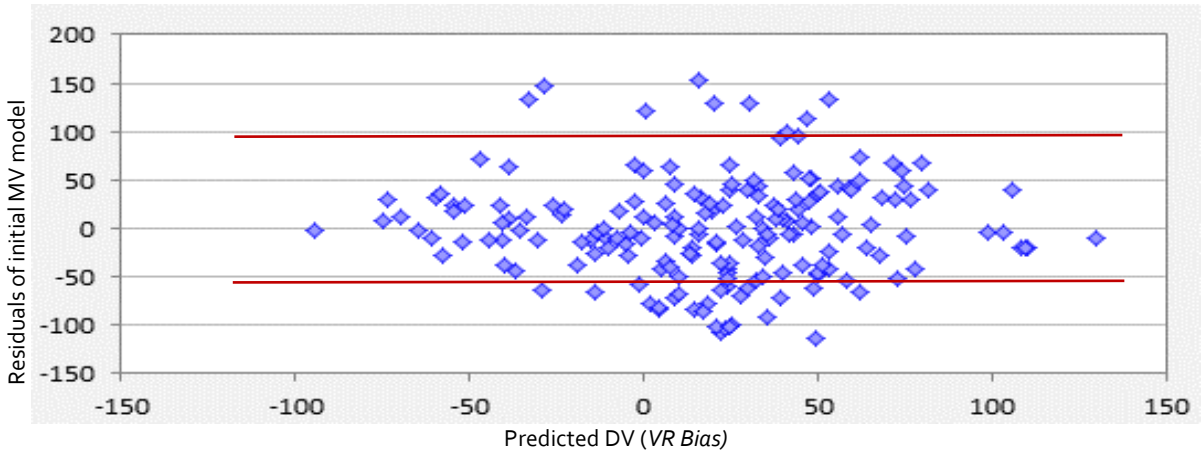
Figure 22: Normal probability plot of actual and theoretical data



Residuals with extreme positive and negative values (i.e. <-120, >120) deviated from a normal distribution; this was not of major concern.

The normal quantile plot in Figure 22 highlights that the residuals of a small number of data points included in the model (*n* = 18) points deviated from a normal distribution. The negative deviations were found to be due to instances where the model predicted small positive values for *VR Bias* (i.e. +22%) but negative values (i.e. -60%) were apparent in the data set. The positive deviations were due to instances where the model predicted small positive values based on factors included in the model (i.e. +21%), but large biases (i.e. +170%) were apparent in the data. These deviations were not drawn disproportionately from any one product category or VR study. The homoscedasity of the error term was investigated by plotting the residuals of the model against the DV in Figure 23. The terms of this model are reported in Appendix 7, Figure 31. They are not substantially different to the terms of the model reported in Table 41.

Figure 23: Plot of residuals against the predicted values of *VR Bias*



The variance of residuals (i.e. the model's error) was larger for predicted values between -50 and +50% than for more extreme values of Y (Predicted *VR Bias).*

Figure 23 demonstrates that the variance of the residuals was not equal across the range of *VR Bias* values. Levene's test confirmed that estimates of *VR Bias* between -10 and -50, -50 to +50 and +50 to +140 differed significantly in their variance (F (2, n = 182) = 5.01, p = .007). Levene's test was used

to assess the equality of variances as it is robust to non-normal distributions (Bar-Ilan, Mat-Hassan et al. 2006).

Residual scatter plots for all IVs are shown in Appendix 7, Figure 31. Residuals were found to be heteroscedastic for:

- *Brand type*, varying more substantially for national brands than private label brands (F (1, n = 182) = 12.7, p = .0004).
- *Real share*, varying more substantially for products with real unit shares of less than 5% (F (1, n = 182) = 10.5, p = .001).
- *Pack colour*, varying slightly more for products with unique pack colours relative to their environment than products with prototypical pack colours (F (1) = 16.2, p = .008).
- *Fielding mode*, varying more substantially for studies fielded in person than estimates for studies fielded online (F (1) = 13.0, p = .0004). This is likely a consequence of collinearity between variables. For example, proportionally more of the estimates for in person studies than the online studies had small real shares (i.e. < 5%), a factor which affected the variance of residuals (F (1, n = 182) = 10.5, p = .001).

Residuals were relatively homoscedastic for other variables including *Relative pack size* (F (2) = 1.4, p = .24), *Pack Orientation* (F (1) = .12, p = .72), *Relative Price* (F (3) = .16, p = .91) and the *Number of trips simulated* (F (1) = 2.1, p = .14). Robust standard errors were calculated to account for the heteroscedastic and non-normal distribution of the residuals (White 1980).

Results: Regression modelling

A significant regression equation was found (F(14,166) = 7.63, p < .001) with an with an R^2 of .64. The parameters of this model are reported in Table 41. A number of the variables included were significant predictors of *VR Bias* at the significance level discussed in Chapter 5 (i.e. p < .10 with no Bonferroni correction). The p and standard error (*SE*) values reported in Table 41 and subsequent tables account for the heteroscedastic and non-normal distribution of the data (White 1980).

Table 41: Regression model: coefficients and *significance*

		Standardised β	β	*SE*	*p* value
Significant variables	***Relative_price*** Av NB = 0	.38	1.2	0.3	.00
	Real_Share percentage points	-.20	-2.1	0.7	.01
	PCatg.6 Rice = 0	.16	34.5	18.3	.06
	Unique_colour No = 0	.13	17.7	9.3	.06
	Brand_type NB = 0 PL = 1	-.14	-28.9	16.2	.07
Variables with not significant individual effects	***PCatg.2*** Chewing gum = 0	.19	43.6	31.3	.17
	Relative_pack_size 0 = Regular	.09	14.0	4.6	.20
	Constant	-	-63.4	94.6	.22
	PCatg.5 Petcare = 0	.09	17.2	18.2	.35
	PCatg.3 Feminine hygiene = 0	.08	12.7	15.6	.42
	Fielding_mode In person = 1	.06	12.2	15.8	.44
	Pack_orientation Vertical = 1	-.03	-4.2	9.6	.66
	PCatg.1 Confectionery = 0	.08	12.5	30.4	.71
	Number_of_trips_simulated	.08	10.9	30.4	.72
	PCatg.4 Baby diapers = 0	-.00	-0.5	30.1	.98
	Price:Hedonic Interaction	-.00	-0.0	0.4	.99

The *p* values reported in Table 41 indicate that five variables were significant predictors of the bias in VR purchase estimates at the nominated level of $p < .10$[14] including:

The relative price of the brand ($p < .001$), with the price of a brand relative to other national brands included in the study having the largest effect (Standardised β = 38%) in the model estimated using unstandardized variables. This is not surprising given the strength of the correlation between this variable and the dependent variable, *VR Bias* ($r = .52$) in the data set and the large range exhibited by this variable (StD 22, range -53 to +63, Table 67, Appendix 6). The interaction term included to determine if the prices of brands had more effect in hedonic product categories, was not significant ($p = .99$). When the interaction term was dropped, the significance of *Relative price* remained high (i.e. $p < .001$).

1. The real market shares of products ($p = .01$). This variable was estimated to have a relatively large effect (Standardised β = -20%) in the model estimated and to reduce the magnitude of *VR*

[14] As the overall objective of this research is to identify variables that should be considered in, and addressed by, future studies, a relatively low level of significance was adopted. This is discussed in Chapter 5.

Biases as its value increased (β = -2.1). This is consistent with the effects observed in linear modelling of this variable (Figure 20).

2. Whether a product featured a unique colour that was present in 5% or less of the environment's total pixels (p = .06, Standardised β = .13). The wide error margins around this variable (SE = 9.3) encompass the variance evident across products and product categories in the impact of this variable on the aggregated shares of product variants.
3. In some instances, the product category from which data was drawn. The significance of the *PCatg.6* variable (p = .06) is likely due to VR data for this category deviating more from real sales benchmarks than data or the other categories studied (M_{VRbias} = 86% points vs. 40-60% points for most other categories).
4. Independent of price, whether a product was a national or private label brand affected *VR Bias* (p = .07, Standardised β =.14), with PL brands typically having negative VR biases i.e. underestimating their real shares.

Additional analysis suggested that how PL brands are marketed (i.e. standalone brands, store or generic brands) did not significantly influence the significance or effect of this variable in the model. This is consistent with the observations in Chapter 6 regarding the effects of PL brand type on VR bias. When standalone PL brands were classified as NBs, the standardised effect of *Brand type* was -.138, a value not significantly different to the effect of *Brand type* when they were classified as PLs (Table 41). The 95% confidence intervals of these effects are not significantly different when estimated via bootstrapping (Geoff 2011).

Significance of other variables

The p values reported in Table 41 indicate that a number of variables explored in descriptive analyses were not significant predictors of *VR Bias* including:

- *PCatg.1-5*, the variables that categorised data from VR studies for confectionery, chewing gum, feminine hygiene, baby nappy and pet care products (p >.10). This was because the bias evident in these studies was very similar.
- Whether products had a horizontal or vertical pack format (p = .75). This is consistent with the weak correlation evident between *VR Bias* and this variable in the aggregated data analysed (see Table 40).

Significance of pack size variable

As preliminary investigations had found that in two of the seven product categories analysed, rice (*PCatg.6)* and chewing gum (*PCatg.2),* the *Relative pack size* variable did not provide a suitable proxy for pack surface area/visibility, the variable hypothesised to be responsible for the over selection of large and extra-large packs in VR studies, a regression equation was estimated without the data from these two categories. A significant regression equation was found (F(13,147) = 7.4, p < .001) with an R^2 of .64. The coefficients and p values of the variables in this model are reported in Table 42.

Table 42: Regression model from selected product categories: coefficients and *significance*

		Standardised β	β	*SE*	*p* value
Significant variables	***Relative_price*** Av NB = 0	.32	0.9	0.4	.00
	Relative_pack_size 0 = Regular	.24	18.7	6.0	.00
	Unique_colour No = 0	.19	23.2	9.6	.01
	Brand_type NB = 0 PL = 1	-.17	-29.5	14.7	.05
	Real_Share percentage points	-.14	-1.5	0.9	.10
Variables with not significant effects individually	***Number_of_trips_simulated***	.19	23.9	17.2	.28
	Constant	-	-16.9	67.8	.41
	Fielding_mode In person = 1	.07	12.1	12.5	.37
	Pack_orientation Vertical = 1	.03	3.0	9.8	.72
	PCatg.3 Feminine hygiene = 0	-.03	-3.5	15.6	.84
	PCatg.5 Petcare = 0	.02	3.0	18.2	.85
	PCatg.1 Confectionery = 0	.02	2.7	28.6	.93
	PCatg.4 Baby diapers = 0	.01	1.7	15.7	.95
	Price:Hedonic Interaction	.00	0.0	0.4	.97

Excluding data from product categories where *Relative pack size* did not capture differences in the visibility of products, increased the statistical significance and impact of this term.

The results in Table 42 demonstrate that the relative volume/weight of products was a significant predictor of *VR Bias* (p = .002) when data from six grocery categories where larger pack sizes had a larger visible surface area (see for example Figure 15). In these product categories, each doubling of the pack size, increased *VR Bias* by 18.7% points (+/- 12% points for 95% of data points). What constituted a 'regular' pack size for each product category (e.g. 3 rolls of paper towels, 44 baby diapers) is listed in Table 39.

Interestingly, the p value for *Real Share* (p = .10) in Table 42 suggests that the significance of this variable in regression equations for VR data specific product categories may depend on the strength of the association between *Real Share* and *VR Bias* in the product categories analysed, or other factors such as the sample size of the study. Rice, a product category omitted in Table 42, was a category for which the correlation between these variables was strong (i.e. r = -.67) compared to most of the product categories studied (i.e. r = -.20 to -.40). Why the correlation between these variables was stronger in the VR study for this product category was not clear. This study did not

have a particularly small sample size (n = 253 purchases) or large number of brands (n = 11) compared to the other studies examined (see Table 11). Further analysis is recommended for this variable.

The p values for *Brand type* (p = .05) and *Unique colour* (p = .05) indicate that the significance of these variables as predictors of *VR Bias* was more robust to the exclusion of data from some VR studies/categories.

Significance: Fielding mode and number of trips

The p values reported in Table 41 and Table 42 indicate that the number of normal shopping trips simulated (p = .72 and .28) and whether the VR study was fielded in person or via online panel (p = .44 and .37) were not significant predictors of *VR Bias.* This is consistent with the results of univariate descriptive analyses in Chapter 6 and 7.

As including the *Number of shopping trips* and *PCatg.1-5,* variables that were strongly correlated, in the same regression model increased the error term and may have affected the significance of the p value for *Number of shopping trips*, a regression equation was estimated without *PCatg.1-5,* product category variables that were not significant (p > .10). A significant regression equation was found (F(10,181) = 10.7, p < .001) with an R^2 of .62. The parameters and p values of the variables included in this model are reported in Table 43 and are similar to those reported in Table 41.

Table 43: Regression model to estimate precisely the significance of the number of trips simulated

	Standardised β	β	*SE*	*p* value
Relative_price Av NB = 0	.41	1.2	0.3	.00
PCatg.6 Rice = 0	.21	45.2	14.8	.01
Real_Share percentage points	-.14	-1.5	0.6	.01
Brand_type NB = 0 PL = 1	-.13	-25.9	12.7	.04
Unique_colour No = 0	.12	15.4	8.6	.08
Relative_pack_size 0 = Regular	.10	6.4	4.3	.13
Pack_orientation Vertical = 1	-.07	-9.4	9.1	.29
Constant All other categories	-	9.9	11.6	.39
Fielding_mode In person = 1	.05	10.8	0.9	.37
Number_of_trips_simulated	.05	10.4	0.6	.49
Price :Hedonic Interaction term	-.02	-.09	0.4	.79

The *Number of trips simulated* remained an insignificant variable (p > .40) when variables that inflated the variance for this term were dropped. It was confirmed that the data set had sufficient statistical power for the impact of the *Number of trips simulated* to be statistically significant at *p* <.10 if the effect size had been consistent with prior studies. That this variable had a very small, non-significant effect on the magnitude of *VR Bias* (i.e. 5%) was consistent with univariate analyses presented in Chapter 6 and 7.

That the relative volume/weight of products was not a significant predictor of *VR Bias* (*p* = .13) in this model was not surprising as it included data for product categories where this variable was not a suitable proxy for the visibility of packs. Other variables including *Relative price, Real Share, Brand type* and *Unique colour* remained significant predictors of *VR Bias* (*p* = .001 to .08). Typically, their estimated impacts on the DV, were higher in this model than in Table 41 as fewer variables were included in the regression equation.

Discussion of results

The regression equation estimated without the product category variables *PCatg.1-5,* had the best fit (R^2 = .62) and the lowest average error (48%, StD 61%) of the three models estimated. However, the estimation errors of this model are likely to exceed what is acceptable to users of VR data (i.e. 20% points deviation, Burke, Harlam et al. 1992). Furthermore, the anomalies evident in the

distribution of the models' residuals (see Figure 23 for example) suggest that the model failed include at least one variable that is important for predicting the magnitude of *VR Bias.*

In all three models estimated, *Relative price* was a significant predictor of *VR Bias* ($p < .001$, Standardised $\beta = .32$ to $.41$). While the causes of biases (RQ3) were not examined directly in this chapter, it is possible examine the evidence that supports different explanations.

Firstly, in VR studies shoppers may select the products they aspire to purchase, the explanation favored by Higaki (2004) to justify why VR data should be calibrated based on the real prices of products. Secondly, consumers select more expensive products in VR simulations to create a favorable impression with researchers (i.e. SD bias). This was suggested in a number of past studies (e.g. Burke et al. 1992, Campo et a. 1999), but the evidence to support this explanation is scarce. For example, fielding mode is known to influence the magnitude of SD biases (Colombotos 1969, Krysan, Schuman et al. 1994, Presser and Stinson 1998, Holbrook, Green et al. 2003, Duffy, Smith et al. 2005, Chang and Krosnick 2010), but was not a significant variable in any of the models estimated.

An explanation for which there is more evidence is that in VR simulations that do not have real purchase consequences, the decision-making process is abbreviated and consumers fail to eliminate products they notice based on their price, size, quality or other factors (Bettman 1979, Russo and Leclerc 1994, Senter and Wedell 1999, Shams, Wästlund et al. 2012, Shi, Wedel et al. 2013). This may be why purchase rates and basket sizes in most VR studies far exceed those of real shopping trips (Burke, Harlam et al. 1992, Kleinschmit, Reid et al. 2012, Needel 2013, van Herpen, van den Broek et al. 2016). This explanation is supported by the significance of distinctive colours and large objects in the models estimated. These variables are known to attract attention but not typically influence the grocery choices of experienced consumers (Janiszewski 1998, Wedel and Pieters 2000, Folkes and Matta 2004, Van Der Lans, Pieters et al. 2008).

That consumers fail to eliminate more expensive products they notice by considering pricing information is supported by visual attention data collected in prior studies and two of the VR studies analysed. Desmet, Bordenave et al. 2013 found that the frequency of price consultation was much lower in a VR simulation compared to a mock store simulation (9% of purchases vs. 39%, $p < .03$), when consumers were asked to recall if they looked at the prices of products. Additionally, in two of the VR studies analysed, less than five percent of the fixations recorded ($n = 667$ consumers, 47,437 fixations), were on the shelf tags that contained pricing information, a lower proportion than in real stores (15%, Lindberg, Lindström et al. 2013). How the eye tracking data was collected and analysed is discussed in more detail in Chapter 5, page 91 and Chapter 7, page 147.

Similarly, a number of explanations are possible for the significance of *Brand type* (PL or NB), in addition to *Relative price*. Firstly, consumers may avoid PL products in VR simulations to create a favourable impression with researchers. However, as discussed in the previous paragraph and

chapters there is limited evidence to support this explanation. Alternatively, consumers may be less likely to notice PL products if the packaging of these products is less visually interesting. There is some evidence to support this conclusion. In the data analysed, 14% of the 21 PL brand groups analysed included a unique colour, whereas 47% of the national brand groups analysed included a unique colour (sig. different, $p < .005$). While *Unique colour* was accounted for in analyses, other packaging design elements that may attract attention (Grossman and Wisenblit 1999, Underwood, Klein et al. 2001, Ampuero and Vila 2006, Silayoi and Speece 2007), such as pictures, graphics, typography and layout were not investigated. The use of these features may differ across NB and PL brands (Romaniuk 2016).

The orientation of packs (horizontal or vertical) was not a significant predictor when data from multiple categories was analysed ($p = .29$ to $.72$). It is possible that this variable will be significant in models for individual product categories,[15] particularly if this variable is coded as a ratio rather than a dichotomous categorical variable. However, due to the small size effects for this variable (Standardised $\beta = -.07$ to $.03$), samples sizes would have to be large (e.g. > 200 products for each category) for results to be statistically significant. Results for individual product categories may be significant as it is possible that the effect of this variable differs across product categories depending on the prototypical orientation for packs in each category. Prior studies have found that objects that are orientated differently attract attention (Itti, Rees et al. 2005), irrespective of their other features.

The significance of *Real share* as a predictor of VR bias was not surprising: deviations in the shares of small products are typically large when reported in percentage terms. More research is needed to determine if other variables such as the number of purchases in a study (sample *n),* moderate the effect of this variable.

Chapter Summary

The results presented in this chapter demonstrate that a number of the variables identified in descriptive univariate analyses were significant predictors of deviations between real and VR market share estimates, including:

- The relative prices of products. This variable had a large effect on *VR Bias* in all of the models estimated. Whether products were functional or hedonic in nature was not found to increase the effect of this variable ($p > .10$).
- Whether the product was a national or a private label brand. Analyses found that this variable had a relatively small effect (e.g. Standardised $\beta = .13$-$.17$) compared to other variables included in analyses.

[15] Or in multi-category models with interaction terms for *Pack orientation* for each of the product categories analysed.

- Pack size, in product categories where packs that contained large volumes/weights of products were more visible on shelf. In these product categories, this variable had a relatively large effect (e.g. Standardised β = .24).
- Uniqueness of pack colour, with colours that are unusual for the product category (e.g. orange for chocolate bags, or purple for paper towels), increasing *VR Biases*. This variable was not identified in prior studies and had a moderate effect on the magnitude of VR biases (e.g. Standardised β = .13 to.19) when compared to other variables.
- In some instances, the product category examined. Product category variables such as repertoire size and category buying frequency that may explain why one product category was significant, were not explored in this chapter but the product category that was significant is examined in Chapter 9. This may assist with identifying variables that should be considered in future research.

Whether the retail environment simulated was a significant predictor of *VR Bias* was not examined due to multicollinearity issues. However, as descriptive analyses reported in Chapter 6 and 7 suggested that biases for large packs were larger in environments where large packs are relatively less popular such as pharmacy stores, so future studies should, if possible, explore if the retail environment simulated significantly affects the magnitude of biases. This has not been researched in any prior studies.

Multi-variable regression modelling confirmed that fielding studies in person rather than online and with just one 'normal' shopping trip, did not affect the magnitude of biases observed. This confirms the results of univariate descriptive analyses in Chapter 6 and 7.

As this chapter used aggregated data to confirm which variables were significant predictors of *VR Bias*, additional testing with product level data is recommended.

CHAPTER 9

STUDY 3: SHELF POSITION RESEARCH

Overview

This chapter examines whether virtual simulations can provide results that are consistent with what happens in real stores when shelf layouts are changed, in spite of the biases evident in shoppers' choices. This is explored using data from two VR studies that examined whether changing the layout would benefit sales of the product category and some brands.

Overview of theory

Effect 1: Organising the category will improve sales

Past studies have found that organising large assortments by product, flavour or even pack colour can make shopping easier for consumers, and increase consumption by as much as 100% in laboratory studies (Simonson and Winer 1992, Kahn and Wansink 2004). In field experiments in real stores, organizing the products on display typically increases category sales by 5% or 6% (Drèze, Hoch et al. 1994; results for multiple categories reported).

Effect 2: Moving a product or segment to the start of an aisle or large assortment will increase sales

In physical grocery stores, products placed in more convenient locations, such as the entrance of an aisle are more likely to be seen and selected by shoppers than products placed in the middle of an aisle (Drèze, Hoch et al. 1994, Larson, Bradlow et al. 2005, Sorensen and Suher 2010). Similarly, on online shopping sites, products included on the first page of search results are 58% more likely to be selected than items listed on subsequent pages (Breugelmans, Campo et al. 2006, Anesbury, Nenycz-Thiel et al. 2016, Mean effect size .58).

In past studies conducted in real physical stores, sales increased by between 2% and 52% when products were moved to more visible locations (Mean effect +17%, SD = 18% ;Drèze, Hoch et al. 1994, Chung, Schmit et al. 2007, van Nierop, Fok et al. 2008). However, in virtual simulations these effects may not be apparent if shoppers take the time to browse the entire aisle and continue to select products after satisfying their inventory requirements (Campo, Gijsbrechts et al. 1999, Kleinschmit, Reid et al. 2012, Massara, Porcheddu et al. 2014, van Herpen, van den Broek et al. 2016).

Effect 3: Moving products to shelves at eye-level will increase sales

In physical stores products on shelves that are around chest height (i.e. 90 to 110 centimeters for most shoppers) are more likely to be noticed than products above or below this height as shoppers generally stand a few feet away from the shelf and cast their gaze slightly downwards (Drèze, Hoch

et al. 1994, Sorensen and Suher 2010, Scamell-Katz 2012). Experiments in physical stores have established that moving products to shelves that are at chest height increases sales of those products by between 10% and 100% (Mean effect +39%, SD = 30% Hubbard 1969, Drèze, Hoch et al. 1994, van Nierop, Fok et al. 2008).

Structured investigations

This chapter examines if:

1: Organising products by type increases sales in a way that is consistent with experiments in real stores. This has not been examined by any prior studies.

2: Moving products to locations that would be more convenient in real stores increases sales of those products in VR.

3: Moving products to shelves that would be at eye-level in real stores increases sales of those products in a way that is consistent with experiments in real stores.

Together these three tests address RQ6: Does VR testing provides generalisable shelf layout results?

Confirming with VR data that moving products to shelves that would be at eye-level positively affects sales is important as studies of visual attention have found that consumers tend to look at the centre of computer screens[16] (Tatler 2007, Bindemann 2010), an area below eye-level. While it would have been preferable to establish whether there were any systematic relationships between VR purchase biases and position on shelf before exploring this question, as discussed in Chapter 5 (page 79), this was not possible.

One of the VR studies validated in Chapter 6 and 7, Study 6, was suitable for validating all three effects. Pertinent details of this study are outlined below. One additional study was suitable for validating Effect 3. It is also discussed briefly below.

VR studies: Pertinent details

Study 6 was commissioned by a well-known consumer packaged goods company to determine if shelf layout changes were likely to benefit sales of a particular type of product (*Heat and serve* products) what layout they should recommend to retailers. The VR environment depicted a grocery store shelf that was 190 centimeters tall and seven meters long and included 167 different products. The study included at three distinct 'segments' or types of products—plain products that need cooking, flavoured products that need cooking and *heat and serve* products that only need heating.

16 The possibility that shoppers' visual attention would follow an "F" or "Z" shape, the reading patterns typically used for webpages and magazine advertisements, was also considered but rejected. It is unlikely that shoppers use the same patterns that they use to make sense of text-based webpages and advertisements with structural elements like headlines and taglines to make sense of visual information that lacks these structural elements i.e. a supermarket shelf with packs arranged equally on shelves. It is however that shoppers use these patterns when viewing individual packs (Clement, 2007), but this was not the focus of the study.

These segments were identified through duplication of purchase (DoP) analysis of purchase panel data (Banelis, Kennedy et al. 2011, Dawes 2016).[17]

The *Sample* selected for Study 6 was representative in terms of the number of participants that had purchased *Plain, Flavored and Heat and Serve* products in the last six months. In Table 44, the proportion of participants that reported purchasing each product type was not significantly different (X^2 (6, n = 268) = 3.09, p > .05) to the in-market penetration of these products based on data obtained from the Nielsen purchase diary panel, Homescan.

Table 44: Penetration different product types in the sample selected

Proportion of shoppers selecting:	**Real Store** 12mths	**Sample** n = 268 6mths
Flavoured	60.0%	57.5%
Plain	41.0%	44.5%
Heat&Serve	21.4%	16.6%
Total	122.4%	118.6%

In this study, consumers were assigned to one of four layouts and instructed to 'shop as they normally would'.

- Layout 1 collected baseline results (n = 85 shoppers). In this layout, products were not organised by type and the best selling brands, Brands A and B, were located on the lowest and highest shelves (see Figure 24). This layout was based on the current layout/merchandising plan for a leading grocery store. Data from this layout was reported in previous chapters.
- Layout 2 tested an alternative arrangement (n = 72) with products grouped by type (i.e. *Flavoured* products, *Plain* products etc.) and the best selling brands on the middle shelves. *Flavoured* products, the most popular type of products, were placed first in the aisle (see Figure 24, Layout 2, dashed box).
- Layout 3 was similar to Layout 2 (i.e. products grouped by *type*, best-selling brands at eye-level, except that *Heat and serve* products, a more expensive product type were placed first in the virtual aisle instead of the more popular *Flavoured* products (see Figure 24, Layout 3, solid line box).

Figure 24: Different shelf layouts tested in Study 6

Layout 1: No organisation; best selling brands above and below eye-level

Layout 2: Organised; *flavoured* products first in aisle; best selling brands at eye-level

[17] Using data, rather than managerial expertise is critical to accurately identifying market segments as managers often presume there are segments (Kennedy and Ehrenberg 2000, Kennedy and Ehrenberg 2001). Suitable data was not available for any of the other product categories examined in this research.





Layout 3: Organised; *heat and serve* first in asile; best selling brands at eye-level



A fourth layout was tested but is not reported to protect the commercial advantage of the company that commissioned this study.

All layouts included the same range of products (n = 167). The samples for each layout were comparable in terms of shoppers' product and brand preferences (see Appendix 4: Study 6: Pantry staple) but differed in how many shoppers were included (e.g. Layout 1 n = 85, Layout 2 n = 72, Layout 3 n = 73). Consequently, total spend/purchase figures cannot be validly compared. Average spend per shopper and market share results are compared and tested for significance using unpaired t tests and contingency tests. Pearson's statistics are used in chi-squared (X^2) tests as counts for all groups and conditions were larger than 10.

Shoppers in each layout purchased multiple products (M_{items} 2.59), increasing substantially but not sufficiently, the statistical power of the study for comparisons of purchase shares. However, as discussed in Chapter 5, the experiment was underpowered given the effects sizes in reported in past studies (Drèze, Hoch et al. 1994, van Nierop, Fok et al. 2008) and number of brands included.

Study 7, the other study validated in this chapter, similar in its scale and level of visual realism to Study 6, the study outlined above. It depicted 200 feminine care products arranged in an aisle approximately 160 cm tall.

Figure 25: Example shelf layout tested in Study 7


Kotex
Kotex
Kotex

Results: Consistency of real and virtual effects

1: Organising the category by type increased sales

Table 45: Impact of organising the category

	Category Spend Av.	Flav'd Pur.Share	Plain Pur.Share	Heat&Serv Pur.Share
Category not organised n=253 purch's	$4.83	53.0%	21.8%	25.2%
Category organised n=444	$5.18	45.5%	19.6%	34.9%
Effect of change	+ 7%	- 16%	-11%	+38%

In the virtual test organising the category by product type increased demand for *Heat and Serve* products, a more expensive product type.

The results in Table 45 demonstrate that organising the category by product type benefited sales of *Heat and Serve* products (Table 45: V_{share} = 34.9% vs 25.2%, statistically significant X^2 (1) = 23.2, p < .01). That organising the category increased sales of a less popular product type is consistent with the effects of sales experiments in real stores (Dreze, Hoch et al. 1994, Kahn and Wansink 2004). Other changes in the shares of products types were not significant (p > .15).

There was no significant increase in the amount shoppers spent overall (Table 45: *Av. Category Spend* $5.18 organised StD = $4.31 vs $4.83 not organised StD = $4.70, t (228) = 0.51 p = .61). The results in Table 45 demonstrate that organising the category did not significantly affect the number of items shoppers selected in the virtual study (M_{items} 2.59 vs 2.65, StD = 1.73 vs. 1.57, t (228) = .27 p = .79) or how long consumers took to select products (M_{time} 71 seconds vs 77 seconds, StD = 56 vs 45 seconds, t (228) = .84, p = .40).

Discussion of organisation results

Although the increase in the share of *Heat and Serve* products following organisation was substantial (+38%) and statistically significant ($p < .01$), substantially more shoppers selected *Heat and serve* products in the VR simulation than reported buying these products in the last six months in the pre-shopping survey (Table 46: 36.7% vs. 16.6%, significantly different, confidence intervals do not overlap at 99% C.L.). Additionally, based on panel data from real purchases, only 14% of *Flavoured* product buyers also purchase Heat and Serve products over a 12 month period (Nielsen *Homescan* data, 2014). In the VR simulation, 28.6% of *Flavoured* rice buyers also selected a *Heat and Serve* product, a significantly larger proportion (C.I. do not overlap at 95% confidence level). These results indicate that a number of shoppers who do not regularly purchase *Heat and serve* products, selected these products in the VR study even when products were not organised by type. This may be because *Heat and Serve* products are a premium product that costs more per ounce on average than *Plain* or *Flavoured* products (Table 46: Av. price $0.21 vs $0.07 per ounce). As analyses in Chapter 8 confirmed, product prices are a significant factor in predicting VR biases. It is also possible that participants were unaware they had chosen a convenience product that was more expensive. Unfortunately, the data necessary to test this theory was not available.

Table 46: Real-VR purchasing of different product types

Proportion of participants who:	**Reported purchasing** n = 268	**Selected in VR** n = 268	**Diff. V-S Prop'n of Shoppers**
Flavoured av. price $0.21 p/oz	57.5%	63.7%	+ 6.2%
Plain av. price per pack $0.08 p/oz	44.5%	40.3%	- 4.2%
Heat&Serve av. price $0.20 p/oz	16.6%	36.7%	+ 19.9%

A larger proportion of shoppers selected *Heat and Serve* products, a more expensive product type, in the VR simulation than reported doing so in real stores.

2: Moving products to the start of the aisle increases sales

In Table 47 the effects of moving a product type to the start of the aisle are shown.

Table 47: Impact of moving products in aisle

Purchase share when:	First	Second	Sales Effect	
Flavoured products	49.1%+	39.8%	+ 23%	Stat.Sig. $p < .01$
Heat&Serve products	36.1%	33.8%	+ 7%	Not.Sig. $p > .10$

Rearranging the aisle did significantly affected the purchase share of *Flavoured* products.

Rearranging the aisle had a significant impact on sales of *Flavoured* products (X^2 (1) = 7.67, p = .005) but not on sales of *Heat and Serve* products. When *Flavoured* products were placed first in the aisle, the purchase share of these products was 23% larger than when they were placed later in the aisle. This effect is consistent in magnitude with sales experiments from real stores (Mean effect 16%, StD = 18%).

In this study there was no strong evidence of consumers selecting more products that expected from the first or middle sections of the product assortment, a pattern apparent in other VR studies that used more restricted shelf assortments (Chandon, Hutchinson et al. 2009, Massara, Porcheddu et al. 2014, van Herpen, van den Broek et al. 2016). In real stores, *Flavoured* products have approximately 55% market share and *Heat and Serve* products have 35% market share.

3: Moving products to eye-level shelves generally benefits sales

The third change that was tested was whether moving popular brands to eye-level shelves increased their sales. As category organisation affected the relative popularity of *Flavoured, Plain and Heat and Serve* products and this change was introduced concurrently with changes to shelf location shelf location, the purchase shares of brands are reported within their product type in Table 48 to control for this effect. Contingency tests (X^2) were performed on purchase counts of each brand to identify significant differences. Tests statistics for changes that were statistically significant are reported in the text that follows Table 48.

Table 48: Impact of moving brands to eye level shelf positions in Study 6

	Shelf location is:	Not eye level	Eye level	Sales Effect	
		Share of VR sales for product *type*			
Flavoured	Brand A	42.5%	45.5%	+ 7%	N.S. $p > .10$
	Brand B	34.3%	46.8%⁺	+33%⁺	Stat diff. $p < .01$
	Brand C	20.0%	18.7%	- 7%	N.S. $p > .10$
Plain	Brand A	21.1%	33.0%	+66%	N.S. $p > .10$
	Brand B	52.7%	64.8%	+23%	N.S. $p > .10$
H&S	Brand A	15.6%	23.4%	+50%	N.S. $p > .10$
Average effect				**+29%**	

Moving products to positions that would be more visible in real stores increased significantly sales of one brand, other results were not statistically significant.

The sample of purchases for each product type ranged between 56 for *Plain* products not at eye-level and 198 for *Flavoured* products at eye-level. Consequently, only large changes in brand shares for *Flavoured* products, the most commonly selected product type, were statistically significant (Table 48: Brand B X^2 (316) = 6.76, p = .01). The magnitude of this change was not statistically different to the average effect of moving products to eye level in experiments conducted in real stores (Mean effect +39%, SD = 30%, p > .10 Hubbard 1969, Drèze, Hoch et al. 1994, van Nierop, Fok et al. 2008).

In Table 49 the impact of moving brands to eye-level shelf positions in Study 7 is reported for four brands and overall. The sample of purchases ranged between 40 and 410 for each brand for which changes were tested; and the overall sample was 638 purchases.

Table 49: Impact of moving brands to eye level shelf positions in Study 7 for feminine care products

Shelf position:	Not eye level	Eye level	Sales Effect	
Brand G	4.5%	7.0%	+ 64%	Stat diff. $p < .10$
Brand C	18.0%	21.0%	+ 16%	N.S. $p > .10$
Brand B	21.0%	21.0%	+ 0%	N.S $p > .10$
Brand F	5.3%	6.0%	+ 13%	N.S. $p > .10$
Total	**48.8%**	**55.0%**	**+ 13%**	Stat diff. $p < .05$

Shelf layout changes had a significant effect on brand sales overall in Study 7.

In this study the four brands who's on shelf locations were varied had a cumulative share 13% larger when located on shelves equivalent to eye level than when located on shelves above/below eye-level. Shopper eye level is typically 90 to 110 centimeters from the floor (Drèze, Hoch et al. 1994, Sorensen and Suher 2010, Scamell-Katz 2012). This difference was statistically significant (X^2 (1) = 4.77, p = .03) as was the large difference in the shares of Brand G (share +64%, X^2 (1) = 3.23, p = 0.07 marginally significant,).

Discussion of shelf position results

In both studies it was anecdotally observed that the brands that benefited most from changes in their shelf position tended to be those that that were moved from low, floor level shelves to eye-level shelves. As the heat map of product interactions in Chapter 7 demonstrated, bottom shelves tend to receive the least amount of attention in VR studies (see Figure 18). Prior studies have also found that purchases of products located on the bottom shelf tend to be underestimated in VR studies (Porcheddu and Venturi 2011, van Herpen, van den Broek et al. 2016). Further investigation is however required.

Chapter Summary

The VR studies presented in this chapter demonstrated that organising the category by product type increased demand for a more expensive product type, a result consistent with the effects of sales experiments in real stores (Dreze, Hoch et al. 1995).

Moving products to more visible locations (i.e. the start of the aisle, eye-level shelves) did have a significant impact on sales in some instances, but further testing with larger samples is required to confirm the results. It is recommended that future studies seeking to test similar effects have at least 1000 purchases per layout to ensure a good level of statistical power (e.g. 80% power, $\alpha < .05$). This could be achieved by increasing the number of participants shown each layout or exposing participants to multiple layouts. Purchase rates from prior studies can be used to estimate the number of products each participant is likely to select (Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016), aiding comparisons.

Overall the results from this chapter cautiously suggest that VR studies may be able to accurately predict what and how much impact shelf layout changes will have in real stores, if the tendencies of shoppers to select more expensive products than they usually buy are adequately controlled.

CHAPTER 10

SUMMARY AND DISCUSSION OF FINDINGS

Overview

Chapter 10 discusses the key findings from this thesis, its implications for marketers and for VR research and makes five recommendations for future research.
It commences with a table summarising the outcomes of investigations undertaken.

Summary of quantitative findings

Table 50: Summary of quantitative investigations

Investigation	Outcome	Contexts tested
Do participants select products from their normal purchase repertoires?	Yes, but a large proportion of shoppers selected more expensive products than they usually buy (51%), larger pack sizes (36%) or different types of products (22%).	4 studies; range of product categories and retail environments.
Are biases consistently present?	Yes, biases are consistently evident for a range of products in VR studies including national brands, PL brands of all types, products with small market shares, large packs and products that feature unique pack colours.	8 studies; range of product categories and retail environments; all US data.
Are sample biases likely to be responsible for the biases observed?	No, while differences were observed between the VR samples and their buying populations in some categories/studies, the same biases evident across a range of demographic groups and behavioural groups, suggesting they are a function of the VR method.	4 studies; range of product categories and retail environments.
Are biases for premium and PL brands are more pronounced in simulations fielded in-person?	No, fielding mode was not found to be a significant factor in regression modelling and contrary to expectations, the shares of more expensive, socially desirable brands were not larger in studies fielded in person.	8 studies; range of product categories and retail environments; US data.
Do shoppers verify pack sizes in hypothetical VR studies?	No, attention to pack size information appears to be very low. Participants were not asked what pack size they had selected so further research is recommended.	4 studies; disposable nappies, confectionery, paper towels and feminine care.
Are pack-size biases because shoppers cannot feel/touch products?	Unlikely, as the same biases were not evident in online shopping data, a context where consumers (also) cannot feel or touch products.	3 studies; categories including chocolate, petcare and paper towels.

Investigation	Outcome	Contexts tested
Are biases related to the prices of products/brands?	Yes, the prices of products relative to others brands in the category was found to be a significant factor in regression modelling.	8 studies; range of product categories and retail environments.
Does the magnitude of bias vary across product categories?	The product category examined was only found to be a significant predictor of bias ($p > .10$), in one product categories examined. The type of product category (hedonic than utilitarian) was not found to increase biases for expensive brands.	8 studies; range of product categories and retail environments.
Are shoppers aware of these biases in their behaviour?	No, most participants claimed to have shopped normally and selected the products they normally buy.	2 studies with pre and post shopping survey data
Are pack-size biases because images of products are scaled down to fit consumers' laptop screens in VR studies fielded online?	No, biases for large packs are evident in studies fielded on larger computer monitors in specialised research facilities. Furthermore, fielding mode was found to be a non-significant factor ($p > .10$) in predicting the size of biases.	2 studies in descriptive analyses (feminine hygiene products); 8 in modelling
Do additional trips improve the accuracy of purchase choice data?	No, whether one or two normal shopping trips were simulated was not found to be a significant predictor of the bias. Analyses had sufficient power to identify even a small effect for this variable (e.g. 10% or greater).	2 studies with multiple trips, 6 with single, range of categories and contexts
Can VR simulations replicate the real effects of shelf layout changes in stores?	Yes, VR simulations can provide accurate results for the impact of organising large assortments, moving products to more convenient locations and eye-level shelves.	2 studies; limited number of effects tested; more research recommended

Discussion of key findings

The objective of this thesis was to determine how realistically consumers' purchase choices are captured in VR shopping simulations and what changes should be made to improve the validity of VR research for decision making in marketing.

Finding 1: Current industry VR studies are conducted in similar ways

In-depth interviews found that VR studies commissioned by marketers often employ similar protocols, instructions and surveys. Given that studies testing different instructions for VR have not been published, the similarity of the instructions used is remarkable. It is possible that high staff turnover and crossover rates in the MR industry contribute to the similarities. This suggests that more validation research could be conducted by following the approach used in this thesis.

Showing the recruitment protocols, survey questions and instructions used in industry studies, and demonstrating they are not sufficient, also makes a substantial contribution as this information is typically proprietary. Discussing how industry studies are conducted also provides a stronger foundation for future research in this domain.

Finding 2: Consumers' actual product, brand and pack size purchase choices are not captured accurately in VR simulations

Previous studies concluded that consumers' VR purchase choices reflected their real purchase behaviours because correlations between real and virtual sales data were high and most participants claimed to shop realistically. However, no studies suitably tested this assumption and a significant proportion (e.g. 90%, n = 21) of the marketing managers interviewed were concerned that VR data may not be wholly accurate. The results of this study address, in part, their concerns.

Comparisons between shoppers' VR purchases and their past purchases found that a substantial proportion of shoppers selected products from outside their usual purchase repertoire—51% of shoppers chose a brand that they had not purchased in the last month, 36% selected a pack size they claim to not normally purchase and 22% selected an entirely different type of product from the category they purchased previously (i.e. product designed to be reheated in the microwave rather than cooked on a stove). Additionally, analyses in Chapter 9 demonstrated that shoppers do not reliably select the same *types* of products as they usually buy in VR studies e.g. microwavable products, compromising duplication, penetration and market share metrics for product segments, SKUs and brands.

In all of the simulations analysed, participants were requested to "shop as they normally would" and most (i.e. 90%+) thought they had selected products consistent with their real purchase choices. This indicates that consumers are poor at assessing the realism of their own behaviours in VR environments, a finding of significance for previous studies that used claimed realism as their main measure of validity for VR data (e.g. Tjostheim and Haugland 2005, Wright, Schnack et al. 2017).

Differences in shoppers' real and VR product choices were evident across all demographic and behavioural groups validated but some groups provided substantially less realistic data, including:

- Infrequent category buyers (i.e. purchase the CPG category every three months rather than monthly, the typical inter-purchase period for the category).
- Shoppers responsible for less than 75% of their households' grocery shopping.

While these findings are not surprising, infrequent buyers are less likely to be familiar with the brands available and which their household usually buys, they indicate it may be possible to improve the accuracy of VR purchase data by recruiting only regular buyers and household shoppers. The implications of this finding are discussed on page 190.

Finding 3: Purchase choices are consistently biased towards expensive brands and products, cheaper products, and larger pack sizes

Modelling confirmed that in data for US grocery products, biases were evident for:

- Products that are cheaper or more expensive than the average national brand in the product category.
- Private label products, some of which are marketed as cheap 'value' brands and some of which are marketed as standalone brands. Modelling confirmed that biases are present for PL products, even when the impacts of price are considered.
- Brands with small shares in real stores when biases were measured in percentage terms (i.e. VR share as a proportion of real share).
- Packs that are more visible on shelf due to their surface area and/or pack colour.

These biases were evident across a range of product categories (e.g. confectionery, chewing gum, pet care, feminine hygiene) and retail environments (e.g. grocery stores, mass merchandisers such as Walmart and drug/pharmacy stores). Notably, analyses found that the price of products, a variable examined in only one prior validation study (Higaki 2004), had a larger impact on biases[18] than whether the product was a national or a private label brand, a key variable of interest in past validation studies (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013, van Herpen, van den Broek et al. 2016).

That biases were evident across a range of VR studies suggests that more methodological experimentation is needed to perfect the design of VR studies. Methodological research to date has been limited and not sufficiently systematic (Campo, Gijsbrechts et al. 1999, Breen 2009, van Herpen, van den Broek et al. 2016). Recommendations for experiments that should be conducted commence on page 195 of this chapter.

[18] In unstandardised data. It is possible that scale differences across variables (see descriptive statistics in Appendix 6) contributed to this result. Further testing is recommended.

Finding 4: Multiple trips do not substantially improve the realism of VR purchase data

Previous studies presumed that additional trips would improve the realism of consumers' purchase choices by reducing the effects of social desirability and the novelty of VR technology, on consumers' purchase choices. Modelling and descriptive comparisons demonstrated that whether one or two normal shopping trips were simulated was an insignificant factor in predicting the magnitude of differences between real and VR purchase shares. While other researchers used a greater number of trips (e.g. 12) to reduce the effects of novelty on purchase behaviours, in these studies a large impact on purchase behaviours was evident between shoppers first and second trips (e.g. 35% reduction in purchase quantities in Campo, Gijsbrechts et al. 1999). That biases for premium brands, cheaper products and large pack sizes were still evident on shoppers' second VR trip, suggests that social desirability and novelty may not be the primary causes of biases in VR choice data as previously presumed (Burke 1996, Campo, Gijsbrechts et al. 1999, Burke and Leykin 2014).

By examining the evidence for explanations that have been proffered by past studies for the biases evident in VR data including social desirability, small screens and the inability to handle products, this thesis makes a substantial contribution by directing future research to worthwhile areas. The marketing managers and industry researchers interviewed expressed a desire for better evidence to inform their decisions about how to design studies and where to focus their efforts. Findings 5, 6 and 7 address these desires directly.

Finding 5: Biases could be explained by the abbreviation of the decision making process and the lack of real purchase consequences

Using attention, gaze and purchase data from a range of VR studies, this thesis demonstrated that biases in VR shopping data may be consequence of the decision making process being abbreviated in VR studies participants failing to eliminate products they notice based on their size, price or other factors that typically influence real purchase choices (Bettman 1979, Russo and Leclerc 1994, Senter and Wedell 1999, Shams, Wästlund et al. 2012, Shi, Wedel et al. 2013).

While the cause of this bias was not directly tested in this thesis, studies in a range of fields have demonstrated that the absence of real purchase consequences, results in shoppers overestimating their preferences for expensive options (see for example Carlsson and Martinsson 2001, Lusk and Schroeder 2004, Hensher 2010). None of the VR studies available for this thesis had real purchase consequences. Testing the impact of real purchase consequences in well-designed methodological experiments is recommended as fielding a study to test this factor was outside the scope (and budget) of this thesis

Gaze tracking or qualitative research to confirm that the decision making process is abbreviated in VR studies without real purchase consequences, is also recommended.

Finding 6: VR purchase data collected online through pre-recruited research panels does not differ to data collected in shopping mall research facilities

While some of the industry researchers interviewed expressed strongly held beliefs that more reliable VR data would be collected in VR research facilities with large screens, researcher supervision and specialised equipment, the mode of fielding was not a significant factor in predicting the magnitude of biases for VR studies. Moreover, the same biases were evident in VR purchase data collected in research facilities with specialised equipment, as in VR data collected from pre-recruited Internet panellists using their own personal computers. This is consistent with findings from the HCI literature showing that screen size does not affect consumers' comprehension of or responses to VR environments (Tan, Gergle et al. 2004, Tyndiuk, Lespinet-Najib et al. 2004). If possible, future studies should validate that SD response tendencies are more apparent in VR simulations fielded online than fielded in person using a questionnaire designed to measure socially desirable response tendencies (e.g. Paulhus' 1991 Balanced Inventory of Desirable Responding).

The similarity of VR data collected online using panellists own computers with data collected using specialised interaction devices (e.g. shopping trolley joysticks) and large screens suggests that these innovations do not improve the accuracy of VR purchase data, although they may enable shoppers to navigate VR environments faster and more naturally (Slater, Linakis et al. 1996, Tyndiuk, Lespinet-Najib et al. 2004, van Herpen, van den Broek et al. 2016).

Finding 7: VR is suitable for shelf layout research

Analyses confirmed that a VR simulation of a grocery aisle shown on a normal computer screen can provide results consistent with the effects of shelf layout changes in real stores, including:

- Organising a large assortment by brand and product type (i.e. flavoured, plain etc.) benefited sales of less popular products.
- Placing premium products at the start of the aisle benefited sales of these products by a moderate amount (e.g. 23%).
- Placing popular products at eye-level increased sales of these products by a large amount (e.g. 33%).

The VR study provided estimates consistent in magnitude with experiments conducted in real stores, extending prior research that found that immersive VR technologies (i.e. head-mounted displays, wide-screen displays) can provide shelf layout results that are directionally consistent with changes in real stores (Porcheddu and Venturi 2011, Precourt 2012, van Herpen, van den Broek et al. 2016).

This research indicates that normal computer screens can be used to collect suitably accurate estimates of what may happen in real stores when shelf layout are changed, opening up future research to Internet research samples. These findings also extend, although cautiously as only two

studies were examined, the variety of shelf layout changes for which VR has been validated, to include organisation by product type and brand.

Implications for the use of VR for shopper research

Implication 1: Marketers should be cautious of using VR for observational research

Overall the findings from this thesis highlight that VR does not capture normal shopping, choice or visual attention behaviours just because the environments used look realistic. The systematic biases in VR purchase data suggest that price and pack visibility influence purchase choices differently to real purchase situations and that the decision-making process may be abbreviated in VR simulations. Coupled with previous research showing that there are substantial differences in how long consumers shop, how many products they choose and recalled attention to prices (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Waterlander, Scarpa et al. 2011, Desmet, Bordenave et al. 2013), marketers should be extremely cautious of using VR to understand how consumers behave in actual stores or what factors influence purchase decisions (e.g. decision tree research). Marketers may be better served by using learnings from observations from real stores (e.g. Scamell-Katz 2012, Sorensen, Bogomolova et al. 2017) to address their research questions.

Implication 2: Without improvements, VR is not recommended for forecasting sales

The magnitude of differences between real and VR purchase data indicate that marketers and researchers should be wary of using virtual simulations to forecast demand for products without calibrating the results to account for the biases evident and perhaps introducing real purchase consequences to counteract biases. The models presented in Chapter 8 demonstrate that calibration models for consumer goods data will likely need to account for the failure of VR shoppers to consider the (real) prices of products, the over-selection of more noticeable products, the tendency for less popular products to have very different real and VR purchase shares when measured in proportional terms and biases in the selection of PL brands. Only one of these variables, price, was used in prior calibration models for VR purchase data (Higaki 2004).

However, further research will be needed to confirm that these biases affect VR data sets collected in other markets and for other types of products, and to identify additional variables which predict biases, if accurate calibration models are to be established. The additional research needed to extend the calibration research reported in this thesis is discussed after page 195.

Until appropriate calibration models are established, VR purchase data should be considered a representation of what shoppers notice and extreme care taken in extrapolating calibrated or uncalibrated sales results to real stores.

Implication 3: VR is a good option for shelf layout research

The findings from Chapter 9 suggest that VR is a useful tool for collecting quantitative data to help convince retailers to implement shelf layout changes in real stores as it may be able to accurately predict what and how much impact shelf layout changes will have in real stores, if the tendencies of shoppers to select more expensive products than they usually buy are adequately controlled. One of the key advantages of VR over other forms of shopper research is that a variety of different shelf layouts can be tested easily to determine which is most likely to succeed in real stores. However, users should be aware that the choices of some shoppers and shares of some products (e.g. premium brands, PL products, large pack sizes etc.) will not be consistent with results from real purchase environments, particularly if VR studies are conducted without real purchase consequences.

Implication 4: VR cannot be used to determine shoppers' real repertoires

The differences between shoppers' real and virtual purchase choices mean that hypothetical VR simulations are unlikely to determine shoppers' usual purchase repertoires, generate accurate market structure data, predict the penetration of products or product types or understand the buying preferences of different groups. Interviews found it is sometimes used for these purposes in markets where reliable panel data is not available, for example in China and India (Romaniuk and Sharp 2016). However, more cost-effective approaches such as short-term purchase diaries or two-purchase surveys, are likely to provide reliable data (Bennett 2004, Bennett and Graham 2010). At this stage, VR simulations are unlikely to be a good option for collecting data about shoppers' repertoire sizes (i.e. the number of products they purchase in a given period) or switching/loyalty rates (e.g. share of category requirements) as shoppers also select two or three times as many products on each simulated shopping trip, switch brands more readily than in real stores (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, Needel 2013, van Herpen, van den Broek et al. 2016) and select products they do not normally buy. However, further methodological experiments may be able to resolve these biases.

Future research should also determine if VR would can provide accurate growth/decline forecasts for segments as grocery categories are often divided into partitions or segments of different products (Ehrenberg 2000, Dawes 2016) and marketers may desire to predict which segments of a category will be affected by changes to category layout or product choice (Treiber and Needel 2000).

Implication 5: Caution is required in using VR for assortment or product choice research

The results presented in this thesis suggest that VR should not be used to make predictions about how changes to product assortments will affect in-store sales, a purpose for which they have been used, and recommended for, in past studies (Burke 1999, Breen 2009, Precourt 2012). The biases evident in results indicate that the importance of having a variety of pack sizes, price points and/or

brand types in a product assortment is likely to be underestimated, even if results are benchmarked against a control cell because consumers fail to adequately consider these factors when selecting products in hypothetical VR experiments.

Results presented in this thesis demonstrate that asking shoppers what pack sizes they usually purchase in pre-shopping surveys and including a range of pack sizes in VR studies, is insufficient to generate accurate size-choice data. Efforts must be made to get shoppers to actively consider the size of the products they choose in VR simulations, especially those that do not have real purchase consequences. For example, by including additional prompts such as "The pack size selected contains *X* servings. Is this how much you usually purchase?", or having shoppers select their preferred product and then their preferred pack size from a modified shelf set. Until such innovations are tested, VR is not recommended for size assortment research. Other innovations will need to be tested to validate VR for the use of price range research including having participants spend, or think they are spending, real money. The methodological research that should be undertaken for VR is outlined after page 195.

In markets where reliable panel data exists, marketers should consider using panel data and the Duplication of Purchase Law to predict how changes at shelf will impact products' sales and market shares (Ehrenberg 1991, Hoek, Kearns et al. 2003, Kennedy and McColl 2012, Tanusondjaja, Nenycz-Thiel et al. 2016) instead of conducing VR experiments for different shelf sets.

Implication 6: VR may overestimate the sales benefits of having a pack that stands out on shelf

While this thesis did not examine directly the suitability of VR for pack design research, the biases evident in shoppers' purchase choices suggest that virtual tests may overestimate the sales benefits of having a pack design that stands out on shelf because of its colours, design or visible surface area. While there is limited evidence validating the accuracy of VR for pack design research (i.e. only Berneburg 2007), VR simulations are likely to be a better option than focus group or survey research as they enable the shelf context, an important factor for pack research (Young 2004), to be incorporated into analyses. Additionally, VR is likely to be an advantageous approach because of the ease with which multiple designs can be tested in a commercially secure environment, a point made by the marketing managers and industry researchers interviewed.

However, marketers should take care to include respondents who are light or infrequent buyers of the brand if using VR for pack research. Most grocery brands have many light and occasional buyers who buy once or twice a year, or less frequently. These buyers are more likely to be confused by pack changes because they have less brand recognition and longer interpurchase intervals (Romaniuk 2011), making them more likely to miss pack change information advertised in store or in market. However, light and occasional buyers make up a large proportion of sales (i.e. 60%), and are critical for brand growth (Sharp 2010, Romaniuk and Sharp 2016). Determining the impact of

category involvement on shoppers' VR purchase choices is an area recommended for future research (see page 195) as it may also influence consumers' purchase choices.

Implications for the design of VR studies

Implication 1: Real and virtual behaviours differ; a control cell is essential

VR studies testing new price promotions, signage, assortments or other changes to see if they should be implemented in real stores, must include a control cell with the same product assortment, shelf layout and pricing of real stores. This is necessary as VR results are not likely to be directly generalisable to real stores, even if their (visual) elements are reproduced accurately by VR environments.

As analyses in Chapter 9 demonstrated, the number of large packs or prominence of expensive product options can affect the popularity of specific product segments and may affect overall spend. Thus, marketers need to be conscious of inadvertently biasing results when designing test and control cells. Additional factors that influence results are likely to be identified in future studies, so caution (and research expertise) will continue to be critical to the design of VR studies.

Implication 2: Shopper studies should have larger samples (or fewer cells)

The ease with which different shelf layouts, signs or displays can be created in VR makes it tempting to test many iterations of an idea or shelf layout in a study. One industry study reported in the literature for example tested 192 different shelf layouts (Kopf and Treiber 2003). Marketers need to be cognisant of the effects of testing additional layouts on sample sizes and ensure that samples for individual layouts are not reduced to point that the sample is insufficient to detect or reject differences in results at conventional levels of significance (Sawyer 1982, Cashen and Geiger 2004).

Planning for VR studies could be improved by considering the effect size(s) likely to occur and ensuring that each layout tested has a sufficiently large sample that comparisons will have good statistical power. Marketers and market researchers should consider establishing databases of the effect sizes achieved in VR studies so accurate estimates of effect sizes to can be obtained.

Implication 3: Extra 'dummy' trips are not necessary for VR

While additional trips may have been necessary when VR environments were first utilised in the 1990s, they do not appear to improve the accuracy of contemporary market share data. The shares of most brands and pack sizes did not differ substantially between consumers' first and second trips and the number of trips simulated was not found to be a significant factor in predicting the magnitude of biases. As additional trips that are conducted for research purposes increase the fielding time for which each respondent needs to be compensated, without enlarging the sample, VR research budgets should instead be used to test other changes that may improve the accuracy of VR data.

Implication 4: VR studies should experiment with real purchase consequences

The gaze tracking data presented in this thesis demonstrated that very few consumers (i.e. less than 10%) paid attention to the prices or pack sizes of the products they selected in VR studies and that biases were consistently evident for more expensive and more noticeable products in VR data, even when the effects of multiple factors were considered. Additionally, the absence of variation in shoppers' product and pack size choices across household and income groups, suggested that economic concerns and consumption rates failed to inform shoppers' VR purchase choices. These issues suggest that researchers need to experiment with giving VR studies real purchase consequences to determine if this improves the validity of VR data and the realism of decision making processes.

Testing the impact of purchase consequences in well designed, methodological experiments is recommended as a priority for future research. This discussion commences on page 195.

Implication 5: Haptic feedback gloves are unlikely to be necessary to obtain realistic VR grocery choice data

Comparisons between VR and real online shopping data in Chapter 7 demonstrated that the purchase shares of larger pack sizes were significantly different in VR studies and real online shopping data. This suggested that the absence of real purchase consequences, rather than the inability handle products in the VR studies tested, was more likely to be responsible for the biases observed in VR data.

This suggests that before testing innovative technologies such as haptic feedback and baskets that enable participants to feel the weight of the products they select, real purchase consequences should be tested in VR experiments to determine if they can resolve the biases observed.

Implication 6: VR environments could be designed with fewer 'bells and whistles'

A number of the findings from this thesis suggest that VR simulations could be designed more simply and cost-effectively. For example, VR studies could be designed so that products look 3D on shelf (i.e. have depth) but are not interactive three-dimensional objects that can be selected and rotated (see Figure 10) as less than 10% of consumers utilise this feature in studies for packaged grocery products. This would be appropriate for grocery categories where few shoppers pay attention to nutrition information on the back or sides of packaging in real stores (Sewell and Komogortsev 2010) and would reduce significantly the time required to prepare VR environments and bandwidth required to run them online.

Similarly, univariate analyses in Chapter 7 found that the shares of large and regular pack sizes were not significantly different in VR studies that simulated different retail environments.[19] This suggests that until more research is conducted, it may be wise for researchers to use relatively generic VR shopping environments, rather than expending resources on visually tailoring VR environments to evoke *specific* retail settings e.g. a specific Walmart store. As discussed in Chapter 3 and advertised in industry publications (Breen 2009), some research agencies have ready-made generic VR shopping environments that can be used for research.

This recommendation is supported by research in psychology and consumer behaviour. Psychology researchers have found that with appropriate priming, fairly rudimentary looking 3D VR environments are sufficient to obtain good psychological responses and that higher levels of visual fidelity do not necessarily evoke more realistic psychological responses (Mania and Robinson 2004, Wilson and Soranzo 2015). Similarly, recent research has shown that many aspects of shopper behaviour do not differ meaningfully across retail formats (Sorensen, Bogomolova et al. 2017). Researchers intent on maximising the validity of VR shopping data, are likely to be better served by getting participants into the right frame of mind, engaging their non-visual senses and screening out participants with a low capacity for 'presence' (Slater 1999, Tyndiuk, Lespinet-Najib et al. 2004, Wilson and Soranzo 2015, Schnak, Holdershaw et al. 2016), than investing their efforts in recreating all the visual elements of individual stores.

Designing more simple VR environments may also help to reduce biases for visually salient packs. Researchers have hypothesised that participants tend to pay attention to the aspects of a research task they perceive matter to researchers (van Herpen, van den Broek et al. 2016). Participants may presume it is the visual elements of the environment that matter in higher quality simulations and select more visually engaging packs. Studies that compared basic and sophisticated VR simulations (e.g. Burke, Harlam et al. 1992, van Herpen, van den Broek et al. 2016) did not investigate this possibility.

However, tailoring VR environments to evoke specific retail settings may still be necessary for studies that aiming to convince retailers to trial or implement changes in real stores, retailers may have strong beliefs about the uniqueness of shopper behaviour in their stores. Additionally, a natural view of the shelf is likely to be important for shelf layout studies.

All of the VR simulations validated in this thesis showed the category in the context of the aisle and commenced with shoppers seeing products from an entrance to the aisle (see for example Figure 1). Some studies also controlled which end of the aisle shoppers started from. While VR studies where shoppers started in the middle of the aisle were not examined, I concur with other researchers that an end of aisle approach is likely to be necessary to avoid biases for the centre of the assortment/aisle (see for example van Herpen, van den Broek et al. 2016). Past studies that

[19] It was not possible to determine if the retail channel simulated contributed to the quantity of bias in multivariable analyses because of collinearity in the small number of data sets analysed.

provided a front-on image of a shelf without the context of an aisle have tended to bias towards products in the middle (e.g. Chandon, Hutchinson et al. 2009, Kim, Nonnemaker et al. 2014).

Implication 7: VR data can be collected online

A key implication of this thesis is that purchase choice and shelf layout studies can be collected through online panels, without negatively affecting the generalisability of the data. Online fielding will substantially increase the time, geographic and cost-advantages of VR over other methods of shopper research such as mock stores, and enable VR studies to be conducted with larger samples and shorter fielding times than was possible with in person fielding. It is also likely to increase the number of research agencies offering VR simulations, a trend that should benefit users of VR.

However, online fielding is unlikely to make VR studies a cheap or quick research option, unless other innovations for the design of VR studies are also adopted. VR environments can for example, be created easily from images or videos of real environments using low-cost software such as Panorama360. This would enable VR projects to be undertaken where consumers view and navigate a 3D VR environment created from photographs of a real environment and indicate in a survey which products they noticed or would purchase, rather than clicking on products. This approach would provide the level of rendering and spatial control that studies have found contribute to VR environments feeling real to participants i.e. a strong sense of presence (Li, Daugherty et al. 2003, Berneburg 2007, Kleinschmit, Reid et al. 2012, Stone and Cogdon, 2013, Roupé, Bosch-Sijtsema et al. 2014, van Herpen, van den Broek et al. 2016) but reduce the quantity of coding required to design VR studies. This approach may be a good option for research aimed at understanding how consumers navigate product assortments and select products. If this type of research is conducted it would be interesting to test if familiarity with the real environment used effects participants sense of presence or behaviour in the VR simulation. Research into HCI has found that prior exposure to the real environment that is simulated, improves participants' spatial understanding of the simulation (Interrante, Ries et al. 2008). Investigating this question may assist with improving the recruitment criteria for VR studies (Implication 8).

To test initiatives that have not been implemented in real stores, VR simulations likely will still need to be created via computer modelling. While the complexity and costs of doing so have reduced substantially since the 1990's, in part due to low-cost and open source software created for 3D modelling (Needel 2013, Carron 2017), large investments of time and/or resources are still likely to be required to create VR environments that are detailed and realistic-looking using computer modelling (Tan, Gergle et al. 2004). This will remain necessary for testing initiatives for which real-world analogues do not exist, unless technologies that enable VR models to be combined with images or videos of real environments are utilised, and will likely mean that VR studies will be more costly than online surveys or focus groups for testing new initiatives, even if VR studies are fielded online.

In-person fielding may still be necessary in some markets as substantial bandwidth is required for VR simulations to run well online, and for studies that require accurate gaze-tracking data to be collected data. Visual attention data can be collected using webcams built into consumers' tablets and computers, and some researchers were beginning to experiment with this approach when this research commenced. However, webcam eye-tracking is unlikely to ever be accurate enough to determine exactly what products and elements shoppers were looking at (Sewell and Komogortsev 2010, Tobii 2012), so may only be suitable for preliminary tests of shelf signage, in-store advertisements and pack designs.

Implication 8: More stringent recruitment criteria may be worthwhile

It may be possible to improve the consistency of shoppers' real and VR purchase choices by only recruiting shoppers who are responsible for the majority of their households' shopping and who purchase the category of interest very frequently (i.e. weekly), as analyses found that these shoppers are the most likely to select products from their self-reported purchase repertoires in VR studies.

However, before implementing this recommendation it would be prudent to confirm that the purchase choices of these shoppers have more external validity by designing an experiment where VR purchase data can be compared to purchase panel, loyalty card or receipt data. These are more reliable sources of data than consumer self-reports (Sudman and Bradburn, 1974, Tourangeau, Rips et al. 2000, Ludwichowska, Romaniuk et al. 2017). Additionally, research should confirm how much benefit can be obtained from this recommendation and to what extent it will affect the generalisability of VR data. As already mentioned in regards to pack design research, light and infrequent buyers make up a large proportion of sales for most CPG brands, so excluding these shoppers may diminish the generalisability of VR data.

Irrespective of recruitment criteria, future studies should ask participants what brands, products and pack-sizes they typically purchase, or collect this data through other means, so that the validity of shoppers' purchase choices can be better ascertained. Analyses demonstrated that consumers' own judgments of the realism of their purchase choices cannot be trusted, but information about the external validity of shoppers' purchase choices is needed if marketers are to trust VR for critical decision-making.

Limitations & Recommendations for future research

Area 1: Observational and experimental research to confirm the causes of the biases evident

It was hypothesised that biases were evident in results because participants failed to eliminate products they noticed but that were not suited to their tastes, budgets or purchase needs, skipping a stage of the decision making process (Russo and Leclerc 1994, Pieters and Warlop 1997, Van Der Lans, Pieters et al. 2008). While gaze tracking data was used to confirm that very few consumers considered price or size information in the VR simulations analysed, further analysis to confirm that the elimination phase was no present, was not possible. This was a limitation using data from VR studies conducted by other research agencies.

To confirm whether the elimination phase is skipped in VR experiments that do not have real purchase consequences, future studies should examine gaze tracking data from VR studies. Coding schema that have been used to classify gaze tracking data from real stores into decision making stages (see for example Russo and Leclerc 1994), should be applied to VR data, and results compared. To ensure that the decision stages identified in real stores reflect consumers thought processes in VR studies, it is recommended that qualitative research be conducted prior to the coding of data. This could be conducted by having participants view the gaze tracking video from their VR shopping trip and explain or narrate their thought processes (Smolentzov, Olsen et al. 2010, Gidlöf, Holmberg et al. 2012).

Ideally this research should test whether the hypothetical nature of VR purchase experiments, is responsible for the abbreviated decision making processes of consumers. To this end, a variety of 'incentive compatible' approaches should be tested, including:

- Telling participants at the start of the experiment they will receive the products they select in the VR simulation, but then providing them with cash incentive of equal or greater value as a substitute. This may be sufficient to resolve the biases observed and would avoid the logistical challenges of providing shoppers with actual products.
- Providing participants with a small cash incentive (e.g. $10) but requiring them to pay for, and then supplying, the products they select.
- Having participants hold onto money from their own wallet while making VR purchase selections or hand money to a researcher, but not actually requiring them to pay for the products they choose.

This research should be a priority as analyses in Chapter 8 indicated that the relative prices of products had the largest effect on the biases evident in VR data when the effects of unstandardised variables were compared.

Area 2: Systematic experimentation will be essential to resolving the observed biases and perfective VR research

More research is needed within the paradigm of VR research to resolve the biases identified in VR data. Ideally this research should focus on the protocols, rather than the technology used in VR studies as was found to be an under researched area (see Finding 1, page 183).

This research should be systematic, testing the impact of each change thoroughly and independently, and should benchmark new to established approaches. In addition to the methodological experiments already mentioned, approaches particularly worthy of testing are:

- Running VR experiments that systematically manipulate factors that have been shown to affect purchase behaviours in real stores including fatigue, time pressure and category involvement (Allsopp 2005, Hui, Fader et al. 2009, Reutskaja, Nagel et al. 2011, Yao and Oppewal 2016), to determine if they affect VR purchases in similar ways.
- Once it is better understood how shoppers interpret the instruction to "shop as you normally would", the most commonly used instruction in industry studies, studies should experiment with more specific prompts designed to encourage participants to behave realistically. This may include instructing participants to: select the products they intend to purchase on their next shopping trip,[20] to repeat their most recent shopping trip or select the brand, product varieties and or pack sizes as they usually buy. To determine if these instructions improve the realism of participants' purchase choices, ideally individual level purchase records e.g. loyalty card or purchase diary, should be examined and comparisons should account for normal variations in the buying behaviours of individuals across occasions due to the existence of purchase repertoires (Banelis, Riebe et al. 2013, Sharp and Romaniuk 2016). Testing instructions that provide a specific task (i.e. repeat your most recent shopping trip) is recommended as prior studies have found that shoppers are more likely notice pricing information and consider a variety of products, behaviours that are likely to improve the realism of VR shopping, when assigned a specific purchase task in a VR study (Burke and Leykin 2014).
- Determining if changes to the physical environment in which a VR study is conducted, for example in lab environment or shopping mall location, affects the realism of VR data. Psychology studies have found that changes to the physical environment in which a VR study is experienced (e.g. the lighting, temperature or sound) can substantially affect behaviour (Wilson and Soranzo 2015). Additionally, simple differences like having participants shop a VR environment standing up rather than sitting down or relevant scents (e.g. the freshly baked bread for a grocery store study, perfume for a department store study), could also be tested to determine if they can improve the external validity of VR data.

[20] Participants could be recruited in advance and told to hold their purchases to ensure they have actual purchase needs at the time of the VR study.

This type of systematic methodological research will require resources to be devoted to testing but could be tackled through an industry wide research effort or series of industry/academic partnerships to test the benefits of specific changes.

Area 3: Additional analysis to establish trustworthy calibration

Additional research will be needed to establish calibration models that can reliably be used to improve the predictive ability of VR data. This should be a priority for research as, in addition to being of practical usefulness to researchers, it may assist with identifying other biases in VR data that can be addressed through improved research design. Furthermore, accurate calibration models have been essential to establishing other types of shopper research including Simulated Test Markets and purchase likelihood questionnaires (Morwitz and Schmittlein 1992, Fitzsimmons and Morwitz 1996, Morwitz 2001, Morwitz, Steckel et al. 2007) as trustworthy and predictive methods of research.

Future research should:

- Identify additional variables that improve the accuracy of VR calibration. This may include factors mentioned in Chapter 8 and 9, such as unit price and category purchase frequency, or variables identified in prior research such as shelf location and promotional displays (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999, van Herpen, van den Broek et al. 2016). The aim should be to develop calibration models that have higher fit statistics and smaller errors that the models presented in this thesis, but are simple enough they are of practical usefulness and generalisable to other VR data. This may require experimentation with other model forms and terms as this was not a focus of this thesis.
- Refine the color and pack size variables that were identified as potentially significant predictors of VR bias. Research should, for example, determine if coding the visible surface area of packs is a better predictor of biases across a range of product categories than the pack size/weight variable coded.
- Confirm that the variables identified in this thesis are significant predictors of bias in VR data for individual products (i.e. SKUs). This should be a priority as marketing managers are often interested in predicting accurately the sales of individual products in their range (Higaki 2004, Precourt 2012, Needel 2013). Determining if the same variables can be used to calibrate other marketing metrics such as product penetration, segment penetration and duplication of purchase, would also be of value (Ehrenberg 2000, Treiber and Needel 2000) as in markets where panel data is unreliable or does not exist, VR would be a useful tool for determining the structure of markets if it can be validated. Similarly, VR would also be a useful tool for obtaining forecasts for segments and undertaking tests of shelf initiatives and cross purchasing, if results can be accurately calibrated.
- Test the generalisability of the variables identified in this thesis to grocery and non-grocery product data. Ideally this research should seek to identify the boundary conditions of the variables identified (Ehrenberg 1995, Uncles and Wright 2004), confirm models and model

parameters were not overfitted and establish what adjustments or transformations are needed for models to apply to a range of products and VR data sets.

- Investigate if the same beta weights can be used to calibrate data from different product categories, countries and retail environments, or if tailored estimates are needed. As cross-country studies have found that US consumers are more prone to overestimating their 'willingness to pay' than consumers in many other markets (Ehmke, Lusk et al. 2008), beta weights for pricing variables are likely to require adjustment. Similarly, in some countries, brand type (PL or NB) may be a less important or insignificant significant predictor of VR bias, as consumers are less willing to pay price premiums for national brands (Steenkamp, van Heerede et al. 2010) than the consumers examined in this thesis.

It may be possible, at least initially, to undertake this research by utilising data from VR studies already conducted by industry researchers and syndicated sales data, the approach used in this thesis. However, it would also be of value to compare the real and VR purchase choices of individuals across multiple occasions, this would provide a more precise estimate of the biases in consumers' purchase choices in VR simulations.

Area 4: Confirming shelf layout findings with larger samples

It is recommended that more research be undertaken to confirm that VR can predict the impact of shelf layout changes in real stores as this is a key area of concern for industry users. While this was investigated in this thesis, only three of the many changes that could be implemented in real stores were tested and the samples were not sufficiently large to validate that shelf layout changes have similar effects on the shares of small brands as in real stores. It is recommended that future research be conducted with at least 1000 purchases per layout to ensure a good level of statistical power (e.g. 80% power, $\alpha < .05$).

Area 5: Other VR approaches will require research and validation

This thesis compared two methods of fielding VR studies: (1) VR displayed simulations on computer screens in specialised research facilities and (2) VR simulations displayed on consumers' personal computers. These were the approaches most commonly used VR research approaches in Europe, the US and Australia at the time this research commenced. However, a range of other technologies are also utilised including multi-projector "CAVE" systems in specialised research facilities, tablet computers, QR codes that augment real environments and head mounted displays (i.e. VR goggles).

Research has already established that head mounted displays and multi-screen viewing systems, approaches that immerse shoppers in the VR environment, provide shelf layout results consistent with this thesis (Porcheddu and Venturi 2011, van Herpen, van den Broek et al. 2016). However, more research is needed to establish the advantages of different VR approaches for other types of shopper research. For example, available motion-sensor tracking devices (e.g. Virtuix Omni) that enable participants to walk through VR environments and use hand movements to select products

(see for example Figure 26), may be advantageous for research aimed at understanding category buying and browsing behaviours (Roupé, Bosch-Sijtsema et al. 2014).

Ideally research utilising different technologies should be conducted so as to determine the impact of each component of the technology (e.g. display type, interaction device, interaction software), rather than comparing VR systems that differ in a number of these aspects. This will enable researchers to make informed decisions about the costs and benefits of different technology and progress to be made in understanding what elements effect results. Furthermore, testing of new technologies should be strategic, aiming to rectify the biases identified in VR data and incorporate learnings from other fields of research to identify promising technological innovations for testing.

Additionally, research should examine the benefits of different technologies for specific types of shopper research. For example, the size of the VR displays may be important for the accuracy of category placement research, but may not be important for pack design, product choice or pricing research.

Figure 26: Example of a VR rig in which participants walk to move forward



This type of VR rig would enable participants to engage more naturally with a VR environment, for example by walking to move closer to a shelf.

Further research will also be needed to determine if high rates of simulation sickness (i.e. participants feeling motion sick in VR) can be resolved. Industry experts have predicted that HMD displays will become less 'jerky' over time but that high simulation sickness rates will continue to affect environments with tall shelves, narrow aisles and large colour palettes found in the typical grocery store (Sparto, Whitney et al. 2004, Wilson and Soranzo 2015).

Area 6: More validated research is needed to establish VR as a trusted method

The body of evidence for VR is very slim, particularly in comparison to more established methods of shopper research. Marketers and industry researchers should play a key role in building the evidence base for VR by documenting results and the outcomes of initiatives tested in VR that are implemented in real stores (e.g. did the layout selected perform well in real stores). This will assist in showing that VR simulations *can* provide generalisable and predictively accurate results.

Chapter Summary

This chapter discussed the key findings of this thesis and their contributions to knowledge, including that:

- VR shopping simulations cannot estimate accurately the real repertoires or purchase choices of consumers as a large proportion of consumers select products they do not buy in real stores in VR simulations that look realistic but do not have real financial consequences.
- Without calibration, VR shopping data is unlikely to provide accurate market share estimates, particularly for premium grocery products, private label products or products that stand out on shelf. Consequently, VR is not recommended for sales forecasting until accurate calibration models that account for these and other variables are established.
- Until research protocols to resolve the biases evident are tested, VR simulations are not recommended for assortment change or observational purchase research. Additionally, care is recommended if using VR for pack design research as the sales benefits of having a pack design that stands out on shelf because of its colours, visible surface area or design will likely be overestimated.
- Despite the biases evident, VR simulations may still be able to provide results that are indicative of what will happen in real stores when shelf layouts are changed if the tendency of participants to select more expensive products than they usually buy is controlled. Studies should however be designed with larger samples and/or fewer test layouts.
- Whether VR studies are fielded in person in specialised research facilities with multi-screen displays or online with panel members' own computers, does not affect the accuracy of VR purchase data and VR studies do not need to be fielded with multiple 'practice' trips. This will assist with making future studies less burdensome to conduct.
- The biases evident in VR purchase data could be explained by the failure of consumers to eliminate products they notice based on price, taste and other real-world concerns. More research is recommended to confirm and resolve this bias.

This thesis provides a number of specific recommendations for research that are informed by interviews with industry practitioners and experts, the effect sizes evident in VR studies analysed and research in fields more advanced in their use of VR simulations, such as psychology and human computer interaction.

APPENDIX 1
COMPANIES USING VR FOR RESEARCH

Based on information reported publicly in and in-depth interviews. Companies marked with an asterisk have a Customer Collaboration Centre with a large screen for displaying virtual environments. This represents a significant investment in the technology although does not necessary correlate with its use for research.

Company name	**Product categories**
Anheuser-Busch InBev	Alcoholic beverages (beer)
Carlton United Breweries	Alcoholic beverages (beer, cider, spirits)
Carlsberg	Alcoholic beverages (beer)
Diageo*	Alcoholic beverages (beer, wine, spirits)
Heineken	Alcoholic beverages (beer)
Molson Coors	Alcoholic beverages (beer)
Energizer Holdings	Batteries
Coty	Beauty products & perfume
Coca-Cola	Beverages (soda, water, juices etc.)
Heinz	Food (soups, packaged vegetables, sauces & dressings, ready meals, desserts, baby food)
Kraft-Mondelez*	Confectionery (chocolate, gum, candy), food (biscuits, chips, crackers, dairy spreads etc.), beverages (cordial, coffee), pharmaceuticals (cough drops)
Mars-Wrigleys*	Confectionery, food (pasta sauce, rice, spices etc.) & pet care
Nestle*	Confectionery, food (cereal, yogurt, ice-cream), beverages (water, coffee, juice), pet care & pharmaceuticals (cough drops)
Dole Fresh Vegetables	Food (fresh & frozen fruit & vegetables)
Crystal Farms Cheese	Food (dairy products, muffins etc.)
ConAgra Foods	Food (popcorn, cream, pasta, flours, ready meals and side dishes, meats)
General Mills*	Food (cereal, pasta sauce, yogurt, ice-cream, frozen vegetables, dough)
Kellogg's	Food (cereal)
PepsiCo	Food (cereals, crisps) & beverages (soda, juice, electrolyte etc.)
Unilever*	Food (ice-cream, spreads, soups & sauces) & beverages (tea), home care (detergents, disinfectant), personal care (hair care, soap etc)
Sara Lee	Food (desserts, packaged meats, bakery) & beverages (coffee)
Colgate-Palmolive	Personal care (toothpaste, hair care, soap etc.)
Clorox	Home care (plastic wrap, laundry, cloths, bleach) & food

Company name	Product categories
	(dressings, seasonings), personal care
Sun Products	Home care (laundry powder, dish soap etc.)
Kimberly-Clark	Home care (paper towels, toilet paper) & personal care (nappys, feminine hygiene)
Glaxo-Smith-Kline	Personal care (toothpaste), beverages (juice), pharmaceuticals
Henkel-Schwarzkopf	Personal care (hair care, skin), home care (detergents, dish soap), adhesives
Johnson & Johnson	Personal care (soap, band aids, oral care, feminine hygiene) & pharmaceuticals
Procter & Gamble*	Personal care (toothpaste, hair care, soap, beauty products, feminine hygiene), home care (detergents, cleaning products), pet care, grocery (Pringles) paper products, pharmaceuticals & batteries
Reckitt Benckiser	Personal care & home care (condoms, dishwasher tablets etc.)
British American Tobacco	Cigarettes
Japan Tobacco International	Cigarettes
Staples	Stationery & office products

APPENDIX 2

IN-DEPTH INTERVIEW MATERIALS

Recruitment material



ANNUAL REPORT 2012

University of South Australia | Ehrenberg-Bass Institute for Marketing Science

VIRTUAL REALITY - NEW RESEARCH FRONTIER

Shopping research: what's the reality when it comes to virtual reality?

Big brands have been using virtual reality technologies since the 1990s to visualise retail environments, foster retailer engagements and undertake shopper research.

While there are various tools on the market, typically virtual reality shopper solutions combine high-resolution images of retail environments with research and gaming technologies to gather insights about how consumers navigate stores and how they respond to in-store initiatives in terms of attention and brand choice.

However, even if virtual environments realistically portray retail environments and consumers are recruited to shop 'normally', virtual shopping behaviours may not reflect typical in-market behaviours.

As a result the Institute started a research initiative to validate if virtual reality is a tool marketers can trust for shopper research.

Our pilot results suggest that metrics of sales uplift, shopper engagement, time-in-aisle and attention may not be reliably correlated with in-market behaviours. We need to understand the conditions when we can and when we should not use such approaches.

For stage two of our project we invite a select group of Corporate Sponsors (experienced with virtual reality or willing to invest in it) to join us to continue to validate virtual shopping approaches across a range of categories, countries, approaches and shopping situations. We aim to determine:

- how accurate results are compared to the real world (in what conditions);
- if and which metrics can be calibrated to in-store behaviours (and how); and
- standards to increase the actionability of results and deliver cost-savings.

For more information or to express your interest in being included in this research initiative contact
Katherine Anderson or Assoc. Prof Rachel Kennedy
Virtual@MarketingScience.info



University of South Australia | Ehrenberg-Bass Institute for Marketing Science



Progressing Virtual Reality

Ehrenberg-Bass Institute Research Initiative

Virtual reality has the potential to advance marketing knowledge and transform shopper research. Virtual reality combines high-resolution images of retail environments and computer gaming technology to deliver shoppable environments for retail visualisation and consumer research. This is an exciting and technologically advanced approach, but for results to be trustworthy, validation and benchmarks for calibration are needed.

Both the Advertising Research Foundation and the Path to Purchase Institute have called for more 'research on virtual research'.

Preliminary research with a select group of industry studies has found that measures of sales lift, shopper engagement, time-in-aisle and attention have varying levels of consistency with real measures, even when market share estimates are highly correlated to real benchmarks. More research is needed to determine if other measures of shopper behaviour are also biased, conditions when biases occur and if results can be calibrated to fully harness the potential of virtual reality for marketing research.

As an independent research institute, the Ehrenberg-Bass Institute for Marketing Science will work with users, research agencies and virtual reality vendors to gather empirical evidence to validate virtual reality and establish best-practice methods and benchmarks.

Project Summary

A research initiative to advance industry knowledge and provide certainty to marketers.

Key Outcomes

- Validation of shopper behaviour metrics.
- Recommended standards for virtual projects to increase the actionability of results and deliver cost-savings.
- Benchmarks for determining the real-world accuracy of results and calibrating metrics effectively.

Register your interest in this industry initiative by contacting:

Virtual@MarketingScience.info

Founding members will have priority access to project outcomes.

Project Leaders

The Ehrenberg-Bass Institute's Associate Professor Rachel Kennedy and Katherine Anderson will coordinate the academic research. James Sorensen from TNS will coordinate efforts from a research industry perspective.

University of South Australia | Ehrenberg-Bass Institute for Marketing Science



Research Agenda

- Which aspects of shopper behaviour are captured accurately in virtual reality simulations?
- What behavioural metrics are consistently biased? Can results be calibrated to reflect in-market impact?
- What level of virtualisation should be employed? What is the cost / benefit of online virtual environments? Does rendering quality improve the accuracy of results?
- How can simulations or studies be enhanced to produce more realistic and reliable metrics?
- Develop and validate industry standards to make virtual studies easier, faster and more cost effective to implement.
- Determine empirically when virtual approaches can be used reliably for research.

Opportunities to Participate

We welcome involvement in this research project in a number of ways:

- Contribute data from your past (or current) virtual reality studies for independent validation and benchmarking against real shopper behaviour data.
- Conduct a study to compare virtual data to in-store data or shopper research data.
- Conduct a study to compare data from sophisticated and basic virtual environments. This may involve testing the impact of rendering quality.
- Contribute results from comparisons of virtual data and post-implementation results.
- Contribute knowledge or ideas. We are aiming to build a network of interested researchers and marketers.

For more information or to register your interest contact:

Katherine Anderson on +61 8 8302 7249 or Virtual@MarketingScience.info

About the Ehrenberg-Bass Institute

The Ehrenberg-Bass Institute, based at the University of South Australia delivers world-class research, scientific knowledge and dramatic discoveries to corporations all over the world.

For over a decade, some of the world's leading marketers like Unilever, Mars, Coca-Cola, and Procter & Gamble have been funding our research program to answer fundamental questions about buyer behaviour and how brands grow.

University of South Australia | Ehrenberg-Bass Institute for Marketing Science

Participant information sheet



Progressing Virtual Reality

Ehrenberg-Bass Institute Research Initiative

Virtual reality has the potential to advance marketing knowledge and transform shopper research. Virtual reality combines high-resolution images of retail environments and computer gaming technology to deliver shoppable environments for retail visualisation and consumer research. This is an exciting and technologically advanced approach, but for results to be trustworthy, validation and benchmarks for calibration are needed.

Both the Advertising Research Foundation and the Path to Purchase Institute have called for more 'research on virtual research'.

Key Researcher: Katherine Anderson, M.Bus
Katherine.Anderson@MarketingScience.info
+61 8 8302 7249
Supervisors: Assoc. Prof. Rachel Kennedy & Dr Anne Sharp
Rachel.Kennedy@MarketingScience.info +61 8 8302 0290
Anne.Sharp@MarketingScience.info +61 8 8302 0637

Preliminary research with a select group of industry studies has found that measures of sales lift, shopper engagement, time-in-aisle and attention have varying levels of consistency, even when market share estimates are highly correlated to real benchmarks. More research is needed to determine if other measures of shopper behaviour are also biased, conditions when biases occur and if results can be calibrated to fully harness the potential of virtual reality for marketing research.

As an independent research institute, the Ehrenberg-Bass Institute for Marketing Science will work with users, research agencies and virtual reality vendors to gather empirical evidence to validate virtual reality and establish best-practice methods and benchmarks.

The Ehrenberg-Bass Institute, based at the University of South Australia delivers world-class research, scientific knowledge and dramatic discoveries to corporations all over the world.

Participant Information

You can contribute knowledge to this research initiative by participating in an informal 30 minute interview, either over the phone or in-person. To register your interest contact Katherine Anderson on +61 8 8302 7249 or Virtual@MarketingScience.info

Participation is voluntary and the confidentiality of participants will be protected at all stages of the project. Audio recordings will be made with the permission of participants but no information which could lead to the identification of any individual will be released without their express consent. Participants are free to withdraw from the project at any stage.

Participants will be provided with a summary of findings via email in early 2014.

This project has been approved by the University of South Australia's Human Research Ethics Committee. If you have any ethical concerns about the project or questions about your rights as a participant contact the Committee Executive Officer, vicki.allen@unisa.edu.au +61 8 8302 3118.

Interview discussion guide

- General introduction to clarify role/position, region, category responsible for
- When did you first start using virtual reality for research?
- What types of organisations are using virtual reality approaches? For what categories?
- What types of projects is virtual reality used for?
- What types of store environments (supermarket, convenience, superstore etc) are being modeled?
- What are the key advantages of using virtual reality approaches? (perceived or actual)
- Are results better for some product categories or products?
- How does virtual reality compare to other methods in terms of its insight, timeliness and cost?
- Does it offer advantages over other methods of research? (actual or perceived)
- What data/metrics are typically collected in a virtual shopping project?
- How is the data from virtual shopping projects reported & used in decision making?
- Which metrics are perceived to be the most important?
- How valid is the data perceived to be?
- Can virtual shopping simulations predict in-market outcomes?
- To what extend does virtual shopping resemble real shopping?
- Do you think subjects shop normally and respond to marketing changes like they would in the real world?
- How valid do you think the data is? Do you have concerns about the data coming from virtual shopping simulations?
- Do results validate against real shopping benchmarks or in-store research?
- How are virtual projects conducted?
- How are virtual shopping environments built? Are whole stores modeled virtually or just single aisles?
- What types of research facilities are used?
- What level of virtual reality is used?
- How are online and central location virtual approaches used differently or for different studies/ categories?
- How are subjects typically recruited?
- How are subjects introduced to the task and the virtual shopping environment?
- What types of controls are put in place to normalize the behaviour of respondents?
- Are there some types of research that virtual reality is not suitable for?
- What are the limits of the method?
- How do you think virtual reality approaches will be used in the future?

Characteristics of industry participants

Table 51: Group 2 Market researchers (n = 16)

Job Title	Years using VR
Managing Partner	20
Founder/ CEO	17
Chief Research Officer	8
Chief Research Officer	7
Knowledge Leader & Head of VR Research	6
Consultant for Advanced Analytics	6
Director of Product Development	5
Sr VP Consumer, Retail & Shopper Research	4
VP, Shopper Analytics	4
Project Manager	4
CEO, Retail Research	3
VP, Consumer & Market Insights	3
Account Director	3
Sr. VP, Retail & Shopper	3
Director, Product Management & Development	2
Director, Global Retail & Shopper Insights	1
Average	6

Table 52: Group 3 Industry experts (n = 3)

Job title	Years using VR
Principal consultant	10
Global Consulting Director Retail and Shopper	10
Research professor: School of Marketing	4
Average	8

Table 53: Group 1 Marketing managers (n = 21)

Job title	Years using VR
Global Director Shopper Insights	22
Sr. Category Manager	7
Category Management Leader, VR specialisation	6
Senior manager, Store usability	6
Director, Shopper Insights	5
Director Consumer & Market Insights	5
Sr. Manager of Virtual Reality Practice	4
Head of Shopper Insights Asia-Pac.	4
Project Manager, VR Specialisation	3
Sr. Consumer Insights Manager	3
Director of Customer Marketing & Category Management	3
Advanced Analytics & Market Research	2
Senior Insights Planner	2
Consultant, Consumer Insight & Strategy	2
Manager Consumer Insight & Strategy	2
Practices Innovation Manager	2
Category Manager	2
Sr Director Consumer & Market Insights	2
Manager Virtual Store Development	2
Insights Manager	2
Manager of Collaboration, Learning & Insights, Europe	1
Average	4

APPENDIX *3*

STRUCTURE OF VR DATA & METRICS AVAILABLE

One of the challenges of VR data is the quantity of data recorded. Detailed data about the behaviours and position of each participant is recorded on a second-by-second basis, in addition to pre-and post shopping survey data.

This data is typically stored in custom databases or multiple data sheets that are linked by different codes i.e. StudyTicket ID, Product ID, Respondent ID.

Many hours of work were needed to prepare and re-join these data sheets.

Figure 27: Structure of VR data

Sheet 1: Session data

Cell	StudyTickID	RespondentID
1	268519	229487

Sheet 2: Product examination data

RespondentID	SurveyID	Sex	Age	etc.
229487	268519	2	4	

(Continued on over page. Data for pre an post-shopping surveys was stored in separate data files as was information about which layout/shelf setup each respondent shopped.)

Sheet 3: Product examination data

ActivityID	StudyTickID	Exercise	ProductID	Action	View	Time	Order
671252	268519	Practice	3700010668	View	Front	43:28.7	1
671253	268519	Practice	3700010668	View	Right	43:55.2	2
671254	268519	Practice	3700010668	View	Back	43:56.2	3
671255	268519	Practice	3700010668	View	Front	43:57.2	4
671256	268519	Practice	3700010668	View	Back	43:58.2	5
671257	268519	Practice	3700010668	View	Bottom	43:59.2	6
671268	268519	Practice	3700010668	View	Front	45:42.5	7
671269	268519	Practice	3700010668	View	Front	45:43.0	8
671270	268519	Practice	3700010668	View	Front	46:17.2	9
671295	268519	Shop	3700010673	View	Front	51:08.7	10
671315	268519	Find	3700010668	View	Front	56:34.3	11

Only data from 'shop' trips was analysed

Sheet 4: Product selection data

ActivityID	StudyTickID	Exercise	ProductID	Action	Time
1246420	268519	SHOP	3700010673	Pickup	51:08.7
1246421	268519	SHOP	3700010673	Add	51:11.3
1246422	268519	SHOP	3700010673	Remove	51:13.2

Sheet 5: Product position data

ProductID	Product Description	PositionX	PositionY	PositionZ	Price
3700010668	BNTY PPR TWL WHT RL SPR MG 2 PLY 112 CT 6 CT	0.4079	0.8156	-9.5516	12.54
3700010673	BNTY PPR TWL WHT RL 2 PLY 56 CT 8 CT	0.3012	0.8156	-9.3431	9.54

The information needed link product coordinates with their location on screen was not available. The position of products was coded manually from shelf layout plans.

In many instances companies did not save high resolution images of the VR environment used.

Table 54: List of metrics available from a typical VR study

	Time
1.	Total Mean Session Time
2.	Dwell time for various areas/products
	Attention
3.	% Noting At Least Once
4.	Total % Re-Noting
5.	Actual % Re-Noting
6.	% Noting Within 5 Seconds
7.	% Noting First
8.	Actual Mean Time to First Fixation –Product Hierarchy
9.	Gaze pick range analysis
10.	Cross-noting at least once
	Engagement
11.	Total Mean Accumulated Engagement Time
12.	% Share Mean Accumulated Engagement Time
13.	Total % Picking up at PH level
14.	% Removing product from basket at PH level
15.	Actual Mean duration of Pick Up at PH level
16.	Total % Examining Front (same as Total % Picking Up)
17.	Actual % Examining Front (will always be 100%)
18.	Actual % Examining Left
19.	Actual % Examining Right
20.	Actual % Examining Back
21.	Actual % Examining Top
22.	Actual % Examining Bottom
23.	Cross-examination (pick up)
	Conversion
24.	Conversion Rate: Notice to Pick Up
25.	Conversion Rate: Pick Up to Purchase
26.	Conversion Rate: Notice to Purchase
27.	Conversion Rate: Notice Signage then Related Product
28.	Conversion Rate: Notice Signage who Pick Up Related Product
29.	Conversion Rate: Notice Signage who Purchase Related Product
	Event Stream
30.	Total Mean Time to first POS Noted
31.	Total Mean Time to first product noted
32.	Total Mean Time to first product pick up
33.	Total Mean Time to first product placed in basket
34.	Total Mean Time to last product placed in basket
35.	Total Mean Time to last product noted
36.	Total Mean Time to last POS Noted
	Findability Exercise
37.	Actual Mean Time to Noting product

38. Actual Mean Time to picking up product
39. Total % noting product within **5** seconds
40. Total % picking up product
41. % picking up product as their first product
42. Mean number of products noted before target product noted
43. Mean number of products picked up before target product noted
44. Mean number of products picked up before target product picked up

Examination Task: Eye-tracking sample ONLY

45. % Noting element at Least Once
46. Actual Mean Time to first noting of element
47. Total % Noting element First
48. Total % Noting element Second
49. Total % Noting element Third
50. Total % Noting element Fourth
51. Total % Noting element Fifth

Purchasing Data-Total sample

52. Total % Purchase Penetration
53. Total Mean Basket Spend
54. Total Mean Basket Units
55. Total Mean Basket Weight/Volume
56. Actual Mean Basket Spend
57. Actual Mean Basket Units
58. Actual Mean Basket Weight/Volume
59. Share of Basket Spend
60. Share of Basket Units
61. Share of Basket Weight/Volume
62. Cross-purchasing

Position Data

% penetration by grid square
% penetration by grid square – 5 seconds
% penetration by grid square – 10 seconds
% penetration by grid square – 20 seconds
% penetration by grid square – 30 seconds

APPENDIX 4

DETAILS OF DATA SETS

Study 1: Confectionery & Gum

Study purpose	Test if displaying products upright in transparent boxes/capsules would increase sales and shopper satisfaction
Scope of simulation	330 chocolate, confectionery & gum products arranged in an aisle (approx. 9 meters); other product categories visible in the background
Product range	2-3 different sizes for each product; expensive and cheap products
Study design	Practice session followed by normal shopping trip in either the test or the control scenario; followed by alternate scenario. This was done to get the most out of the sample. Starting position (left vs right end of aisle) alternated across participants.
Sample criteria	Purchased confectionery (i.e. candy) from grocery store in the past 6 months; not asked about typical purchase quantity, spend or repertoire
Recruitment	Shopping malls in 7 US locations incl. Ca, Ga, NJ, Pa, Tx, Wi
Fielding	In person in a commercial research facility
Interface	Normal computer mouse to select products on 20 to 28 inch computer screen
Sample size	300 per scenario / 1500 purchases
Instructions Trip 1 (displayed on screen)	Please imagine you are going to the grocery store for a typical shopping trip. You have written candy on your shopping list and are planning on going to the candy section during your trip. As you shop the candy section, take note of how easy or difficult it is to identify and find the items you are looking for on the shelf. Once you have completed your shopping and checked out you will be asked a series of questions. Begin by clicking on the next button.
Instructions Trip 2 (displayed on screen)	You are now going on your second shopping trip. Please imagine this is a new day and a new trip to your regular store and you have decided you need to buy candy once again. As you shop the candy section, take note of how easy or difficult it is to identify and find the items you are looking for on the shelf.

Study 2: Petcare

Study purpose	To test if organising category by need rather than brand, and signposting the position of certain products would increase sales and shoppablity (time to find)
Scope of simulation	184 canine snacks and treats arranged as though in an aisle (approx. 8 meters); other categories visible in the background
Product range	2-3 different sizes for each product; expensive and cheap products
Study design	Practice session, normal shopping trip then 'time to find' exercise for specified product. Starting position (north vs south end of aisle) alternated but not recorded Eye-tracking data collected
Sample criteria	Over 18 years of age; main or joint household shopper; purchased dog treats in the past 3 months; shop at Walmart frequently
Recruitment	Shopping malls in 8 US locations incl. Az, Fl, Ga, Il, Mo, Pa, Tx;
Fielding mode	In person in a commercial research facility
Interface	Normal computer mouse to select products on 24 inch computer screen with an aspect ratio of 16:10
Sample size	50 for control group; 300 total; 131 purchases
Eye-tracking (calibration)	Shoppers' eye movements were recorded using a desk-mounted Tobii eye-tracking system. This recorded the position of the fovea 50 times per second using a small, infrared camera under the screen on which simulations where displayed. To calibrate the system, shoppers were seated 65 centimeters away from the screen, adjusted so their eyes were level with the centre of the screen and asked to look at the cross on the screen. This cross moved to five points on the screen. Shoppers could use this system whilst wearing glasses. The system recorded eye fixations (movements of at least 50 milliseconds on a point of interest) and saccades (movements between points of interest) for the duration of their shopping trip.
Instructions (displayed on screen)	Please imagine that you are going to your local Walmart for a typical trip to purchase treats for your dog. Please shop the [section] as if you were using your own money, buying as much or as little as you normally would. Please spend as much or as little time as you typically would when shopping this section. Once you have completed your shopping and have checked out, you will be asked a series of questions.

Study 3: Feminine care products

Study purpose	Understand how shoppers navigate to the category and select products.
Scope of simulation	200 feminine hygiene products arranged in an aisle approx. 9 meters long; different layouts and shelf fixtures for Walmart and CVS setups
Product range	2-3 different sizes for each product; expensive and cheap products
Sample screeners	Female; aged 21-54; purchased feminine care products from a drug or Walmart store in the past 3 months
Recruitment	Shopping malls in 6 US locations incl. Az, Fl, Ga, Il, Mo, Pa, Tx
Fielding mode	In person in a commercial research facility
Interface	Joystick modelled on a shopping trolley, set below two 24 inch computer screens arranged to provide a 140° field of view
Sample size	300 consumers
Instructions for trip validated	Please imagine that you are going to your local [store name] for a typical trip and feminine hygiene products (pads, tampons or wipes) are on your list. Please shop the [section] as if you were using your own money, buying as much or as little as you normally would. Please spend as much or as little time as you typically would when shopping this section. Once you have completed your shopping and have checked out, you will be asked a series of questions.

Study 4: Disposable nappies

Study purpose	Understand how shoppers navigate to the category and select products.
Scope of simulation	115 products arranged in as though in the aisle approx. 9 meters long; different layouts and shelf fixtures for Walmart and Kroger setups
Product range	3 different sizes for each product in Kroger, 2 in Walmart; expensive and cheap products
Sample criteria	Purchased nappies and/or training pants in the last 3 months at Walmart or grocery store (any brand); responsible for the majority of household's grocery shopping; asked about typical purchase quantity, pack size and brand repertoire
Recruitment	Shopping malls in 12 US locations incl. Fl, Ga, Ma, NJ, Oh, Pa, Tx, Wa
Fielding mode	In person; commercial research facility on 24 inch computer screen with an aspect ratio of 16:10
Fielding mode	In person in commercial research facility
Interface	Joystick modelled on a shopping trolley, set below two 24 inch computer screens arranged to provide a 140° field of view
Sample size	400 consumers
Instructions for trip validated	Please imagine that you are going to your local [store name] to purchase disposable nappies. Please shop the [section] as if you were using your own money, buying as much or as little as you normally would. Please spend as much or as little time as you typically would when shopping this section. Once you have completed your shopping and have checked out, you will be asked a series of questions.

Study 5: Paper towels

Study purpose	Understand how shoppers navigate to the category and select products.
Scope of simulation	27-45 different products arranged in as though in an aisle approx. 11 meters long; different layouts and fixtures for Walmart and Kroger setups
Product range	2 different sizes for each product in Kroger, 3 in Walmart; expensive and cheap products
Study design	Practice session followed by normal shopping trip in either the Walmart or the grocery setup; followed by time-to-find exercise. Eye-tracking data collected
Sample screeners	Responsible for the majority of household's grocery shopping; have purchased paper towels in the last month; use 2+ rolls of paper towel per month

Brand buy most often?	**Sales Data**	**Virtual Sample**
Expensive > $7.00 per pack	42%	38%
Mid-priced $5.50 to $7.00 per pack	35%	29%
Cheap < $5.50 per pack	30%	33%

Recruitment location	Shopping malls in 7 US locations
Fielding mode	In person in commercial research facility
Interface	Joystick modelled on a shopping trolley, set below two 24 inch computer screens arranged to provide a 140° field of view
Sample size	195 Walmart; 165 grocery
Instructions for trip validated	Assume for a moment that you have come to the store for paper towels. Please start here and go shop as you would normally shop if you were in a store shopping for paper towels. You may buy as many or as few products as you wish.

Study 6: Pantry staple

Study purpose	Test new shelf organisation schema
Scope of simulation	167 rice products arranged in an aisle (approx. 7 meters); other product categories visible in surrounds but not shoppable
Product range	2-4 different sizes for each product; range of cheap and expensive products
Study design	Shopping exercises in promotional area (3 alternative assortments) and main aisle (4 layouts). Starting position (left vs right end of aisle) alternated across alternative layouts. Groups for different layouts were comparable in terms of brand and product purchasing (claimed), height and product assortment
Sample size	80 per layout; 320 total
Fielding mode	In person in specialised research facility
Interface	Joystick modelled on a shopping trolley, set below two 24 inch computer screens arranged to provide a 140° field of view
Recruitment location	Shopping malls in 10 locations incl. Ca, Ga, NJ, Pa, Tx, Wi
Recruitment criteria (sample)	Purchased rice in the past 2 months; Regularly purchase groceries from grocery stores; Purchased one of 6 target brands in past 6 months; Spend at least $25 on a trip that includes rotisserie chicken

Brand(s) purchased in the past 6 months?	**Layout 1**	**Layout 2**	**Layout 3**
Uncle Ben's (Brand A)	63%	56%	54%
Rice-a-Roni	55%	62%	66%
Minute	50%	54%	47%
Zatarain's	47%	37%	45%
Knorr (Brand B)	28%	33%	24%
Goya	18%	23%	22%
Success	13%	18%	16%
Mahatma	10%	10%	19%
Private Label/Store Brand	9%	12%	13%
Near East	8%	7%	6%
Carolina	4%	2%	4%

Products purch'in the past 6 months?	**Layout 1**	**Layout 2**	**Layout 3**

	Long Cooking White Rice	59%	64%	60%
	Instant White Rice	57%	57%	58%
	Boil in Bag White Rice	43%	45%	46%
	Jasmine Rice	27%	34%	27%
	Basmati Rice	26%	26%	28%
	Long Cooking Brown Rice	22%	21%	22%
	Instant Brown Rice	19%	23%	16%
	Boil in Bag Brown Rice	19%	17%	10%
	Flavored Rice prepared on stove top	17%	17%	16%
	Ready To Heat Rice in pouch or tray	13%	8%	16%
	Av. height of participants	**Layout 1**	**Layout 2**	**Layout 3**
		5″ 5′	5″ 3′	5″ 6′
	Starting position	**Layout 1**	**Layout 2**	**Layout 3**
	Left-end of aisle	58%	60%	58%
	Right-end of aisle	42%	40%	42%
Instructions Trip 1 (Not analysed)	Imagine that today you are going shopping and rotisserie chicken is on your shopping list. Please purchase rotisserie chicken, and shop the produce/ deli section for other items to make a meal tonight. Shop as you might normally when purchasing a rotisserie chicken, as if you were using your own money, purchasing as much or as little in this part of the store as you normally would on a regular shopping trip that would include a rotisserie chicken.			
Instructions Trip 2 (Analysed)	You are now going to shop the rice section of the grocery store with the mindset that this is a new day and a new shopping trip. Please shop as you would normally, as if you were using your own money, buying as much or as little in the rice section of the store as you normally would on a regular shopping trip. Please spend as much or as little time as you typically would when shopping this aisle. There are other sections in the store, but you are to shop the rice section only for this shopping exercise today. Once you have completed your shopping and have checked out, you will be asked a series of questions.			

Study 7: Feminine care products

Study purpose	Determine if brand or segment blocking would increase sales for the category, brand and across segments purchasing; test if reducing the number of SKUs would impact category and brand sales
Scope of simulation	204-221 feminine care products arranged in an aisle (approx. 7 meters); other product categories visible in background but not shoppable; Walgreens store signage visible

Product range	2-3 different sizes for each product; range of cheap and expensive products
Study design	Practice session, normal shopping trip then 'time to find' exercise for specified product. Starting position (north vs south end of aisle) alternated but not recorded
Sample criteria	Female; aged 21-54; shop at Walgreens at least once a month; purchased feminine care products from a Pharmacy in the past 3 months
Recruitment	Commercial online panel; respondents from 50 US states
Fielding mode	via the internet on shoppers home computers; no download software to participate
Interface	Consumers' home computer mice/trackpads and screens
Sample size	300 per assortment; 1200 total
Instructions for trip validated	In the interest of your time, we are just going to have you shop for feminine care products today. However, you should shop as though you have other items you are purchasing on this trip to a Walgreens. Please shop for any/all of the following product types as you would on a typical Walgreens visit. Available to purchase will be tampons, pads, and liners. Please shop these products as you normally would, buying as few or as many as you usually do on a Walgreens visit. When you are finished shopping for these items, check out and we will have some additional questions for you.

Study 8: Disposable nappies

Study purpose	Determine if organising category by brands or size is optimum; best location for categories in the aisle and in the store (adjacent categories)
Scope of simulation	Navigate perimeter of entire store; shop 16 pallets of products; arranged in aisle (approx. 15 meters long); Costco branding visible in background
Product range	Range of cheap and expensive products; one pack size
Sample criteria	18+ years of age; child in household that is 0-48 months; have shopped at Costco in the last 12 months; have purchased nappys and/or training pants in the last 12 months
Recruitment	Commercial online panel; respondents from 50 US states
Fielding mode	via the internet on shoppers home computers; no download software to participate
Interface	Consumers' home computer mice/trackpads and screens
Sample size	200 per layout; 1402 total
Instructions for trip validated	For this exercise, assume you are out of nappies, wipes and/or training pants. You've decided to make a trip to Costco to do some shopping and you've come upon the baby care aisle. Please shop the baby care aisle as you would if you were out of nappies, wipes and/or training pants. Please note you are not required to make a purchase in order to complete the exercise.

APPENDIX 5

EXAMPLES OF QUESTIONNAIRES USED

Package Design Study

1. Screening questionnaire

Good morning/afternoon/evening. My name is _____ from [research agency], a national research organization. We are conducting research on consumers' opinions about various products. I wonder if you could spare a few minutes to answer some questions for me?

1a. Have you been interviewed for a market research study in the past 6 months?

Yes -- **(TERMINATE)** No -- **(CONTINUE)**

1b. Do you, or does anyone in your family work in any of the following occupations or companies: - **(IF YES, TERMINATE)**

A market research, marketing advertising or public relations company

A manufacturer, retailer or wholesale distributor of pharmaceutical or over-the-counter health care or beauty aids

A manufacturer, retailer or wholesale distributor of household paper goods

A health care professional such as a doctor, **)** nurse, pharmacist, optometrist or health care administrator

A Pharmacy pharmacy or grocery store

A mass merchandiser such as Wal-Mart

A manufacturer or distributor of household cleaning products -

A manufacturer or distributor of food products

2. **(HAND RESPONDENT CARD .)** In which of the following age groups do you belong? **(CIRCLE BELOW.)**

A. Under 18

B. 18-29 -1

C.	30-39	-2
D.	40-49	-3
E.	50-59	-4
F.	60-65	-5
G.	66 and over	

(DO NOT READ) - Refused

3. Are you the head of household responsible for the majority of your family's grocery shopping?

Yes -1 - (**CONTINUE)** No -2 - **(TERMINATE)**

4a. Now, I am going to read you a list of various products. As I read each one, please tell me which of these products you have personally purchased and used in the past month. **(READ LIST. CIRCLE RESPONSES UNDER Q4a BELOW.)**

4b. (**ASK FOR EACH PRODUCT MENTIONED IN Q4a.)** In a typical month, how many (**INSERT PRODUCT FROM Q4a)** do you and your family use**? (RECORD EXACT NUMBER BELOW Q4b ACROSS FROM APPRPRIATE PRODUCT)**

4c. **(HAND RESPONDENT CARD . ASK FOR EACH PRODUCT MENTIONED IN Q4a.)** You mentioned that you have used **(INSERT PRODUCT FROM Q4a)**. Please tell me which brands of **(INSERT PRODUCT FROM Q4a)** you see listed on this card that you have purchased and used in the past month? **(CIRCLE RESPONSE(s) UNDER Q4c.)**

4d. **(RESPONDENT STILL HAS CARD). ASK FOR EACH PRODUCT MENTIONED IN Q4c.)** Which brand of **(INSERT PRODUCT FROM Q4a)** do you purchase most often? **(CIRCLE ONE RESPONSE FOR EACH PRODUCT UNDER Q4d.)**

Q4a Products Purchased And Used In The Past Month		Q4b # Purchase in typical month	Q4c Brands Purchased And Used In The Past Month		Q4d Brand Purchased Most Often	
Rolls of paper towels*	-1	_____^	Bounty	-1	Biunty	-1
			Bounty Basic	-2	Bounty Basic	-2
			Bounty Extra Soft	-3	Bounty Extra Soft	-3
			Brawny	-4	Brawny	-4
			Mardi Gras	-5	Mardi Gras	-5
			Scott	-6	Scott	-6
			Sparkle	-7	Sparkle	-7
			Viva	-8	Viva	-8
			Store Brand	-9	Store Brand	-9
			Other	-10	Other	-10
Bottles of water	-2	_____	Aquafina	-1	Aquafina	-1
			Dasani	-2	Dasani	-2
			Evian	-3	Evian	-3
			Fugi	-4	Fugi	-4
			Nestle	-5	Nestle	-5
			Poland Spring	-6	Poland Spring	-6
			Store brand	-7	Store brand	-7
			Other	-8	Other	-8
Plastic garbage bags	-3	_____	Glad	-1	Glad	-1
			Hefty	-2	Hefty	-2
			Ruffies	-3	Ruffies	-3
			Store brand	-4	Store brand	-4
			Other	-5	Other	-5

REFER TO NEXT PAGE FOR QUALIFICATION INSTRUCTIONS

QUALIFICATION INSTRUCTIONS

QUALIFICATION INSTRUCTIONS :

- MUST BE THE HEAD OF HOUSEHOLD RESPONSIBLE FOR THE MAJORITY OF FAMILY'S GROCERY SHOPPING (Q_3)
- MUST HAVE PURCHASED AND USED PAPER TOWELS IN THE PAST MONTH (Q_4a)
- MUST TYPICALLY USE AT LEAST 2 OR MORE ROLLS OF PAPER TOWELS IN A TYPICAL MONTH (Q_4b)
- RECORD MOST OFTEN USAGE (Q_4d) ON FRONT OF SCREENER AS:
    - VIVA – IF VIVA SELECTED IN Q_4D
    - OTHER – IF ANOTHER BRAND (OTHER THAN VIVA SELECTED IN Q_4D)

NOTE: WE EXPECT TO THAT THE RANDOM SAMPLE WILL YIELD ~15-20 VIVA MOST IFTEN USERS AND AS SUCH WILL LIKELY HAVE TO SUPPLEMENT ~55. HOWEVER, WILL NEED TO KEEP TRACK OF BRAND USAGE IN ORDER TO DETERMINE EXACTLY HOW MANY SUPPLEMENTAL/AUGMENT INTERVIEWS TO CONDUCT. WILL NEED TO OBTAIN A TOTAL OF 75 VIVA MOST OFTEN USERS PER CELL

5. (HAND RESPONDENT CARDS) I am going to read you a list of statements. Using the scale on this card, please tell me how much you agree or disagree with each. **(READ EACH STATEMENT. SELECT ONE RESPONSE FOR EACH STATEMENT)**

	Agree Strongly				Strongly Disagree
The main thing I consider when purchasing paper towels, is the quality of the product, not the price	-5	-4	-3	-2	-1
When purchasing paper towels, I usually prefer the cheapest option	-5	-4	-3	-2	-1
I inquire about others well-being	-5	-4	-3	-2	-1
I neglect my duties	-5	-4	-3	-2	-1
It is worth it to me to buy higher quality paper products because it takes less product to do the job	-5	-4	-3	-2	-1
I do not plan ahead	-5	-4	-3	-2	-1
When buying toilet paper, I prefer the least expensive product that is adequate but inexpensive	-5	-4	-3	-2	-1
Using a high quality paper towel actually makes cleaning more fun	-5	-4	-3	-2	-1
I tend to spend more money each month than I have available, so I end up owing more on credit cards and loans	-5	-4	-3	-2	-1
I know how to comfort others	-5	-4	-3	-2	-1

NOTE: IN ORDER TO DETERMINE EACH SHOPPER'S "SEGMENT" THE ANSWERS TO THE ABOVE QUESTION MUST BE ENTERED INTO THE EXCEL SPREADSHEET PROVIDED. PLEASE RECORD SHOPPERS SEGMENT ON FRONT OF SCREENER. THEY WILL FALL INTO ONE OF THE FOLLOWING CATEGORIES:

- **Pragmatic**
- **Uninvolved**
- **Performance**

NOTE: WE EXPECT TO THAT THE RANDOM SAMPLE WILL YIELD ~50 PERFORMANCE USERS PER CELL AND AS SUCH WILL LIKELY HAVE TO SUPPLEMENT ~25 PER CELL. HOWEVER, WILL NEED TO KEEP TRACK OF USAGE IN ORDER TO DETERMINE EXACTLY HOW MANY SUPPLEMENTAL/AUGMENT INTERVIEWS TO CONDUCT. WILL NEED TO OBTAIN A TOTAL OF 75 PERFORMANCE USERS PER CELL.

SCREENER'S NAME_____ time_____:_____ (am/pm)

2. VR shopping task & post-shopping questions

6. Assume for a moment that you have come to the store for paper towels. Please start here and go shop as you would normally shop if you were in a store shopping for paper towels. You may buy as many or as few products as you wish.

Findability exercise:

7. Now, imagine that someone has sent you to the store to look for a specific paper towel product. When you are ready please hit "next" and locate the "1 count VIVA Big Roll White" in the shelf set and put it into your cart. Again, you will not be shopping as your normally would but are going to the shelf set to pick up the "1 count VIVA Big Roll White". Once your find this 1 count VIVA Big Roll White– please place it in your cart.

PROGRAMMING: PLEASE DO NOT ALLOW SHOPPER TO SELECT MULTIPLE ITEMS. ONCE THEY SELECT AND ITEM AND PLACE IT IN THEIR CART – THE EXERCISE IS OVER.

8. Using the scale below, please select the statement which best describes, how easy or hard do you think it was to find the "Viva Big Roll White "? (SELECT ONE RESPONSE.)

- Very easy to find ☐
- Somewhat easy to find ☐
- Neither easy nor hard to find ☐
- Somewhat hard to find ☐
- Very hard to find ☐

FOR Q 8 – 9B SHELF IS STILL IN VIEW

9a Now please focus on the entire line of "Viva" products you see here. Using the scale below, please select the statement which describes how easy or difficult you feel it is to tell the difference among the roll sizes in this "Viva" line you see on the shelf. (SELECT ONE RESPONSE.)

- Very easy to tell the difference ☐
- Somewhat easy to tell the difference ☐
- Neither easy nor difficult to tell the difference ☐
- Somewhat difficult to tell the difference ☐
- Very difficult to tell the difference ☐

9b Now please focus on the entire line of "Viva" products you see here. Using the scale below, please select the statement which describes how easy or difficult you feel it is to tell the difference among the various types of products (i.e., prints, choose a size) in this "Viva" line you see on the shelf. (SELECT ONE RESPONSE.)

- Very easy to tell the difference ☐
- Somewhat easy to tell the difference ☐
- Neither easy nor difficult to tell the difference ☐
- Somewhat difficult to tell the difference ☐
- Very difficult to tell the difference ☐

9c. Again, still looking at the "Viva" products you see here, do they look like a family of products, that is, how well do they fit together as a line? (SELECT ONE RESPONSE.)

- Extremely well ☐
- Very well ☐
- Somewhat well ☐
- Not very well ☐
- Not well at all ☐

The packaging that products come in can often convey certain impressions of the products inside. In the photos before you are two brands of paper towels. Please take a moment to examine these products and then we will discuss these packages and your perceptions of the products based on their overall packaging. While this is just a selection of the products that each brand makes, please assume that all of the varieties you are familiar with would be available in the store where you usually shop.

Please note – regardless rotation the above paragraph should appear before shoppers are asked about either Viva or Bounty/Brawny. Right now in the odd #'d rotations this does not occur - thanks

PROGRAMMER: EXPOSE IMAGES OF VIVA AND COMPETITION PHOTOS ACCORDING TO PHASE:

PHASE 1 = CURRENT VIVA & BOUNTY OR BRAWNY

PHASE 2 =PROPOSED VIVA DESIGN 1 & BOUNTY OR BRAWNY

PHASE 3 = PROPOSED VIVA DESIGN 2 & BOUNTY OR BRAWNY

ALLOW RESPONDENT TIME TO EXAMINE.

~~**SEQUENCE 1**~~

For the next few questions I'd like you to focus on these **"Viva"** packages you see here.

PROGRAMMER: EXPOSE LINE PHOTO OF VIVA ACCORDING TO PHASE:

PHASE 1 = CURRENT VIVA **PHASE 2 =PROPOSED VIVA DESIGN 1**

PHASE 3 = PROPOSED VIVA DESIGN 2

10a. Based on what this overall packaging shows and tells you, what are your impressions of these "Viva" products? **(PLEASE TYPE YOUR RESPONSE BELOW)**

BE SURE TO FOCUS ON WHAT THE PACKAGE TELLS YOU ABOUT THE PRODUCT ITSELF.

10b. And, what, if anything, is different about these products relative to other brands? **(PLEASE TYPE YOUR RESPONSE BELOW)**

BE SURE TO FOCUS ON WHAT THE PACKAGE TELLS YOU ABOUT THE PRODUCT ITSELF.

11a. Now, we'd like to know how appealing these packages are to you. Please select the one statement below that best describes how appealing you find the overall packaging of these "**Viva**" paper towels. **(SELECT ONE RESPONSE.)**

Like them very much	☐
Like them somewhat	☐
Neither like nor dislike them	☐
Dislike them somewhat	☐
Dislike them very much	☐

11b. Why do you say that? **(PLEASE TYPE YOUR RESPONSE BELOW)**

12. Below is a list of words and phrases which could be used to describe packaging for various products. Please select the words and phrases, if any, you would use to describe these "Viva" packages. You may select as many or as few as you like. **(SELECT WORDS BELOW. AGAIN, YOU MAY SELECT AS MANY OR AS FEW AS YOU LIKE)**

PROGRAMMER: ROTATE LIST

Unattractive	☐	Unique	☐
Is environmentally friendly	☐	Hard to read	☐
Vivacious/cheerful looking	☐	Looks out of date	☐
High quality looking	☐	Informative	☐
Modern/up-to-date	☐	Attractive	☐
Easy to read	☐	Easy to understand	☐
Appropriate for this product	☐	Generic looking	☐
Premium	☐	Dull/drab	☐
Colorful	☐	Feminine	☐
Cluttered	☐	Vibrant	☐
Simple	☐	Looks creative	☐

1. Below is a list of statements which could be used to describe these "**Viva**" products. As you read each one, please select the <u>one</u> statement which best describes how much you agree or disagree that the phrase applies to "Viva" <u>based on the packaging you see here</u>. **(SELECT ONE PHRASE FOR EACH STATEMENT. CONTINUE UNTIL ALL ITEMS ARE RATED.)**

PROGRAMMER: ROTATE LIST

Let's begin. "Viva paper towels"…	**Agree Strongly**	**Agree Somewhat**	**Neither Agree Not Disagree**	**Disagree Somewhat**	**Disagree Strongly**
Are made by a brand I can trust	☐	☐	☐	☐	☐
Absorb quickly	☐	☐	☐	☐	☐
Absorb a lot	☐	☐	☐	☐	☐
Are high quality products	☐	☐	☐	☐	☐
Are better than other brands of paper towels	☐	☐	☐	☐	☐
Are suitable for a wide range of tasks	☐	☐	☐	☐	☐
Are strong and durable	☐	☐	☐	☐	☐
Have thick sheets	☐	☐	☐	☐	☐
Are soft	☐	☐	☐	☐	☐
Don't tear or fall apart when wet	☐	☐	☐	☐	☐
Are for someone like me	☐	☐	☐	☐	☐
Are like cloth	☐	☐	☐	☐	☐
Are worth paying more for	☐	☐	☐	☐	☐
Are a premium product	☐	☐	☐	☐	☐
Are soft enough to use on your hands and face	☐	☐	☐	☐	☐
Could be used in place of a cloth	☐	☐	☐	☐	☐

14a. Thinking about the papertowel products you are familiar with, are these **"Viva"** packages the same as or different from the **"Viva"** packages that you have seen in the store where you usually shop? (**SELECT <u>ONE</u> RESPONSE.)**

Same ☐ **(SKIP TO <u>Q15</u>)** ()

Different ☐ **(ASK Q14b & 14c)**

(DO NOT READ) Don't Know ☐ **(SKIP TO Q15)**

14b. **(IF "DIFFERENT" IN Q14a, ASK:)** Do you feel the change is for the better or worse? **(SELECT ONE RESPONSE.)**

Better ☐ Worse ☐ Same ☐

14c. **(IF "DIFFERENT" IN Q14a, ASK:)** Do you think the product inside has changed? **(SELECT ONE RESPONSE.)**

Yes, the product has changed ☐ **(ASK Q14d)**

No, the product is the same ☐ **(SKIP TO Q15)**

14d. **(IF "YES" IN Q14c, ASK):** And do you feel the change is for the better or worse? **(SELECT ONE RESPONSE.)**

Better ☐ Worse ☐ Same ☐

15. Based upon what these **"Viva"** packages show and tell you, how likely would you be to buy any of these **"Viva"** products if they were available where you generally shop. **(SELECT ONE RESPONSE.)**

Definitely would buy ☐
Probably would buy ☐
Might or might not buy ☐
Probably would not buy ☐
Definitely would not buy ☐

16. Now, we would like to know if your interest in buying **"Viva"** has changed from before coming here today.

If seeing these packages has…

- **Not changed your interest** in buying this brand, choose the middle box (0).
- Made you **more likely than before** to buy this brand, choose the appropriate box on the right.
- Made you **less likely than before** to buy this brand, choose the appropriate box on the left.

Less Likely Than Before To Buy				About **As Likely As** Before To Buy	**More** Likely Than Before To Buy			
A lot less		A little less			A little more		A lot more	
-4	-3	-2	-1	0	+1	+2	+3	+4
☐	☐	☐	☐	☐	☐	☐	☐	☐

17. If one roll of **"Bounty Big Roll White"** were priced at **$2.07**, how much would you expect to pay for one roll of **"Viva Big Roll White"** product? **(RECORD RESPONSE ON LINES PROVIDED BELOW.)**

(ASSUME PRODUCTS ARE NOT ON SALE – THERE IS NO COUPON.)

$ _____ . _____ _____
(DOLLARS) (CENTS)

For the shoppers who see Viva and Brawny –please substitute "Brawny Big Roll for $1.85"

18. The actual price of one roll of **"Viva Big Roll White"** like the ones you see here is **$2.22**. Considering the price, which statement on this card best describes how you feel about the VALUE of these products? **(SELECT ONE RESPONSE.)**

A very good value	☐
A fairly good value	☐
An average value	☐
A somewhat poor value	☐
A very poor value	☐

19a. Based upon what these **"Viva"** packages show and tell you, how likely would you be to buy one roll of "**Viva Big Roll White"** if it were priced at **$2.22,** assuming you were in the store shopping for paper towel in single rolls? **(SELECT ONE RESPONSE.)**

Definitely would buy	☐	
Probably would buy	☐	
Might or might not buy	☐	
Probably would not buy	☐	ASK Q19b
Definitely would not buy	☐	

19b. (ASK Q19b IF SELECTS "1" OR "2" IN Q19a) Why do you say you **(INSERT RESPONSE FROM Q19a)? (TYPE RESPONSE BELOW.)**

REMOVE PHOTO FROM VIEW

PROGRAMMER: EXPOSE PHOTO OF COMPETITION:

BOUNTY **OR** **BRAWNY**

For the next set of questions, please focus your attention on these **"Bounty"** packages you see here.

NOTE: the sample will be split so that half of the shoppers will be asked about Viva and Bounty and half will be asked about Viva and Brawny.

20. Now, we'd like to know how appealing these packages are to you. Please select the one statement below that best describes how appealing you find the overall packaging of these **"Bounty"** paper towels. **(SELECT ONE RESPONSE.)**

Like them very much	☐
Like them somewhat	☐
Neither like nor dislike them	☐
Dislike them somewhat	☐
Dislike them very much	☐

21. Below is a list of words and phrases which could be used to describe packaging for various products. Please select the words and phrases, if any, you would use to describe these "**Bounty"** packages. You may select as many or as few as you like. **(SELECT WORDS BELOW. AGAIN, YOU MAY SELECT AS MANY OR AS FEW AS YOU LIKE)**

PROGRAMMER: ROTATE LIST

Unattractive	☐	Unique	☐
Is environmentally friendly	☐	Hard to read	☐
Vivacious/cheerful looking	☐	Looks out of date	☐
High quality looking	☐	Informative	☐
Modern/up-to-date	☐	Attractive	☐
Easy to read	☐	Easy to understand	☐
Appropriate for this product	☐	Generic looking	☐
Premium	☐	Dull/drab	☐
Colorful	☐	Feminine	☐
Cluttered	☐	Vibrant	☐
Simple	☐	Look creative	☐

22. Below is a list of statements which could be used to describe these **"Bounty"** products. As you read each one, please select the <u>one</u> statement which best describes how much you agree or disagree that the phrase applies to **"Bounty"** <u>based on the packaging you see here</u>. **(SELECT ONE PHRASE FOR EACH STATEMENT. CONTINUE UNTIL ALL ITEMS ARE RATED.)**

PROGRAMMER: ROTATE LIST

Neither

Let's begin. "Viva paper towels"...	Agree Strongly	Agree Somewhat	Agree Not Disagree	Disagree Somewhat	Disagree Strongly
Are made by a brand I can trust	☐	☐	☐	☐	☐
Absorb quickly	☐	☐	☐	☐	☐
Absorb a lot	☐	☐	☐	☐	☐
Are high quality products	☐	☐	☐	☐	☐
Are better than other brands of paper towels	☐	☐	☐	☐	☐
Are suitable for a wide range of tasks	☐	☐	☐	☐	☐
Are strong and durable	☐	☐	☐	☐	☐
Have thick sheets	☐	☐	☐	☐	☐
Are soft	☐	☐	☐	☐	☐
Don't tear or fall apart when wet	☐	☐	☐	☐	☐
Are for someone like me	☐	☐	☐	☐	☐
Are like cloth	☐	☐	☐	☐	☐
Are worth paying more for	☐	☐	☐	☐	☐
Are a premium product	☐	☐	☐	☐	☐
Are soft enough to use on your hands and face	☐	☐	☐	☐	☐
Could be used in place of a cloth	☐	☐	☐	☐	☐

23. Based upon what these **"Bounty"** packages show and tell you, how likely would you be to buy any of these **"Bounty"** products if they were available where you generally shop. (**SELECT ONE RESPONSE.)**

Definitely would buy ☐
Probably would buy ☐
Might or might not buy ☐
Probably would not buy ☐
Definitely would not buy ☐

24. Now, we would like to know if your interest in buying **"Bounty"** has <u>changed</u> from before coming here today.

If seeing these packages has...

- **<u>Not changed your interest</u>** in buying this brand, choose the middle box (0).

- Made you **more likely than before** to buy this brand, choose the appropriate box on the right.
- Made you **less likely than before** to buy this brand, choose the appropriate box on the left.
-

Less Likely Than Before To Buy | About **As Likely As** Before To Buy | **More** Likely Than Before To Buy

A lot less | A little less | | | | A little more | A lot more

-4	-3	-2	-1	0	+1	+2	+3	+4		
☐	☐	☐	☐	☐		☐	☐	☐	☐	

25. The actual price of one roll of **"Bounty Big Roll White"** like the ones you see here is **$2.07**. Considering the price, which statement on this card best describes how you feel about the VALUE of these products? **(SELECT ONE RESPONSE.)**

A very good value	☐
A fairly good value	☐
An average value	☐
A somewhat poor value	☐
A very poor value	☐

26. Based upon what these **"Bounty"** packages show and tell you, how likely would you be to buy one roll of **"Bounty Big Roll White"** if it were priced at **$2.07** assuming you were in the store shopping for paper towel in single rolls? **(SELECT ONE RESPONSE.)**

Definitely would buy	☐
Probably would buy	☐
Might or might not buy	☐
Probably would not buy	☐
Definitely would not buy	☐

For the shoppers who see Viva and Brawny –please substitute "Brawny Big Roll for $1.85"

3. Usage and segmentation questions

Now, just a few questions for classification purposes.

C1. Which of the following types of paper towel do you generally purchase? **(SELECT ONE MENTION)**

White	☐

Prints ☐

C2a. And, what roll size due you typically purchase? **(SELECT ONE MENTION)**

Regular roll ☐

Large/mega roll☐

C2b. Do you typically purchase standard size paper towels or ones where you can create/choose a size? **(SELECT ONE MENTION)**

Standard sized paper towels ☐

Choose-a-size/Pick-a-size/Select-a-size/ ☐

C3. Do you typically buy paper towels in single rolls or as part of a bundle or multi pack? **(SELECT ONE MENTION)**

Single roll ☐

Bundle/multi pack (and how many in rolls are included?) ☐

rolls per bundle/multi pack _____

C4. Where do you purchase your paper towels? **(SELECT <u>ALL</u> THAT APPLY)**

<u>A mass merchandiser (please specify:)</u>

Target ☐

Wal-Mart ☐

Kmart ☐

A convenience store ☐

A Pharmacy (i.e. Eckerd, CVS) ☐

A supermarket or grocery store (i.e. Kroger, Publix) ☐

A warehouse club store (i.e. Costco, Sam's Club) -7

Via the internet -8

Other _____ -9
(specify)

C5. What is your marital status? **(SELECT <u>ONE</u> RESPONSE.)**

Married ☐

Single ☐

Separated/divorced ☐

Other ☐

Refused ☐

C6. What is the last year of schooling that you completed? **(SELECT <u>ONE</u> RESPONSE.)**

Some high school or less	☐	Graduated college (4 years)	☐
Graduated high school	☐	Post-graduate work	☐
Some college (less than 4 years)	☐	Refused	☐

C7a. Including yourself, how many people are there in your household? (**TYPE IN RESPONSE BELOW)**

_____

C7b. Including yourself, how many are…

18 or over _____

13-17 _____

9-12 _____

6-8 _____

Under 6_____

C8. Which **best** describes your current work/employment status? **(SELECT ONE RESPONSE.)**

Working full-time ☐

Working part-time ☐

Currently not working ☐

Full-time homemaker ☐

Full-time student ☐

Retired ☐

C9. Which of the following ranges best represents your household's total annual income. **(SELECT ONE RESPONSE.)**

Under $15,000	☐	$100,000-$124,999	☐
$15,000-$24,999	☐	$125,000-$149,999	☐
$25,000-$34,999	☐	$150,000-$199,999	☐
$35,000-$49,999	☐	$200,000-$249,999	☐
$50,000-$74,999	☐	$250,000 or more	☐
$75,000-$99,999	☐	Prefer not to answer	☐

C10. Please select the one of these races/ethnicities best describes you. **(SELECT ONE RESPONSE)**

A. Asian / Other Pacific Islander ☐

B. Black / African-American ☐

C. Hispanic / Latin / Mexican ☐

D. Middle Eastern / North African ☐

E. Native American / Alaskan Native ☐

F. White / Caucasian ☐

G. Other ☐

RESPONDENT'S NAME:_____
ADDRESS (STREET):_____(TOWN)_____
ADDRESS (STATE):_____(ZIP)_____
TELEPHONE NUMBER: (AREA CODE) ( )_____

INTERVIEWER'S NAME:_____(DATE)_____

TIME ENDED:_____
LENGTH OF INTERVIEW:_____
(IN MINUTES)

APPENDIX 6

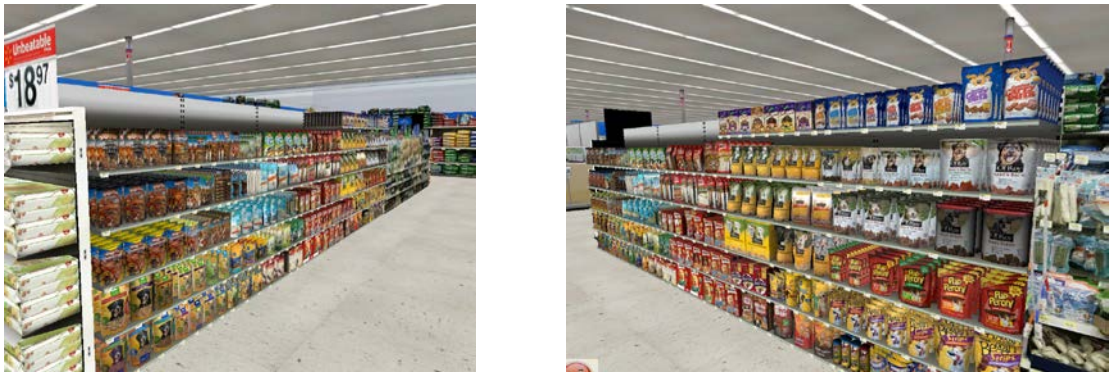
DESIGN CONSIDERATIONS FOR VR STUDIES

Consideration 1: Shoppers' starting positions

As discussed in Chapter 5, most of the virtual reality studies commissioned by marketers focus on one product category, aisle or area of the store. Consequently, decisions about how and where shoppers will commence their trip must be made. Figure 28 provides an illustration of how shoppers' perspective differs when they are assigned to start at one end of a virtual aisle instead of the other.

Figure 28: Alternative starting positions in a category-focused virtual simulation

Example: View with left starting v. right starting position



In virtual studies, shoppers starting positions are sometimes manipulated to control the direction shoppers approach the category.

In many studies, shoppers starting positions are alternated (i.e. 50% start at right; 50% at left) to ensure that results will be generalisable to stores where shoppers can approach the category from either end of the aisle. However, in only two of the studies examined, was the starting position of participants recorded. This suggests that researchers do not routinely analyse results based on the starting position of respondents. A problematic finding given that past studies have found shoppers to be significantly faster and more accurate when assessing the quantity of stock available on virtual shelves in their left visual field than in their right visual field (accuracy +7% $p < .01$; speed +6% $p < .05$; (Massara, Porcheddu et al. 2014).

Before proceeding with analyses for Chapter 9, it was important to validate that the proportion of shoppers starting at each end of the aisle was not significantly different across the three layouts tested. This was confirmed with a chi-squared test of independenence on the 2x3 contingency table ($X^2(2, n = 236) = 4.29$, $p = 0.12$). It was also interesting to see how shoppers starting position affected shopping behaviours in a realistic shelf environment (Massara, Porcheddu et al. used plain white three-dimensional boxes in place of real products).

Results: Impact of starting positions

Shoppers who started at the left-end of the assortment appeared to select more items and shopped more quickly than shoppers who started at the right end of the aisle in this and other virtual simulations. While the differences were not large or significantly different—on average shoppers selected 8% more items when starting on the left rather than the right (M_{items} = 3.25 vs 3.0, StD = 1.64 vs 1.55; t (284) = 1.58 p = .11) and took 6% less time (M_{time} = 84s vs 89s, StD = 60 vs. 58 ; N.S. different t (284) = 0.28 p = .77)— small differences could be enough to obscure the effects of shelf layout changes if the number of shoppers starting on the left and right end of the aisle was not carefully controlled in VR studies.

The impact of this factor was not confirmed with modelling as so few studies recorded the starting position of participants.

Discussion & recommendations for future research

That the starting position of shoppers was recorded in so few studies, suggests that many marketers are unaware of impact that starting position can have on visual perception and purchase behaviour in virtual studies. More research is needed to establish why participants navigate virtual simulations more quickly and select more items when starting with particular views of the aisle.

It is possible that there is a natural preference for left to right movement in computer-simulated environments—shoppers' eyes tend to move from left to right when viewing product assortments in computerised settings (Orquin, Bagger et al. 2013, Shi, Wedel et al. 2013, Burke and Leykin 2014) and cartoons, movies and video games typically portray motion from left to right, even in countries with languages that written in the opposite direction (Walker 2015). When shoppers are assigned to a left-end starting position they must move right to shop the rest of the category but when they are assigned to a right-end starting position, they must move left to see other products, a motion that may feel unnatural and may explain the small, non-significant differences in results.

While much more research is needed in real stores, explorative studies have found that shoppers who travel with the dominant flow of traffic in a store spend more on average and travel shorter distances than shoppers who go against the flow (Groeppel-Klein and Bartmann 2008, Kholod, Takai et al. 2011). However, the serpentine nature of many shopping paths means that the dominant direction of traffic will differ for categories depending on their position in store (Larson, Bradlow et al. 2005).

Past studies have found that when stimuli are presented in viewers' left field of view, behavioural responses to global features (i.e. scenes and patterns) are faster and more accurate (Massara, Porcheddu et al. 2014). In contrast, when stimuli are presented in viewers' right field of view responses to local features (i.e. specific details are faster), perhaps due to hemispherical specialisation in the brain (Robertson and Lamb 1991, Fink, Marshall et al. 1997). Whether shoppers were left or right handed, a metric that often indicates which side of the brain is more dominant, was not recorded in this study.

Consideration 2: Tailoring VRs to shoppers' real heights

In more sophisticated VR simulations shoppers' angles of view are tailored to match their real height. How this affected the generalisability of results in other VR studies is discussed in Appendix 6. Figure 29 provides an example of what a virtual shelf would look like when adjusted for a taller shopper (175 cm / 5 foot 10 inches) —products on the middle and highest shelves are easy to read/see, while products on the lower shelves are quite difficult read/see due to the steep angle of viewing.

Figure 29: View of virtual shelf adjusted for a moderately tall shopper



Tailoring the angle of view to match the (real) height of shoppers/subjects, makes products on lower shelves are more difficult to see for taller shoppers than shorter shoppers.

In Table 54 and Table 55 the virtual purchases of taller and shorter shoppers are compared from two independent to provide insight into whether tailoring the view of shoppers' affects the products they choose. For the purposes of this analysis, products on the middle shelves were between 75 and 150cm from the floor (i.e. between eye and hand height for most shoppers).

Results: Effect of angle of view on purchase choices

Results suggest that shoppers' angle of view does not influence substantially what items they select. Shoppers taller than 170 centimeters selected slightly more products from the top shelves in a virtual simulation for confectionery products than shoppers shorter than 170 centimeters (Table 55: 35.3% of purchases vs 38.8%, not statistically different, X^2 (2, n = 44) = .58, p = .74). Similarly, in a study for pet products, taller shoppers selected more products from the top shelves than shorter shoppers (Table 56: 22.8% of purchases vs 20.6%, not statistically different X^2 (2, n = 28) = 1.2, p = .55).

Table 55: Impact of viewing angle on purchase choices in a confectionery study

Purchases from:	Total Sample n=86	Taller Shoppers >170cm / 5`7" n=44	Shorter Shoppers <170cm / 5`7" n=42
Top Shelf >150cm	32.1%	35.3%	38.8%
Middle Shelves ~ 100cm	43.5%	43.1%	42.2%
Lower Shelves <75 cm	24.4%	21.6%	19.0%
Total	100.0%	100.0%	100.0%

Tall shoppers appeared to select slightly more products from the top shelves than shorter shoppers but results were not statistically significant given the sample sizes for each group.

Table 56: Impact of viewing angle on purchase choices in a petcare study

Purchases from:	Total Sample n=50	Taller Shoppers >170cm / 5`7" n=28	Shorter Shoppers <170cm / 5`7" n=22
Top Shelf >150cm	21.8%	22.8%	20.6%
Middle Shelves ~ 100cm	37.5%	32.8%	43.5%
Lower Shelves <75 cm	40.7%	44.4%	36.1%
Total	100.0%	100.0%	100.0%

Tall shoppers appeared to select slightly more products from the top shelves than shorter shoppers but results were not statistically significant given the sample sizes for each group.

Interestingly, in both of these studies between 20% and 35% of the products selected by shoppers were from the top shelves, a greater proportion than in real stores. For example, the products located on the top shelves of the VR pet care study had a market share equivalent to 17.4% in real stores, 25% less than in the virtual study. This suggests that while adjusting the angle of view to reflect shoppers' real heights influences what products they choose, more shoppers choose products from the top shelves in virtual simulations than real stores, a result consistent with past studies (e.g. in van Herpen, van den

Broek et al. 10% more shoppers selected products from the top shelves in the VR simulation than the real store).

Discussion & recommendations for future research

Based on the results in Table 55 and Table 56 between nine and 16% of shoppers a select product that they would have difficulty reaching in real stores without assistance. In future studies it may be worth identifying these shoppers and removing or screening their purchases to see if the overall accuracy of VR purchase data can be improved.

The possibility that more shoppers select products from the bottom shelves in virtual studies than in real stores because it is easier to do so than in real stores should also be considered. In a VR shopping study in the Netherlands, 46.7% of shoppers selected a product from the lowest shelf, 57% more than in the real store on which the simulation was modelled (van Herpen, van den Broek et al. 2016). As with the studies validated here, shoppers viewed the VR simulation on a normal computer monitor and selected products using a mouse; no physical exertion was required for them to select products. It may be possible to overcome these biases by using VR goggles or life-size motion sensitive screens to display VR shopping environments.

Consideration 3: Number of product categories shopped

The main body of this thesis examined only studies where the participants shopped just one category or area of a virtual store e.g. the confectionery aisle. These are the types of simulations most commonly used by marketing managers, as they are generally only interested consumer behaviour as it relates to the product category or categories they manage.

In 2014 a study fielded with 1680 products across eleven product categories arranged as though in three aisles of a grocery store became available for analysis. This study was commissioned by a research agency interested in showing marketing managers how VR could be used for more extensive research. This study was fielded in shopping malls across the US and included many of the same brands and product categories as were validated in the single-category studies. It is examined to see if having participants shop more categories improves the realism of VR data.

Results: Number of product categories shopped

In Table 57 the purchase shares of premium, private label and large packs in this study are benchmarked to shares of these products in real stores based on sales data from retailers. Comparable results from VR studies with one product category or area of the store visible were reported in Chapters 6 and 7.

The purchase shares of large pack sizes and premium brands were consistently larger in the VR simulation where multiple categories could be shopped than in real stores (Table 57: 18% and 55% larger on average). The shares of private label products were also smaller than in real stores (Table 57: 75% smaller on average). X^2 tests were used to compare observed to expected counts of premium

brand purchases and differences were highly significant ($p < .01$) for the majority of studies. These are the same biases that were evident in virtual simulations were participants only one product category could be shopped.

In Table 41, the averages for each group of products shown is not weighted by the number of purchases for each product category as this thesis aimed to show how results vary across product categories.

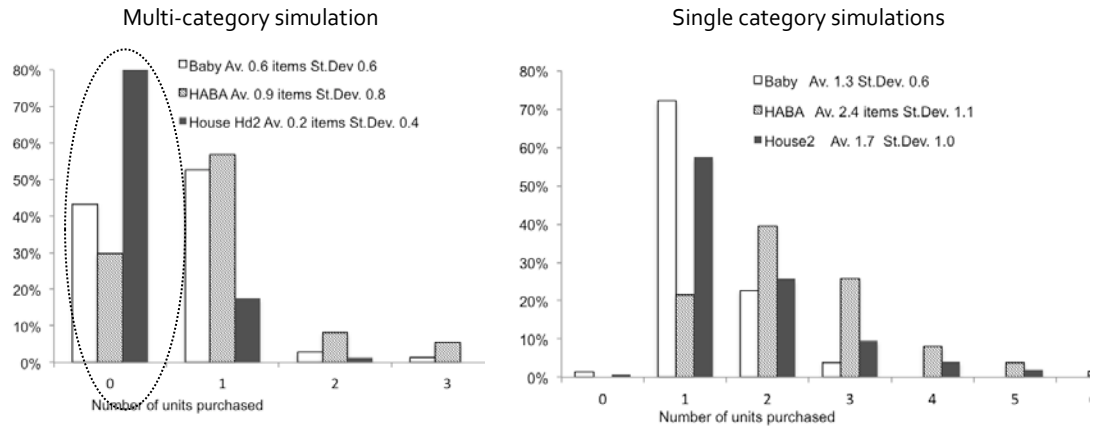
Table 57: Purchase choice results from a VR simulation with multiple categories shown

Purchase shares of:	Product Category	Real Benchmark n >10,000 purchases	VR Multi Category n ~ 201 purchases per	Statistical Results	Average Effect
Larger packs	**Baby care**	42%	54%	X^2 (1, n=201) = 12.5, $p < .01$	
Larger packs	**Feminine care**	45%	53%	X^2 (1, n=201) = 5.77, $p < .01$	
Larger packs	**Paper towels**	56%	62%	X^2 (1, n=201) = 3.05, $p < .10$	
Larger packs		48%	56%	X^2 (1, n=201) = 16.2, $p < .01$	+18% greater
Premium brand	**Baby care**	59%	62%	X^2 (1, n=201) = 0.73, $p = .38$ N.S.	
Premium brand	**Feminine care**	15%	24%	X^2 (1, n=201) = 8.9, $p < .01$	
Premium brand	**Paper towels**	39%	90%	X^2 (1, n=201) = 589, $p < .01$	
Premium brand		38%	59%	X^2 (1, n=201) = 110, $p < .01$	+55% greater
Cheap P/L brand	**Feminine care**	22%	6%	X^2 (1, n=201) = 90.7, $p < .01$	
Cheap P/L brand	**Baby care**	23%	-		
Cheap P/L brand	**Paper towels**	28%	-		
Cheap P/L brand		24%	6%	X^2 (1, n=201) = 115, $p < .01$	25% of real share

More shoppers selected large pack sizes and premium brands in the virtual simulation than do in real stores; fewer selected cheap private label brands.

In Figure 30 the number of items that shoppers selected from each product category is compared for single and multi-category VR simulations in three product categories.

Figure 30: Purchase quantities from multi and single category simulations



As Figure 30 shows, the number of categories shopped affected the number of shoppers who purchase from each category (i.e. prevalence of '0 unit' trips) and the number of products that shoppers choose from each product category. Differences were highly significant for all three categories (Baby: t (529) = 13.0 p < .01, HABA: t (354) = 14.8 p < .01, Household product: t (428) = 19.9 p < .01).

Implications: Number of product categories to include

The results from this investigation suggest that the findings from thesis regarding biases in VR purchase data are generalisable to VR simulations where multiple categories can be shopped. Having consumers shop for more products does not improve the realism of VR choice data, although it does affect the number of items purchase a metric that was overestimated in past studies (Burke, Harlam et al. 1992, Campo, Gijsbrechts et al. 1999).

Whether participants should shop multiple categories in a VR study will depend on the objectives of each study—for studies aimed at measuring penetration or purchase rates accurately, including additional categories will aid in normalising these metrics. (In real life, having shoppers walk away from the shelf without choosing a product is perfectly normal (Sorensen and Suher 2010).) For studies interested primarily in purchase choice data, additional categories will just lead to wastage in the sample as fewer purchases will be made overall. This validates the decisions of marketing managers to restrict shopping VR studies to the product category(s) to which they are most interested.

Consideration 4: Precise wording of instructions

Across the studies selected for validation there was some variation in the wording of participant instructions. For example, some studies emphasised that consumers did not need to purchase anything in order to complete their virtual shopping trip whereas others did not. The impact of this variation is examined here to confirm that using data from studies with different studies did not undermine the validity of the conclusions from this thesis.

As the results in Table 58 show, the wording of the instructions and whether shoppers were told they did not need to purchase did not affect the number of shoppers who walked away from the virtual shelf without purchasing or the number of items shoppers chose on average (M_{units} not significantly different, $p > .05$).

Table 58: Results from single category simulations with (slightly) different instructions

Study	Instructions to participants	Units Purchased	
		>1	Av.
8	Please start here and go shop as you would normally shop if you were shopping for disposable nappies… you do not need to purchase to complete your shopping trip.	99%	1.28
4	Imagine that you are shopping today at your regular store and have decided that you need to buy disposable nappies…. Please shop as you normally would for yourself or family.	99%	1.37

Consideration 5: Recruitment criteria

This investigation examined the impact of the recruitment screening criteria on the representativeness of the samples obtained. All of the studies selected for analysis in this thesis recruited regular category buyers who had shopped at the outlet simulated recently (i.e. three or six months). However, some studies imposed additional recruitment criteria, e.g. participants must be the primary household shopper, must purchase at least two units of the product per month to ensure the sample aligned with the manufacturers' or retailers' target market. These studies were not analysed because the VR purchase data was unlikely to match to sales data from real stores for which a general (i.e. all purchases) sample is used. However, they can provide insight into appropriate recruitment criteria for VR studies, a question on which there is no literature to guide decision making.

The recruitment criteria used in six VR studies (see Table 59 for details) were applied to members of Nielsen's Homescan panel, a panel that tracks the characteristics and grocery purchases of 60,000 households. The Homescan panel is designed to be representative of the US population and can provide an indication of how representative the resulting samples are likely to be of buyers generally.

In Table 59, VR studies are listed in from most to least restrictive recruitment criteria and the estimated proportion of category buyers included in these criteria is reported.

Results: Impact of recruitment criteria on the generalisability of VR data

Table 59: Proportion of buyers and sales represented by VR recruitment criteria

Study	Recruitment criteria specified	% Buyers
1	Purchased category at grocery store in past 6 months	77
4	Purchased category at grocery store or Walmart in past 3 months	68
-	Purchased category at convenience store in last 3 months	40
-	Female, aged 15-50, purchased category at Target in past 3 months	20
-	Aged 18-65, main household shopper & purchased category in last month	19
	Average	**45**

The more criteria set for recruitment, the less representative the sample is of category buyers.

The results in Table 59 demonstrate that as additional criteria are applied, i.e. specific age, specific gender, shopped at specific retailer and timeframes are narrowed; fewer category buyers are represented by the VR samples. In some instances, 80% of category buyers are excluded by the recruitment criteria. Studies with broader recruitment criteria i.e. long purchase windows, no age restrictions; generally include a greater proportion of category buyers (e.g. 77%). In some product categories e.g. feminine hygiene items, applying age or gender criteria had little impact on the representativeness of samples.

Implications: Recruitment criteria

The results suggest that less restrictive recruitment criteria to VR studies for commonly purchased grocery items and that marketers should consider using consumer panel data model the impact of different recruitment criteria to ensure their decisions will not undermine the representativeness, and usefulness of the sample. Generally marketers will be better served by results that apply broadly to category buyers (Sharp 2013).

Furthermore, when establishing recruitment criteria marketers should consider how recall and telescoping errors are likely to affect the samples obtained (Tourangeau 2000). For example, the tendency of consumers recall events as occurring more recently than they actually did (Cohen and Martin 2008), may make a sample slightly broader than a marketer intends (e.g. result in some consumers who last purchased four months ago being included in a study with a 'last three months' purchase window). Similarly the tendency of shoppers to overestimate how much they purchased in a period (Wind and Lerner 1979, Nenycz-Thiel, Beal et al. 2012), may mean that some consumers who buy less frequently than the purchase criteria are included in samples.

APPENDIX 7

ADDITIONAL RESULTS FOR CHAPTERS 6,7 AND 8

Table 60: Statistical tests for premium grocery brands

Virtual Study	Environment simulated	Real Share	Virtual Share	Statistical Results
6	Grocery store n=253	13.2%	17.5%	X^2 (1, n=253) = 3.33, $p < .10$
2	Mass retailer n=131	39.7%	51.0%*	X^2 (1, n=131) = 6.87, $p < .01$
5	Grocery store n=359	31.1%	38.4%*	X^2 (1, n=359) = 7.96, $p < .01$
3	Mass retailer n=155	45.2%	55.4%*	X^2 (1, n=155) = 6.63, $p < .01$
1	Grocery store n=1500	51.3%	62.3%*	X^2 (1, n=1500) = 77.25, $p < .01$
4	Grocery store n=359	58.6%	65.1%*	X^2 (1, n=359) = 6.50, $p < .05$
5	Mass retailer n=229	38.8%	42.2%	X^2 (1, n=229) = 1.15, N.S. $p > .10$
3	Pharmacy n=155	52.7%	56.6%	X^2 (1, n=155) = 0.95, N.S. $p > .10$
	Average grocery study total n=2442	38.6%	45.8%*	X^2 (1, n=2442) = 50.52, $p < .05$
	Average mass study n=515 purchases	41.2%	49.5%*	X^2 (1, n=515) = 14.36, $p < .05$
	Average drug study n=155 purchases	52.7%	56.6%	X^2 (1, n=155) = 0.95, N.S. $p > .10$

Table 61: Premium brand shares of shoppers of different ages in VR study

Purchase shares of premium grocery brands	VR Sample n=400	Statistical results Contingency test 5x2
Aged 18-34 yrs (n = 188)	63.8%	X^2 (4, n = 400) = 1, $p = .90$
35-44 yrs (n = 100)	60.3%	N.S. Different
45-54 yrs (n = 40)	65.9%	
55-64 yrs (n = 36)	59.5%	
65+ yrs (n = 36)	58.7%	

The VR study the share of premium brand purchases was similar across all age groups.

Table 62: Statistical tests for PL grocery brands

Virtual Study	PL brand in…		Real Share	Virtual Share	Statistical Results
2	Mass retailer	Pet care	4%	1%	Fisher's exact unconditional two-sided p-value with 95% CI > .10
1	Grocery store	Confection	3%	1%	X^2 (1, n=1500) = 60.60, p < .01
4	Mass retailer	Disposable nappies	35%	9%	X^2 (1, n=350) = 248.7, p < .01
6	Grocery store	Rice	18%	4%	X^2 (1, n=179) = 92.9, p < .01
3	Mass retailer	Fem. care	14%	10%	X^2 (1, n=155) = 3.62, p < .10
3	Pharmacy	Fem. care	22%	11%	X^2 (1, n=155) = 19.1, p < .01
5	Grocery store	Paper towels	28%	15%	X^2 (1, n=359) = 48.15, p < .01
2	Mass retailer	Pet care	16%	3%	Fisher's exact unconditional two-sided p-value with 95% CI < .01
4	Grocery store	Disposable nappies	23%	10%	X^2 (1, n=330) = 62.26, p < .01
5	Mass retailer	Paper towels	25%	7%	X^2 (1, n=229) = 112, p < .01

Table 63: Statistical tests for brands fielded online and in person

Virtual Study	Retail environment	Brand	Real Share	Virtual Share	Statistical Results
3	Pharmacy n=155	PL brand	22%	11%	X^2 (1, n=155) = 19.1, p < .01
7	Pharmacy n=300	PL brand	23%	9%	X^2 (1, n=330) = 62.26, p < .01
4	Mass store n=359	PL brand	26%	13%	X^2 (1, n=359) = 44.8, p < .01
8	Mass store n=200	PL brand	27%	17%	X^2 (1, n=200) = 17.1, p < .01
3	Pharmacy n=155	Brand B	15%	19%	X^2 (1, n=155) = 1.1, p > .10
7	Pharmacy n=300	Brand B	29%	32%	X^2 (1, n=330) = 0.66, p > .10
4	Mass store n=359	Brand A	14%	18%	X^2 (1, n=359) = 2.7, p < .10
8	Mass store n=200	Brand A	22%	28%	X^2 (1, n=200) = 3.5, p < .10

Table 64: Statistical tests for pack size choices

Virtual Study	Retail environment	Product(s)	Real Share	Virtual Share	Statistical Results

1	Grocery store n=1083	Large & XL	20.3%	34.9%	X^2 (1, n=1083) = 101, p < .01
3	Pharmacy n=155 purch'	Large & XL	35.9%	51.5%	X^2 (1, n=155) = 14.8, p < .01
2	Mass retailer n=131 purch'	Large & XL	28.4%	39.3%	X^2 (1, n=131) = 6.2, p < .05
4	Grocery store n=330 purch	Large & XL	41.9%	57.6%	X^2 (1, n=330) = 33.5, p < .01
6	Grocery store n=253 purch	Large & XL	21.5%	28.2%	X^2 (1, n=253) = 5.6, p < .05
5	Grocery store n=359 purch	Large & XL	55.6%	70.2%	X^2 (1, n=359) = 36.0, p < .01
3	Mass retailer n=155 purch'	Large & XL	45.3%	55.6%	X^2 (1, n=155) = 6.7, p < .01
4	Mass retailer n=330 purch'	Large & XL	71.7%	79.5%	X^2 (1, n=330) = 11.6, p < .01
5	Mass retailer n=229 purch'	Large & XL	60.9%	70.9%	X^2 (1, n=229) = 11.1, p < .01
7	Pharmacy n=300 purch'	Large & XL	34.5%	49.1%	X^2 (1, n=300) = 27.2, p < .01
Average drug study n=610 purchases		Large & XL	35.7%	50.3%	X^2 (1, n=610) = 51.9, p < .01
Average grocery study not weighted, total n=2025		Large & XL	34.8%	47.7%	X^2 (1, n=2025) = 138, p < .01
Average mass study n=845 purchases		Large & XL	51.6%	61.3%	X^2 (1, n=845) = 33.5, p < .01

Table 65: Results from 'seek and find' tasks

Product Assigned	Category sample N.B.from control cell	Identify correctly	Av.time to find
Product B	Disposable nappies n=100	95%	47secs
Product A	Disposable nappies n=100	91%	47secs
Brand B	Confectionery n=429	90%	11secs
Brand C	Confectionery n=429	90%	43secs
Product A	Paper towels n=170	89%	28secs
Product A	Pet care n=50	88%	28secs
Brand D	Confectionery n=429	87%	30secs
Product A	Feminine care n=100	86%	50secs
Brand A	Confectionery n=429	86%	31secs
Product B	Feminine care n=100	82%	51secs
Product C	Feminine care n=100	63%	24secs
	Average	86%	33secs

Table 66: Colour analysis results

Study 1 See Chapter 7

Study 2

Cluster colors, sized by number of pixels:



cluster	pixels	name	HEX	RGB	HSV	LCH	Lab
	46.31%	134,131,122 friar grey ΔE=1.4	#8A867D	138 134 125	40 9 54	56 5 91	56 0 5
	23.28%	198,195,181 ash ΔE=1.1	#C8C4B9	200 196 185	43 8 79	79 6 94	79 0 6
	11.60%	74,66,55 double mondo ΔE=0.9	#494035	73 64 53	32 27 28	27 8 76	27 2 8
	8.24%	175,136,74 dark tan ΔE=1.9	#B08B4B	176 139 75	38 58 69	60 40 81	60 6 39
	8.18%	151,66,45 tia maria ΔE=3.3	#983F30	152 63 48	9 69 60	39 46 37	39 36 28

Study 3

Cluster colors, sized by number of pixels:



cluster	pixels	name	HEX	RGB	HSV	LCH	Lab
	24.82%	196, 196, 196 grey ΔE=0.6	#C4C4C5	196 196 197	245 1 77	79 1 293	79 0 -1
	13.59%	107, 106, 108 scarpa flow ΔE=1.1	#68686A	104 104 106	242 2 42	44 1 292	44 0 -1
	11.51%	102, 111, 180 chetwode blue ΔE=3.0	#5F70B3	95 112 179	228 47 70	48 40 288	48 12 -38
	9.22%	28, 57, 187 persian blue ΔE=4.7	#3735B2	55 53 178	241 70 70	31 77 302	31 41 -66
	8.89%	184, 184, 231 pale light grayish blue ΔE=3.2	#B6B4DD	182 180 221	243 19 87	75 22 294	75 9 -20
	7.11%	95, 160, 82 muted green ΔE=4.6	#589B46	88 155 70	107 55 61	58 53 135	58 -38 38
	6.77%	180, 70, 104 blush ΔE=5.2	#C55176	197 81 118	341 59 77	51 50 4	51 50 3
	6.22%	74, 156, 168 moderate arctic blue ΔE=2.6	#559BA6	85 155 165	187 49 65	60 23 212	60 -19 -12
	6.19%	63, 48, 127 minsk ΔE=4.0	#332878	51 40 120	248 66 47	22 53 303	22 29 -44
	5.67%	185, 184, 128 pea soup ΔE=3.0	#C3BF83	195 191 131	56 33 76	76 32 104	76 -8 31

Study 4

Cluster colors, sized by number of pixels:



cluster	pixels	name	HEX	RGB	HSV	LCH	Lab
	22.03%	82, 79, 70 touchstone ΔE=2.0	#4E4C41	78 76 65	51 17 30	32 7 101	32 -1 7
	17.80%	25, 24, 24 charcoal ΔE=1.6	#191816	25 24 22	41 10 10	8 1 91	8 0 1
	12.97%	143, 125, 107 squirrel ΔE=0.8	#92806E	146 128 110	31 24 57	55 13 74	55 3 12
	11.80%	113, 51, 60 ribbon ΔE=2.9	#793741	121 55 65	351 55 47	32 31 14	32 30 7
	10.88%	120, 118, 151 rhythm ΔE=4.2	#6A7292	106 114 146	227 27 57	49 18 283	49 4 -18
	6.52%	74, 59, 106 meteorite ΔE=3.8	#554070	85 64 112	266 43 44	31 32 309	31 20 -25
	5.44%	82, 139, 139 dark slate grey ΔE=2.9	#44898A	68 137 138	181 51 54	53 22 199	53 -21 -7
	4.91%	173, 53, 57 get reddy ΔE=2.0	#B3383A	179 56 58	359 69 70	42 56 28	42 50 27
	4.02%	196, 170, 77 sundance ΔE=2.4	#C2AD4D	194 173 77	49 60 76	71 51 94	71 -4 51
	3.64%	46, 45, 136 cosmic cobalt ΔE=4.1	#3C2E8C	60 46 140	249 67 55	26 60 303	26 33 -50

Study 5

Cluster colors, sized by number of pixels:



cluster	pixels	name	HEX	RGB	HSV	LCH	Lab
	22.57%	208, 201, 195 quarter cloudy ΔE=1.6	#D2CCC7	210 204 199	29 5 82	82 3 74	82 1 3
	17.37%	166, 162, 154 dawn ΔE=0.7	#A7A49B	167 164 155	46 7 66	68 5 98	68 -1 5
	15.63%	137, 124, 121 eighth felix ΔE=1.3	#877E7B	135 126 123	18 9 53	54 4 51	54 3 3
	12.51%	74, 75, 70 gravel ΔE=0.8	#4D4B46	77 75 70	47 9 30	32 4 98	32 -1 4
	6.52%	119, 139, 165 shadow blue ΔE=2.9	#818FA8	129 143 168	219 23 66	59 15 273	59 1 -15
	6.40%	188, 184, 138 sage ΔE=1.3	#BBB889	184 182 137	58 25 72	73 24 106	73 -7 23
	5.95%	88, 113, 86 cactus ΔE=2.5	#5B7253	91 114 83	106 27 45	46 21 136	46 -15 14
	5.10%	106, 81, 119 jupiter ΔE=3.1	#69537E	105 83 126	271 34 49	39 28 311	39 18 -21
	5.04%	168, 97, 74 moderate vermilion ΔE=3.8	#A96353	169 99 83	11 51 66	49 34 38	49 27 21
	2.90%	198, 168, 75 roti ΔE=3.9	#D1B14F	209 177 79	45 62 82	73 53 89	73 1 53

Study 6

Cluster colors, sized by number of pixels:



cluster	pixels	name	HEX	RGB	HSV	LCH	Lab
	20.54%	108, 99, 78 double stonewall ΔE=1.3	#6A6350	106 99 80	43 25 42	42 12 92	42 0 12
	15.84%	161, 152, 139 settlement ΔE=0.5	#A0978B	160 151 139	35 14 63	63 8 82	63 1 8
	14.34%	219, 208, 203 quarter milestone ΔE=2.0	#DBCDCB	219 205 203	9 8 86	84 5 34	84 4 3
	11.59%	41, 41, 36 jungle green ΔE=2.0	#25241F	37 36 31	51 15 14	14 3 102	14 -1 3
	10.46%	193, 160, 125 quarter sorrell brown ΔE=1.5	#C5A37E	197 163 126	31 36 77	69 25 73	69 7 24
	8.33%	192, 128, 129 old rose ΔE=3.5	#BE8480	190 132 128	4 33 75	61 24 28	61 22 11
	7.19%	150, 90, 62 coconut ΔE=2.1	#94583F	148 88 63	18 58 58	44 34 48	44 22 25
	6.20%	205, 93, 52 tangerine ΔE=2.8	#D2643B	210 100 59	17 72 82	55 59 47	55 40 43
	3.51%	119, 118, 150 rhythm ΔE=0.8	#797898	121 120 152	241 21 60	52 18 293	52 7 -17
	2.01%	207, 181, 59 old gold ΔE=3.8	#D5B746	213 183 70	47 67 84	75 59 91	75 -1 59

Study 7

Cluster colors, sized by number of pixels:



cluster	pixels	name	HEX	RGB	HSV	LCH	Lab
	15.97%	118,115,111 aluminium ΔE=0.9	#76736D	118 115 109	41 8 46	49 4 92	49 0 4
	15.21%	29,26,29 magentaish black ΔE=1.4	#1D1C1E	29 28 30	273 6 12	10 1 311	10 1 -1
	13.82%	81,78,74 winchester ΔE=0.7	#4F4E4A	79 78 74	48 6 31	33 2 99	33 0 2
	12.47%	55,48,33 birch ΔE=1.7	#353123	53 49 35	47 35 21	20 10 96	20 -1 10
	11.66%	64,81,105 fiord ΔE=3.6	#3C4962	60 73 98	219 39 39	31 17 276	31 2 -17
	10.40%	160,159,156 mountain mist ΔE=1.1	#9D9D9B	157 157 155	47 1 62	65 1 99	65 0 1
	7.87%	42,37,81 hip hop ΔE=1.5	#262450	38 36 80	242 54 31	17 30 298	17 14 -27
	5.40%	113,99,56 yellow metal ΔE=1.2	#706138	112 97 56	43 50 44	42 25 90	42 0 25
	3.66%	53,76,40 lush ΔE=4.4	#3A5429	58 84 41	97 51 33	33 29 131	33 -19 22
	3.53%	102,42,44 red devil ΔE=1.5	#64282C	100 40 44	356 60 39	25 30 21	25 28 11

Table 67: Descriptive statistics for variables included in regression analysis

	n	*Mean*	*St.D*	*Min*	*Max*
Brand type	182	.12	0.32	0	1
Relative Price	182	3.6	22	-53	63
Relative Pack Size	182	1.1	1.0	1	536
Real Share	182	5.8	6.1	0	40
Unique Colour	182	0.4	0.5	0	1
Pack orientation	182	0.6	0.5	0	1
P.Catg 1.	182	0.1	0.4	0	7
P.Catg 2.	182	0.1	0.3	0	1
P.Catg 3.	182	0.3	0.4	0	1
P.Catg 4.	182	0.1	0.3	0	1
P.Catg 5.	182	0.2	0.4	0	1
P.Catg 6.	182	0.1	0.3	0	1
Hedonic Catg.	182	0.3	0.5	0	1
Grocery Enviro.	182	0.5	0.5	0	1
Mass. Enviro.	182	0.4	0.5	0	1
VR Bias	182	19	66	-97	187
Fielding mode	182	0.2	0.3	0	1
Number of trips simulated	182	0.3	0.5	0	1

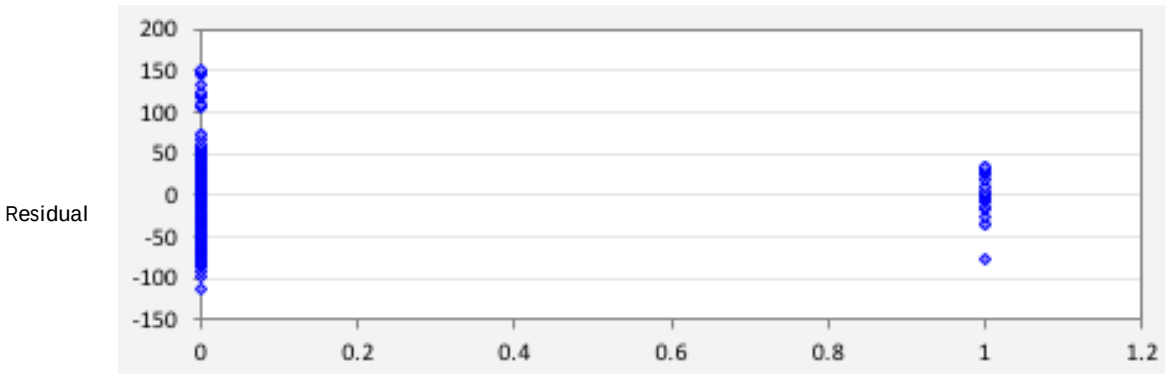
Table 68: Correlation matrix for all variables (*n* = 182)

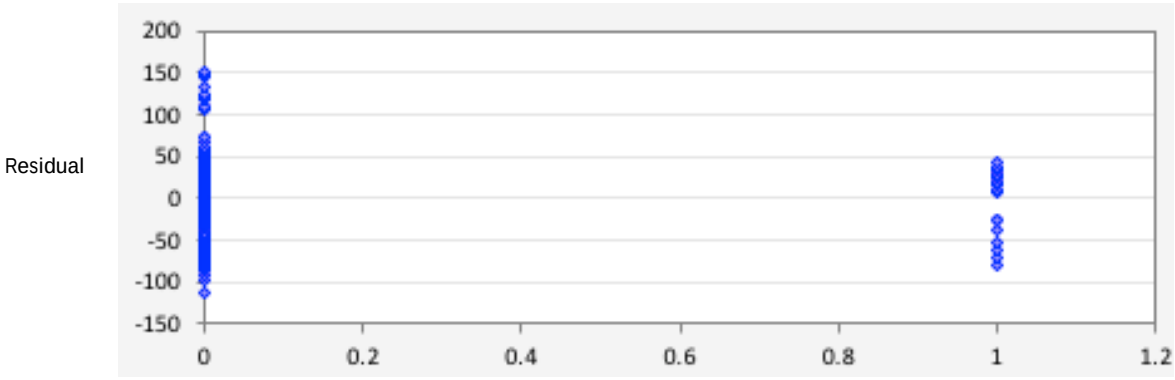
	Brand_Type	Fielding_mode	Grocery_Enviro.	Hedonic:Price	Mass_Enviro.	Number_of_trips	Pack_Orientation	PCatg1	PCatg2	PCatg3	PCatg4	PCatg5	PCatg6	Real_MS	Relative_Pack_Size	Relative_Price
Brand_Type	-0.39															
Fielding_mode	-0.02	1.00														
Grocery_Enviro.	0.21	-0.33	1.00													
Hedonic:Price	-0.20	-0.04	0.12	1.00												
Mass_Enviro.	-0.16	-0.17	-0.70	0.01	1.00											
Number_of_trips	0.05	-0.26	0.54	0.15	-0.29	1.00										
Pack_Orienta'n	0.01	-0.24	0.05	-0.01	0.14	0.01	1.00									
PCatg1	0.10	-0.15	0.46	0.23	-0.32	0.57	-0.13	1.00								
PCatg2	-0.06	-0.11	0.34	0.01	-0.24	0.43	0.19	-0.07	1.00							
PCatg3	-0.06	0.45	-0.54	-0.07	-0.03	-0.43	-0.24	-0.25	-0.18	1.00						
PCatg4	-0.02	0.06	-0.04	-0.04	0.16	0.41	-0.07	-0.17	-0.12	-0.24	1.00					
PCatg5	-0.15	-0.15	-0.39	0.04	0.56	-0.31	0.34	-0.18	-0.13	-0.25	-0.17	1.00				
PCatg6	0.27	-0.12	0.37	-0.04	-0.26	-0.25	0.15	-0.14	-0.11	-0.20	-0.14	-0.15	1.00			
Real_MS	-0.23	0.14	-0.12	-0.39	0.10	-0.06	-0.14	-0.19	-0.25	0.05	0.31	-0.16	0.02	1.00		
RelativePack_Size	0.23	-0.01	0.11	-0.27	-0.04	0.03	-0.14	-0.12	-0.23	-0.14	0.02	-0.08	0.40	0.05	1.00	
Relative_Price	0.52	-0.03	0.11	0.72	-0.08	0.72	0.05	0.13	-0.02	0.00	-0.03	-0.06	0.10	-0.15	0.07	1.00
Unique_Colour	0.18	0.03	0.17	0.04	-0.17	0.15	0.02	0.07	0.27	0.03	-0.10	-0.35	-0.01	-0.04	-0.04	0.06

Figure 31: Residual plots for IVs

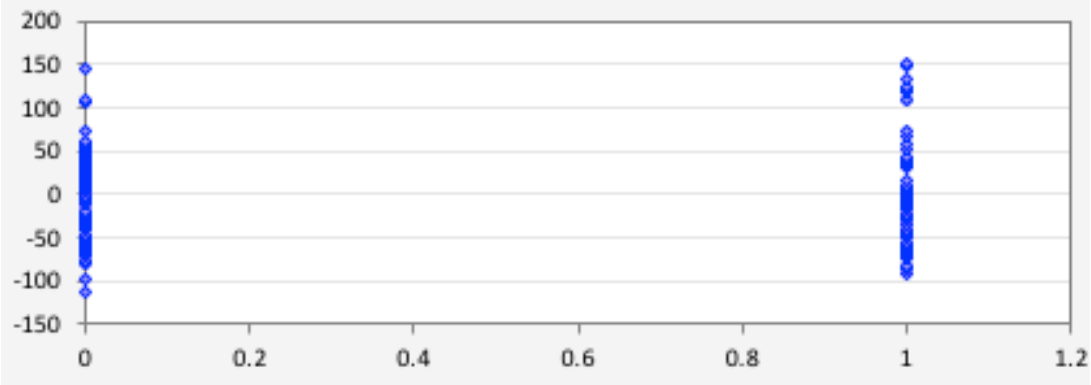
Predicted VR_Bias = 21.949 - 28.921*Brand_Type + 12.2*Fielding_mode + 10.937*Number_of_trips_simulated - 4.234*Pack_Orientation - 12.454*PCatg1 - 43.676*PCatg2 - 12.565*PCatg3 + 0.501*PCatg4 - 17.151*PCatg5 + 34.542*PCatg6 - 0.003562*Price_Hedonic - 2.182*Real_MS + 5.885*Relative_Pack_Size + 1.154*Relative_Price + 17.727*Unique_Colour

R^2 = .40 Variable adjusted R^2 = .36 Robust St.Error = 53 p < .001



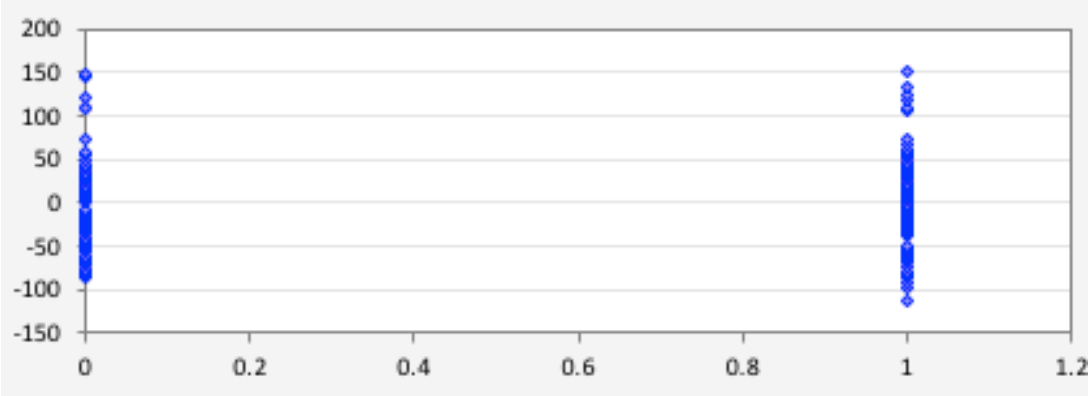


Residual



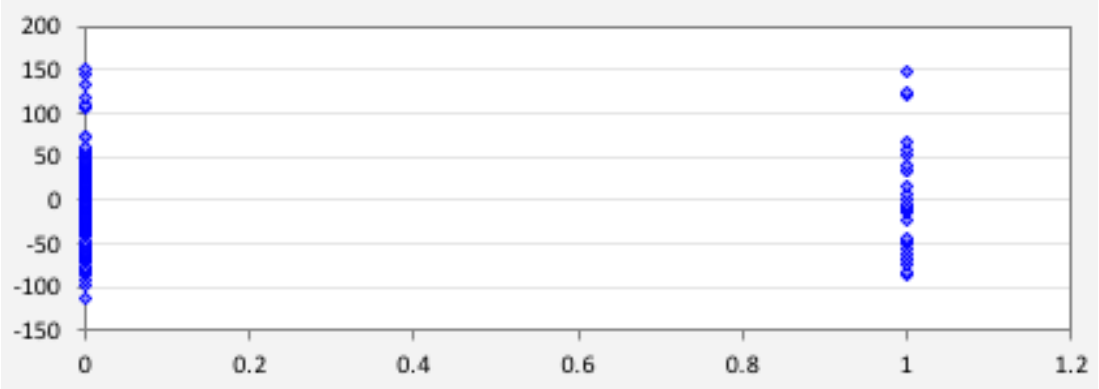
Number of trips simulated

Residual



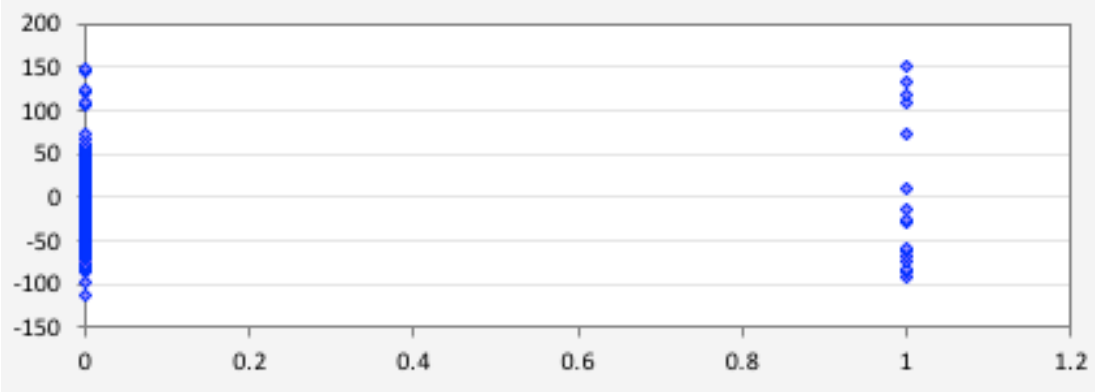
Pack orientation

Residual



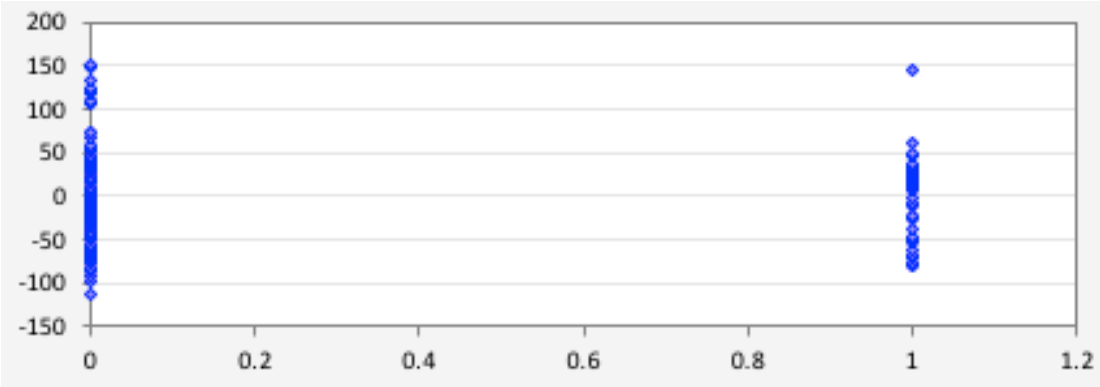
PCatg.1

Residual



PCatg.2

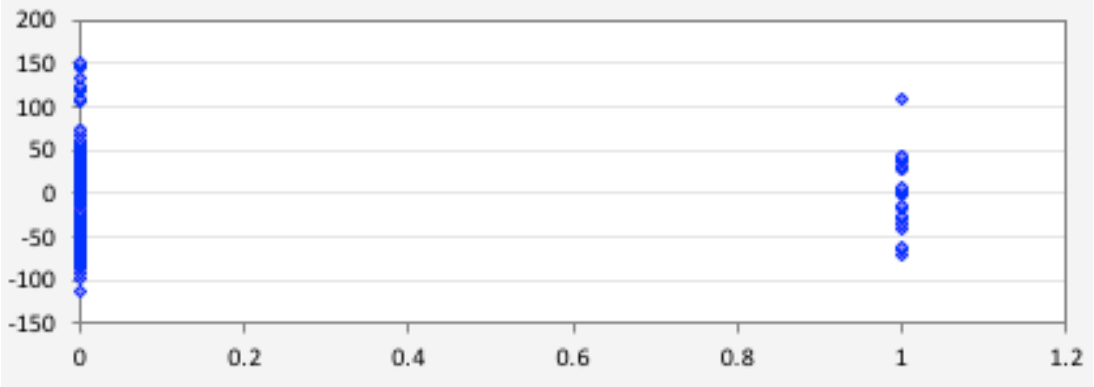
Residual



PCatg.3

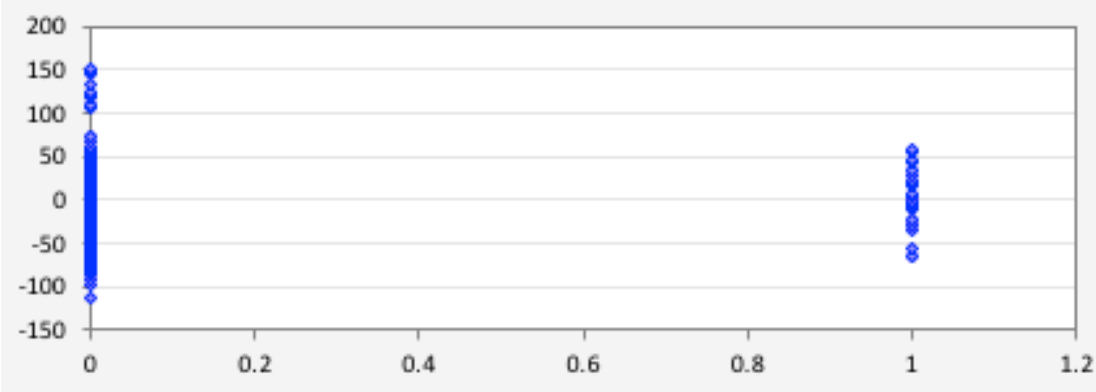
.

Residual



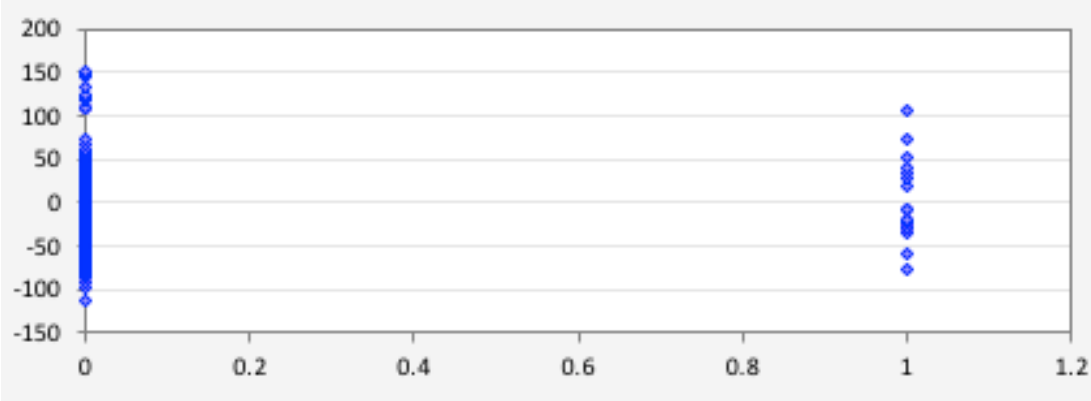
PCatg.4

Residual



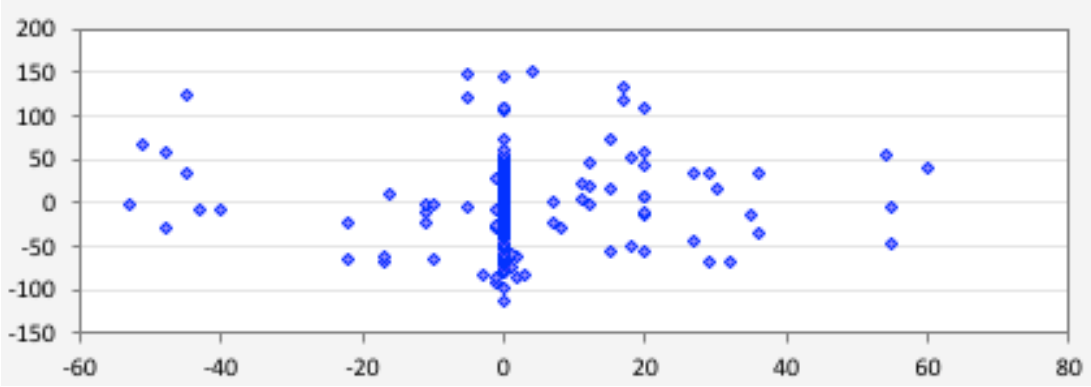
PCatg.5

Residual



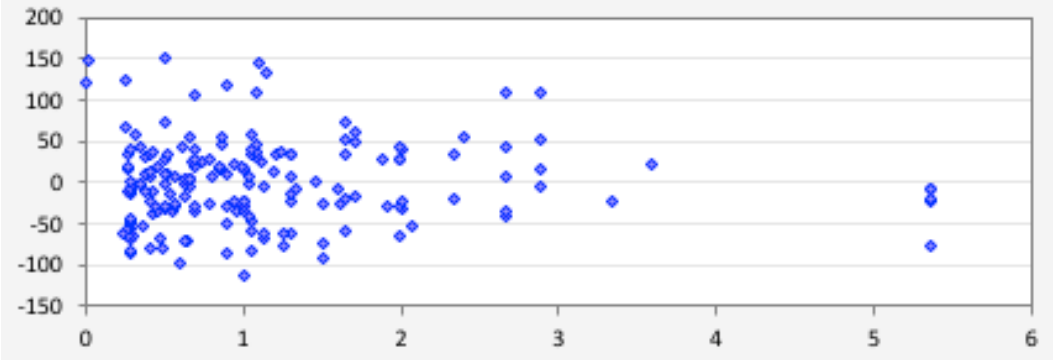
PCatg.6

Residual



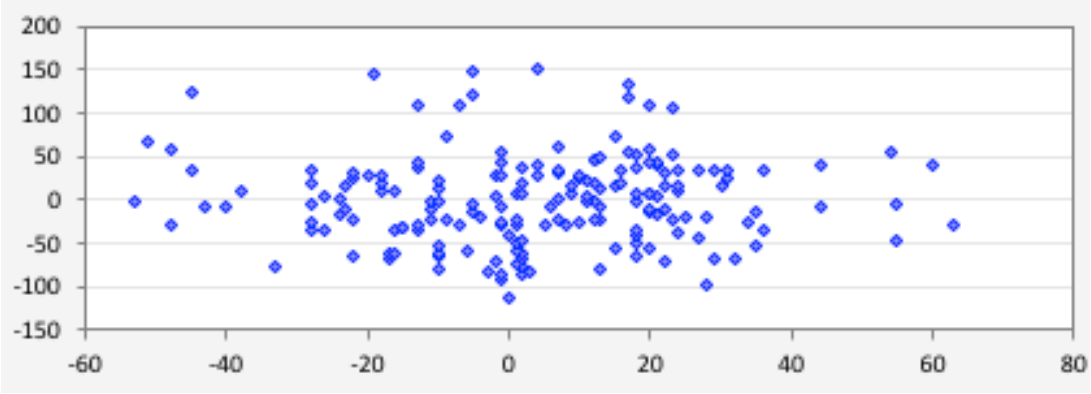
Price:Hedonic (Interaction variable)

Residual



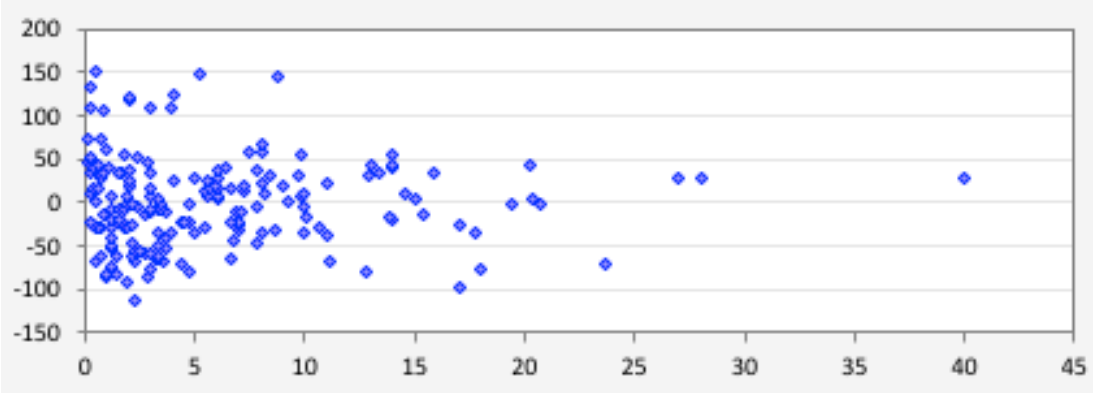
Relative pack size

Residual



Relative price

Residual



Real share

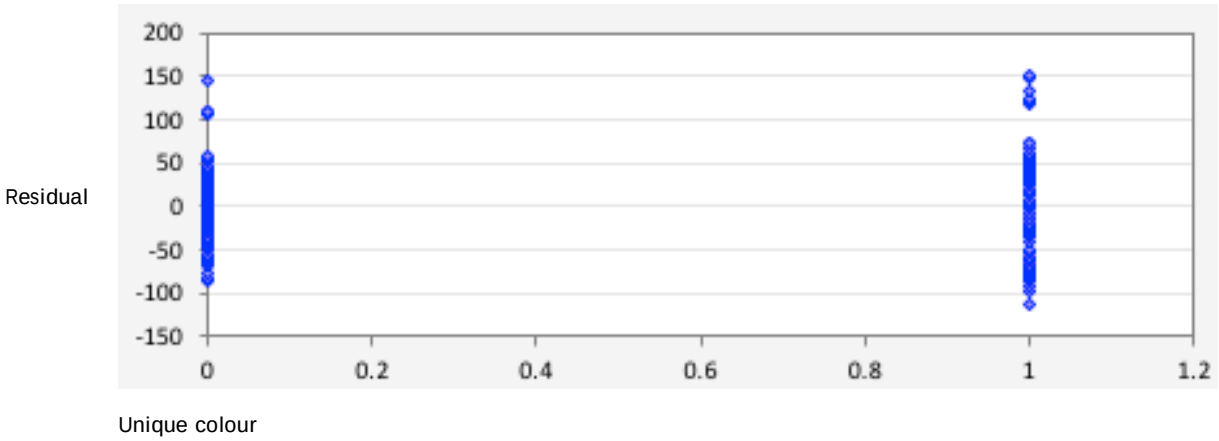


Table 69: Comparability of samples for shelf layout study reported in Chapter 9

Purchased last six mths (claimed):	**Layout 1** n = 84	**Layout 2** n = 68	**Layout 3** n = 72
Flavoured products	57.2%	56.9%	58.2%
Plain products	42.8%	45.3%	45.5%
Heat&Serve products	16.6%	16.8%	16.4%
Brand R	55.2%	62.0%	66.4%
Brand U	62.8%	56.2%	54.5%
Brand M	49.7%	54.0%	47.0%
Brand Z	46.9%	37.2%	44.8%
Brand K	27.6%	32.8%	23.9%
Brand G	17.9%	23.4%	21.6%
Brand S	13.1%	18.2%	15.7%
Brand PL	9.0%	12.4%	12.7%

REFERENCES

Aadland, D. and A. J. Caplan (2003). "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias." American Journal of Agricultural Economics 85(2): 492-502.

Abdul-Muhmin, A. G. (1999). "Contingent decision behavior: effect of number of alternatives to be selected on consumers' decision processes." Journal of Consumer Psychology 8(1): 91-111.

Advertising Research Foundation (2009). Knowledge Brief 1: Overlap, Duplication, and Multi Panel Membership. The Foundations of Quality Study, Advertising Research Foundation.

Aglioti, S., J. F. X. DeSouza and M. A. Goodale (1995). "Size-contrast illusions deceive the eye but not the hand." Current biology 5(6): 679-685.

Ailawadi, K. L., S. A. Neslin and K. Gedenk (2001). "Pursuing the Value-Conscious Consumer: Store Brands Versus National Brand Promotions." Journal of Marketing 65(January): 71-89.

Allenby, G., G. Fennell, J. Huber, T. Eagle, T. Gilbride, D. Horsky, J. Kim, P. Lenk, R. Johnson, E. Ofek, B. Orme, T. Otter and J. Walker (2005). "Adjusting choice models to better predict market behavior." Marketing Letters 16(3/4): 197-208.

Allsopp, J. (2005). "Additional practice papers: Premium pricing: Understanding the value of premium." Journal of Revenue and Pricing Management 4(2): 185-194.

Allsopp, J. (2005). "Premium pricing: Understanding the value of premium." Journal of Revenue and Pricing Management 4(2): 185-194.

Ampuero, O. and N. Vila (2006). "Consumer perceptions of product packaging." Journal of Consumer Marketing 23(2): 100-112.

Amrouche, N., T. B. Rhouma and G. Zaccour (2014). "Branding decisions for retailers' private labels." Journal of Marketing Channels 21(2): 100-115.

Anesbury, Z., M. Nenycz-Thiel, J. Dawes and R. Kennedy (2016). "How do shoppers behave online? An observational study of online grocery shopping." Journal of Consumer Behaviour 15(3): 261-270.

Anselmsson, J., U. Johansson and N. Persson (2007). "Understanding price premium for grocery products: a conceptual model of customer-based brand equity." Journal of Product & Brand Management 16(6): 401-414.

Atalay, A. S., H. O. Bodur and D. Rasolofoarison (2012). Shining in the center: Central gaze cascade effect on product choice. Journal of Consumer Research. 39: 848-866.

Baim, J., M. Galin, M. R. Frankel Risa Becker and J. Agresti (2009). Sample surveys based on internet panels: 8 years of learning. Worldwide Readership Research Symposium. Valencia: 1-14.

Baker, R., S. Blumberg, M. Brick, M. Couper, M. Courtright, M. Dennis, D. Dillman, M. Frankel, P. Garland, R. Groves, C. Kennedy, J. Krosnick, S. Lee, P. Lavrakas, M. Link, L. Piekarski, K. Rao, D. Rivers, R. Thomas and D. Zahs (2010). AAPOR Report on Online Panels: 81.

Baltas, G. and P. C. Argouslidis (2007). "Consumer characteristics and demand for store brands." International Journal of Retail & Distribution Management 35(5): 328-341.

Banelis, M., R. Kennedy, C. Driesener, A. Tanusondjaja and S. Lion (2011). Are partitions a law unto themselves and how would we know? Using the Duplication of Purchase law and behavioural data to identify market structure. Adelaide, Ehrenberg-Bass Institue.

Banelis, M., E. Riebe and C. Rungie (2013). "Empirical evidence of repertoire size." Australasian Marketing Journal 21(1): 59-65.

Bar-Ilan, J., M. Mat-Hassan and M. Levene (2006). "Methods for comparing rankings of search engine results." Computer Networks 50(10): 1448-1463.

Bass, F. M. (1974). "The Theory of Stochastic Preference and Brand Switching." Journal of Marketing Research 11(1): 1-20.

Bass, F. M. (1995). "Empirical Generalizations and Marketing Science: A Personal View." Marketing Science 14(3): G6-G18.

Baumgartner, H. and J.-B. E. M. Steenkamp (2001). "Response styles in marketing research: A cross national investigation." Journal of Marketing Research 38(May): 143-156.

Bell, D. R., D. Corsten and G. Knox (2011). "From point of purchase to path to purchase: How preshopping factors drive unplanned buying." Journal of Marketing 75(1): 31-45.

Bellizzi, J., H. Krueckeberg, J. Hamilton and W. Martin (1981). "Consumer Perceptions of National Private and Generic Brands." Journal of Retailing 57(4): 56-70.

Bennett, D. (2004). "The Taiwanese are just like Australians in their loyalty to fast food outlets." Australasian Marketing Journal 12(3): 97-103.

Bennett, D. and C. Graham (2010). "Is loyalty driving growth for the brand in front? A two-purchase analysis of car category dynamics in Thailand." Journal of Strategic Marketing 18(7): 573-585.

Berger, J., M. Draganska and I. Simonson (2007). "The influence of product variety on brand perception and choice." Marketing Science 26(4): 460-472.

Berneburg, A. (2007). "Interactive 3-D Simulations in Measuring Consumer Preferences: Friend or Foe to Test Results?" Journal of Interactive Advertising 8(1): 1-37.

Bettman, J. R. (1979). An information processing theory of consumer choice. Reading, Massachusetts, Addison-Wesley Publishing Company.

Bijmolt, T. H. A., H. J. van Heerde and R. G. M. Pieters (2005). "New empirical generalizations on the determinants of price elasticity." Journal of Marketing Research XLII(May): 141-156.

Bindemann, M. (2010). "Scene and screen center bias early eye movements in scene viewing." Vision Research 50(23): 2577-2587.

Blattberg, R. C., T. Buesing, P. Peacock and S. K. Sen (1978). "Identifying the deal prone segment." Journal of Marketing Research 15(3): 369-377.

Bock, D. and B. Treiber (2003). Studying New Product Performance in Virtually Created Retail Environments. ESOMAR Technovate Conference. Barcelona.

Bound, J. A. and A. Ehrenberg (1989). "Significant sameness." Journal of the Royal Statistical Society 152: 241-247.

Bowen, G. A. (2008). "Naturalistic inquiry and the saturation concept: a research note." Qualitative Research 8(1): 137-152.

Bowman, D. A. and R. P. McMahan (2007). "Virtual reality: How much immersion is enough?" Computer 40(7): 36-43.

Breen, P. (2009). Shaping Retail: The Use of Virtual Store Simulations in Marketing Research and Beyond. Industry White Paper.

Breetz, C. (2014). Product packaging as tool to demand a price premium: Does packaging enhance consumers‘ value perception to justify a price premium. Hamburg, Germany, Anchor Academic Publishing.

Bressoud, E. (2013). "Testing FMCG innovations: Experimental real store versus virtual." Journal of Product & Brand Management 22(4): 286-292.

Breugelmans, E. and K. Campo (2011). "Effectiveness of in-store displays in a virtual store environment." Journal of Retailing 87(1): 75-89.

Breugelmans, E. and K. Campo (2016). "Cross-Channel Effects of Price Promotions: An Empirical Analysis of the Multi-Channel Grocery Retail Sector." Journal of Retailing 92(3): 333-351.

Breugelmans, E., K. Campo and E. Gijsbrechts (2006). The effects of shelf display on online grocery choices, Katholieke Universiteit Leuven.

Bruwer, J. d. W., N. E. Haydam and B. Lin (1996). "Reducing bias in shopping mall-intercept surveys: The time-based systematic sampling method." South African Journal of Business Management 27(1 & 2): 9-16.

Burke, R. (1996). "Virtual Shopping: Breakthrough in Marketing Research." Harvard Business Review: 120-131.

Burke, R. R. (1998). Real Shopping in a Virtual Store. Sense & Respond: Capturing Value in the Network Era. S. P. Bradley and R. L. Nola, Harvard Business Press.

Burke, R. R. (2002). "Technology and the customer interface: what consumers want in the physical and virtual store." Journal of the Academy of Marketing Science 30(4): 411-432.

Burke, R. R., B. A. Harlam, B. E. Kahn and L. M. Lodish (1992). "Comparing Dynamic Consumer Choice in Real and Computer-Simulated Environments." Journal of Consumer Research 19(1): 71-82.

Burke, R. R. and A. Leykin (2014). Identifying the drivers of shopper attention, engagement, and purchase. Shopper marketing and the role of in-store marketing. D. Grewal, J. Nordfalt and A. L. Roggeveen. Online, Emerald Group Publishing Limited. 11: 147-187.

Bush, A. J. and J. F. Hair Jr (1985). "An assessment of the mall intercept as a data collection method." Journal of Marketing Research 22(2): 158-167.

Campo, K., E. Gijsbrechts and F. Guerra (1999). "Computer Simulated Shopping Experiments for Analyzing Dynamic Purchasing Patterns: Validation and Guidelines." Journal of Empirical Generalisations in Marketing Science 4: 22-61.

Carlsson, F. and P. Martinsson (2001). "Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment." Journal of Environmental Economics and Management 41(2): 179-192.

Carron, C. (2017, 8th May 2017). "Immersive virtual-reality creation software for everyone." Retrieved November 2018, from https://actu.epfl.ch/news/immersive-virtual-reality-creation-software-for-ev/.

Cashen, L. H. and S. W. Geiger (2004). "Statistical Power and the Testing of Null Hypothese: A Review of Contemporary Management Research and Recommendations for Future Studies." Organizational Research Methods 7(2): 151-167.

Chandon, P., J. W. Hutchinson, E. T. Bradlow and S. H. Young (2009). "Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase." Journal of Marketing 73(6): 1-17.

Chandon, P., V. G. Morwitz and W. J. Reinartz (2005). "Do intentions really predict behavior? Self-generated validity effects in survey research." Journal of Marketing 69(2): 1-14.

Chandon, P. and B. Wansink (2002). "When Are Stockpiled Products Consumed Faster? A Convenience--Salience Framework of Postpurchase Consumption Incidence and Quantity." Journal of Marketing Research 39(3): 321-335.

Chang, L. and J. Krosnick (2010). "Comparing oral interviewing with self-administered computerized questionnaires: An experiment." Public Opinion Quarterly 74(1): 154-167.

Charmaz, K. (2006). Constructing grounded theory: A practical guide through qualitative analysis. London, Sage Publications Ltd.

Chen, T., B. Sun and V. Singh (2009). "An Empirical Investigation of the Dynamic Effect of Marlboro's Permanent Pricing Shift." Marketing Science 28(4): 740-758.

Chu, J., M. Arce-Urriza, J.-J. Cebollada-Calvo and P. Chintagunta (2010). "An empirical analysis of shopping behavior across online and offline channels for grocery products: The moderating effects of household and product characteristics." Journal of Interactive Marketing 24: 251-268.

Chung, C., T. M. Schmit, D. Dong and H. M. Kaiser (2007). "Economic evaluation of shelf-space management in grocery stores." Agribusiness 23(4): 583-597.

Clement, J., J. Aastrup and S. C. Forsberg (2015). "Decisive visual saliency and consumers′ in-store decisions." Journal of Retailing and Consumer Services 22: 187-194.

Clement, J., T. Kristensen and K. Grønhaug (2013). "Understanding consumers' in-store visual perception: The influence of package design features on visual attention." Journal of Retailing and Consumer Services 20(2): 234-239.

Cobb, C. J. and W. D. Hoyer (1986). "Planned versus impulse purchase behavior." Journal of Retailing 62(4): 384-409.

Cohen, G. and C. A. Martin (2008). Memory in the real world. New York, Psychology Press.

Cohen, S. and M. Gadd (1996). "Virtual reality shopping simulation for the modern marketer." Marketing and Research Today February: 19-26.

Colombotos, J. (1969). "Personal versus telephone interviews: effect on responses." Public Health Reports 84(9): 773-782.

Corbin, J. and A. Strauss (2008). Basics of Qualitative Research. California, Sage Publications.

Corsten, D. and T. Gruen (2003). "Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks." International Journal of Retail & Distribution Management 31(12): 605-617.

Croft, M. (1998). Private label in Europe: A revolution in brand perception. London, Financial Times Business Ltd.

Davies, B. J., D. Kooijmanb and P. Warda (2003). "The sweet smell of success: olfaction in retailing." Journal of Marketing Management 19(5-6): 611-627.

Dawes, J. (2004). "Assessing the impact of a very successful price promotion on brand, category and competitor sales." Journal of Product & Brand Management 13(5): 303-314.

Dawes, J. G. (2012). "Brand-pack size cannibalization arising from temporary price promotions." Journal of Retailing 88(3): 343-355.

Dawes, J. G. (2016). "Testing the robustness of brand partitions identified from purchase duplication analysis." Journal of Marketing Management 32(7): 695-715.

Dekimpe, M. G., D. M. Hanssens, V. R. Nijs and J.-B. E. Steenkamp (2005). "Measuring short and long run promotional effectiveness on scanner data using persistence modelling." Applied Stochastic Models in Business and Industry 21(4 5): 409-416.

Desmet, P., R. Bordenave and J. Traynor (2013). "Differences in purchasing behaviour between physical and virtual laboratory stores." Research and Applications in Marketing 28(2): 71-86.

Dick, A., A. Jain and P. Richardson (1995). "Correlates of Store Brand Proneness: Some Empirical Observations." Journal of Product & Brand Management 4(4): 15-22.

Dickson, P. R. and A. G. Sawyer (1990). "The price knowledge and search of supermarket shoppers." Journal of Marketing 54(3): 42-53.

Drèze, X., S. J. Hoch and M. E. Purk (1994). "Shelf management and space elasticity." Journal of Retailing 70(4): 301-326.

Duchowski, A. (2007). Eye tracking methodology: Theory and practice, Springer-Verlag New York Inc.

Duffy, B., K. Smith, G. Terhanian and J. Bremer (2005). "Comparing data from online and face-to-face surveys." International Journal of Market Research 47(6): 615-639.

Dunne, D. and C. Narasimhan (1999). "The new appeal of private labels." Harvard Business Review 77(3): 41-52.

Dunning, T. (2012). Natural experiments in the social sciences: A design-based approach. New York, US, Cambridge University Press.

Eales, T. (2014). Private Label in Western Economies - Closing the price gap, losing share. Online, IRi Worldwide: 1-40.

Easley, R. W., C. S. Madden and M. G. Dunn (2000). "Conducting marketing science: the role of replication in the research process." Journal of Business Research 48(1): 83-92.

East, R., K. Hammond, P. Harris and W. Lomax (2000). "First-store loyalty and retention." Journal of Marketing Management 16(4): 307-325.

Eckersley, N. (2011) "Private Label: Key Threat to Big Name Brands." 3.

Economic Research Service (2008). On the Accuracy of Nielsen Homescan Data, United States Department of Agriculture.

Ehmke, M. D., J. L. Lusk and J. A. List (2008). "Is hypothetical bias a universal phenomenon? A multinational investigation." Land economics 84(3): 489-500.

Ehrenberg, A. (1991). "New Brands and the Existing Market." Journal of the Market Research Society 33(4): 285-299.

Ehrenberg, A. (1995). "Empirical generalisations, theory, and method." Marketing Science 14(3): G20-G28.

Ehrenberg, A. (1997). "Description and Prescription." Journal of Advertising Research 37: 17-22.

Ehrenberg, A. (2000). "Data reduction - Analysing and interpreting statistical data." Journal of Empirical Generalisations in Marketing Science 5: 1-391.

Ehrenberg, A. (2000). "Repeat-buying: facts, theory and applications." Journal of Empirical Generalisations in Marketing Science 5(2): 392-770.

Ehrenberg, A., N. Barnard and B. Sharp (2001). "Decision Models or Descriptive Models: A Brief Critique." Marketing Research 13(3): 34-37.

Ehrenberg, A. and J. A. Bound (1993). "Predictability and prediction." Journal of the Royal Statistical Society 156(2): 167-206.

Einav, L., E. Leibtag and A. Nevo (2010). "Recording discrepancies in Nielsen Homescan data: Are they present and do they matter?" Quantitative Marketing Economics 8(2): 207-239.

Fink, G. R., J. C. Marshall, P. W. Halligan, C. D. Frith, R. S. J. Frackowiak and R. J. Dolan (1997). "Hemispheric specialization for global and local processing: The effect of stimulus category." Proceedings of the Biological Sciences 264(1381): 487-494.

Fitzsimmons, G. J. and V. G. Morwitz (1996). "The effect of measuring intent on brand-level purchase behaviour." Journal of Consumer Research 23(1): 1-11.

Fitzsimons, G. J. (2008). Death to dichotomizing. United Kingdom, The Oxford University Press.

Fok, D., R. Paap and P. H. Franses (2012). "Modeling dynamic effects of promotion on interpurchase times." Computational Statistics & Data Analysis 56(11): 3055-3069.

Folkes, V. and S. Matta (2004). "The Effect of package shape on consumers' judgments of product volume: attention as a mental contaminant." Journal of Consumer Research 31(2): 390-401.

Fox, E. J., A. L. Montgomery and L. M. Lodish (2004). "Consumer shopping and spending across retail formats." Journal of Business 77(2): 25-60.

Gabor, A., C. Granger and A. Sowter (1970). "Real and hypothetical shop situations in market research." Journal of Marketing Research 7(November): 355-359.

Gadalla, E., K. Keeling and I. Abosag (2013). "Metaverse-retail service quality: A future framework for retail service quality in the 3D internet." Journal of Marketing Management 29(13-14): 1493-1517.

Garber, L. L., R. R. Burke and J. M. Jones (2000). The role of package color in consumer purchase consideration and choice. Cambridge, Massuchusetts, Marketing Science Institute.

Garber, L. L., E. M. Hyatt and Ü. Ö. Boya (2009). "The effect of package shape on apparent volume: an exploratory study with implications for package design." The Journal of Marketing Theory and Practice 17(3): 215-234.

Geoff, C. (2011). Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis (Multivariate Applications Series) United Kingdom Routledge Publishing.

Gidlöf, K., N. Holmberg and H. Sandberg (2012). "The use of eye-tracking and retrospective interviews to study teenagers' exposure to online advertising." Visual Communication 11(3): 329-345.

González-Benito, Ó., Z. I. Loyola-Galván and P. A. Muñoz-Gallego (2010). "Inter-size and inter-brand competition analysis within a product category: scope of cannibalization effects." The Journal of Brand Management 17(4): 254-265.

Göritz, A. S. (2006). "Cash lotteries as incentives in online panels." Social Science Computer Review 24(4): 445-459.

Gosling, S. D., S. Vazire, S. Srivastava and O. P. John (2004). "Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires." American Psychologist 59(2): 93-104.

Graham, F. K. (1997). Afterword: Pre-attentive processing and passive and active attention. Attention and orienting: Sensory and motivational processes. P. J. Lang, F. P. Simons and M. Balaban. Erlbaum Hillsdale, NJ, Erlbaum: 417-452.

Granger, C. W. J. and A. Billson (1972). "Consumers' attitudes toward package size and price " Journal of Marketing Research 9(3): 239-248.

GreenBook (2015). GreenBook research industry trends report. L. F. Murphy. Online: 1-88.

Griffith, R., E. Leibtag, A. Leicester and A. Nevo (2009). "Consumer shopping behavior: how much do consumers save?" Journal of Economic Perspectives 23(2): 99-120.

Grinberg, A. M., J. S. Careaga, M. R. Mehl and M.-F. O'Connor (2014). "Social engagement and user immersion in a socially based virtual world." Computers in Human Behavior 36(1): 479-486.

Groeppel-Klein, A. and B. Bartmann (2008). "Anti-Clockwise or Clockwise? The Impact of Store Layout on the Process of Orientation in a Discount Store." European Advances in Consumer Research 8: 415-416.

Grossman, R. P. and J. Z. Wisenblit (1999). "What we know about consumers' color choices." Journal of Marketing Practice: Applied Marketing Science 5(3): 78-88.

Groves, R. M. and L. Lyberg (2010). "Total Survey Error: Past, Present, and Future." Public Opinion Quarterly 74(5): 849-879.

Gupta, S. (1991). "Stochastic models of interpurchase time with time-dependent covariates." Journal of Marketing Research 28(1): 1-15.

Hall, L., P. Johansson, B. Tärning, S. Sikström and T. Deutgen (2010). "Magic at the marketplace: Choice blindness for the taste of jam and the smell of tea." Cognition: 8.

Hansen, S. S. (2013). "Exploring real-brand meanings and goods in virtual-world social interaction: Enhanced rewards, rarity, and realism." Journal of Marketing Management 29(13-14): 1443-1461.

Harwood, T. and M. Jones (2011). Using eye-gaze visual technologies to compare consumer response in real and 3D virtual worlds: An exploratory application to retail. Academy of Marketing Annual Conference. Liverpool, UK.

Harwood, T., M. Jones and A. Carreras (2013). Shedding light on retail environments. Eye Tracking South Africa, Cape Town, ACM.

Hawkes, W. J. and F. W. Piotrowski (2003). Using scanner data to improve the quality of measurement in the consumer price index. Scanner data and price indexes. R. C. Feenstra and M. D. Sharpiro. US, University of Chicago Press: 17-38.

Hendel, I. and A. Nevo (2006). "Measuring the implications of sales and consumer inventory behavior." Econometrica 74(6): 1637-1673.

Hendrickson, K. and K. L. Ailawadi (2014). "Six lessons for in-store marketing from six years of mobile eye-tracking research." Shopper Marketing and the Role of In-Store Marketing 11: 57-74.

Hensher, D. A. (2010). "Hypothetical bias, choice experiments and willingness to pay." Transportation Research Part B 44(6): 735-752.

Higaki, A. (2004). Product launch decision testing and demand forecasting using virtual store shelves. ESOMAR. Shanghai, China.

Hoek, J., Z. Kearns and K. Wilkinson (2003). "A New Brand's Behaviour in an Established Market." Journal of Product & Brand Management 12(1): 52-65.

Hoffman, H. G., J. Prothero, M. J. Wells and J. Groen (1998). "Virtual chess: Meaning enhances users' sense of presence in virtual environments." International Journal of Human-Computer Interaction 10(3): 251-263.

Holbrook, A. L., M. C. Green and J. A. Krosnick (2003). "Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias." Public Opin Quarterly 67(1): 79-125.

Holman, L., M. L. Head, R. Lanfear and M. D. Jennions. (2015, 8th July 2015). "Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording." Retrieved 26th May 2017, from http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002190.

Hornik, J. and S. Ellis (1988). "Strategies to secure compliance for a mall intercept interview." Public Opinion Quarterly 52(4): 539-551.

Hoyer, W. D. (1984). "An examination of consumer decision making for a common repeat purchase product." Journal of Consumer Research 11(3): 822-829.

Hubbard, C. W. (1969). "The “shelving” of increased sales." Journal of Retailing 45(4): 1969-1970.

Hubbard, R. and J. S. Armstrong (1994). "Replications and Extensions in Marketing: Rarely Published but Quite Contrary." International Journal of Research in Marketing 11(3): 233-248.

Hubbard, R., R. J. Brodie and J. S. Armstrong (1992). Knowledge Development in Marketing: The Role of Replication Research. Proceedings of the Sixth New Zealand Marketing Educators' Conference, Dunedin, Dept of Marketing, Otago University.

Hubbard, R. and R. M. Lindsay (2012). "From significant difference to significant sameness: Proposing a paradigm shift in business research." Journal of Business Research 66: 1377-1388.

Huberty, C. and J. Morris (1989). "Multivariate analysis versus multiple univariate analyses." Psychological Bulletin 105(2): 302-308.

Hui, P. S. Fader and E. T. Bradlow (2009). "The Traveling Salesman Goes Shopping: The Systematic Deviations of Grocery Paths from TSP Optimality." Marketing Science 28(3): 566-572.

Hunter, J. E. (2001). "The Desperate Need for Replications." Journal of Consumer Research 28: 149-158.

Interrante, V., B. Ries, J. Lindquist, M. Kaeding and L. Anderson (2008). "Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments." Presence: Teleoperators and Virtual Environments 17(2): 176-198.

Irwin, J. R. and G. H. McClelland (2003). "Negative consequences of dichotomizing continuous predictor variables." Journal of Marketing Research 40(3): 366-371.

Itti, L., G. Rees and J. Tsotsos (2005). "Models of bottom-up attention and saliency." Neurobiology of attention 582: 1-11.

Iyengar, S. S. and M. R. Lepper (2000). "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?" Journal of Personality and Social Psychology 79(6): 995-1006.

Janiszewski, C. (1998). "The influence of display characteristics on visual exploratory search behavior." Journal of Consumer Research 25(3): 290-301.

Jarvis, W., C. Rungie, S. Goodman and L. Lockshin (2006). "Using polarisation to identify variations in behavioural loyalty to price tiers." Journal of Product & Brand Management 15(4): 257-264.

Kahn, B. E. and B. Wansink (2004). "The influence of assortment structure on perceived variety and consumption quantities." Journal of Consumer Research 30(4): 519-533.

Kennedy, R. and B. McColl (2012). "Brand growth at Mars, Inc.: How the global marketer embraced Ehrenberg's science with creativity." Journal of Advertising Research 52(2): 270-276.

Kennedy, R. and J. Singh (2002). "A Case for Replication: Fitting Product Variants Data to the Dirichlet Model." Asia Pacific Advances in Consumer Research 5: 374-378.

Kholod, M., K. Takai and K. Yada (2011). "Clockwise and anti-clockwise directions of customer orientation in a supermarket: evidence from RFID data." Knowledge-Based and Intelligent Information and Engineering Systems 6883: 304-309.

Kim, A. E., J. M. Nonnemaker, B. R. Loomis, P. R. Shafer, A. Shaikh, E. Hill, J. W. Holloway and M. C. Farrelly (2014). "Influence of point-of-sale tobacco displays and graphic health warning signs on adults: evidence from a virtual store experimental study." American journal of public health 104(5): 888-895.

King, M. F. and G. C. Bruner (2000). "Social Desirability Bias: A Neglected Aspect of Validity Testing." Psychology & Marketing 17(2): 79-103.

Klein, R. A., K. A. Ratliff, M. Vianello, R. B. Adams Jr, Š. Bahník, M. J. Bernstein, K. Bocian, M. J. Brandt, B. Brooks and C. C. Brumbaugh (2014). "Investigating variation in replicability: A "many labs" replication project." Social psychology 45(3): 142.

Kleinschmit, M., A. Reid and R. Rizzo (2012). Virtual future real life or fantasy: Innovations in virtual testing technologies and applications. ESOMAR. Atlanta, USA.

Klink, R. R. and D. C. Smith (2001). "Threats to the external validity of brand extension research." Journal of Marketing Research 38(3): 326-335.

Koelemeijer, K. and H. Oppewal (1999). "Assessing the effects of assortment and ambience: A choice experimental approach." Journal of Retailing 75(3): 319-345.

Kopf, T. and B. Treiber (2003). The Role of Virtual Shopper Research in Category Management. ESOMAR. Lausanne, Switzerland.

Krosnick, J. A. (1999). "Survey Research." Annual Review of Psychology 50(1): 537-567.

Krysan, M., H. Schuman, L. J. Scott and P. Beatty (1994). "Response rates and response content in mail versus face-to-face surveys." The Public Opinion Quarterly 58(3): 381-399.

Kumar, N. and J. Steenkamp (2007) "Premium Store Brands: The Hottest Trend in Retailing."

Larson, J. S., E. T. Bradlow and P. S. Fader (2005). "An exploratory look at supermarket shopping paths." International Journal of Research in Marketing 22(4): 395-414.

Le Boutillier, J., S. S. Le Boutillier and S. A. Neslin (1994). "A replication and extension of the Dickson and Sawyer price-awareness study." Marketing Letters 5(1): 31-42.

Lee, S., S.-J. Kim, O.-B. Kwon, J. H. Lee and J.-H. Kim (2013). "Inhibitory networks of the amygdala for emotional memory." Frontiers in Neural Circuits 7(129): 1-10.

Leister, W., I. Tjøstheim and J. Lous (2007). Market Research Using a Virtual Test Store on Gaming Technology. SimVis, Germany.

Li, H. R., T. Daugherty and F. Biocca (2003). "The role of virtual experience in consumer learning." Journal of Consumer Psychology 13(4): 395-407.

Lindberg, S., A. Lindström, C. Cederström, A. From and C. Westerlind (2013). Tue 5.1 shoppers' attention to packaging and in-store media. IAPRI Symposium on Packaging. Stockholm, Sweden: 73-80.

Lindsay, R. M. and A. Ehrenberg (1993). "The design of replicated studies." The American Statistician 47(3): 217-228.

Lingonblad, M. (2015). Lost in the Supermarket: Studying the effects of familiarity on consumer decision making. Masters, Lund University.

Litvack, D. S., R. J. Calantone and P. R. Warshaw (1985). "An Examination of Short-Term Retail Grocery Price Effects." Journal of Retailing 61(3): 9-25.

Lombard, M. and T. Ditton (1997). "At the heart of it all: The concept of presence." Journal of Computer-Mediated Communication 3(2): 0.

Ludwichowska, G., J. Romaniuk and M. Nenycz-Thiel (2017). "Systematic response errors in self-reported category buying frequencies." European Journal of Marketing: 1-32.

Lusk, J. L. and T. C. Schroeder (2004). "Are choice experiments incentive compatible? A test with quality differentiated beef steaks." American Journal of Agricultural Economics 86(2): 467-482.

Malhotra, N., M. R. Shaw, J. Hall and M. Crisp (1996). Marketing Research: An Applied Orientation. Sydney, Prentice Hall.

Manganari, E. E., G. J. Siomkos, I. D. Rigopoulou and A. P. Vrechopoulos (2011). "Virtual store layout effects on consumer behaviour: Applying an environmental psychology approach in the online travel industry." Internet Research 21(3): 326-346.

Mania, K. and A. Robinson (2004). The effect of quality of rendering on user lighting impressions and presence in virtual environments. ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry, ACM.

Mannan, S. K., K. H. Ruddock and D. S. Wooding (1997). "Fixation sequences made during visual examination of briefly presented 2D images." Spatial Vision 11(2): 157-178.

Martin, N. and K. Morich (2011). "Unconscious mental processes in consumer choice: Toward a new model of consumer behavior." Journal of Brand Management 18(7): 483-505.

Massara, F., D. Porcheddu and R. D. Melara (2014). "Asymmetric perception of sparse shelves in retail displays." Journal of Retailing 90(3): 321-331.

Mauro, R. (1990). "Understanding LOVE (left out variables error): A method for estimating the effects of omitted variables." Psychological Bulletin 108(2): 314-329.

McCort, D. J. and N. Malhotra (1993). "Culture and consumer behaviour: Toward an understanding of cross-cultural consumer behaviour in international marketing." Journal of International Consumer Marketing 6: 91-128.

McFadden, D. (1980). "Econometric Models for Probabilistic Choice Among Products." Journal of Business 53(No. 3, Pt. 2): S13-S29.

Mendenhall, W. and T. Sincich (1996). A Second Course in Statistics: Regression Analysis. New Jersey, Prentice-Hall International, Inc.

Mitchell, V.-W., D. Lennard and P. McGoldrick (2003). "Consumer awareness, understanding and usage of unit pricing." British Journal of Management 14(2): 173-187.

Mocanu, A., L. Lockshin and S. Bogomolova (2012). Objective measurement of prototypicality in packaging design. ANZMAC, Adelaide, Australian and New Zealand Marketing Academy.

Mormann, M. M., V. Navalpakkam, C. Koch and A. Rangel (2012). "Relative visual saliency differences induce sizable bias in consumer choice." Journal of Consumer Psychology 22(1): 67-74.

Morwitz, V. G. (2001). Methods for forecasting from intentions data. Principles of Forecasting - A Handbook for Researchers and Practitioners. J. S. Armstrong. Boston, Kluwer Academic Publishers: 33-55.

Morwitz, V. G., E. Johnson and D. Schmittlein (1993). "Does measuring intent change behavior?" Journal of Consumer Research 20(1): 46-61.

Morwitz, V. G. and D. Schmittlein (1992). "Using Segmentation to Improve Sales Forecasts Based on Purchase Intent: Which "Intenders" Actually Buy?" Journal of Marketing Research 29(November): 391-405.

Morwitz, V. G., J. H. Steckel and A. Gupta (2007). "When do purchase intentions predict sales?" International Journal of Forecasting 23(3): 347-364.

Mueller Loose, S. and L. Lockshin (2013). "Testing the robustness of best worst scaling for cross-national segmentation with different numbers of choice sets." Food Quality and Preference 27(2): 230-242.

Mueller Loose, S., P. Osidacz, L. Francis and L. Lockshin (2010). The relative importance of extrinsic and intrinsic wine attributes. AWBR 2010 5th International Conference of the Academy of Wine Business Research. New Zealand.

Mueller, S., L. Lockshin, Y. Saltman and J. Blanford (2010). "Message on a bottle: The relative influence of wine back label information on wine choice." Food Quality and Preference 21(1): 22-32.

Mueller, S. and W. Umberger (2009). "What drives the Australian cask wine consumer?" Australian & New Zealand Wine Industry Journal 24(2): 44-47.

Needel, S. (2005). Can you reduce product assortment? Of course you can! ESOMAR: Retail Conference. Budapest, ESOMAR.

Needel, S. (2013). 4D Shopper - 4D Online: Validating Virtual Reality Technologies, Advanced Simulations.

Needel, S. P. (2013). Pricing research using virtual reality. Kennesaw, GA, Advanced Simulations: 1-9.

Needel, S. P. and B. Bean (2002). Advances in Category Management Research. ESOMAR. Sao Paulo.

Nelder, J. A. (1986). "Statistics, science and technology." Journal of the Royal Statistical Society 149(2): 109-121.

Nelder, J. A. (1999). "From statistics to statistical science." Journal of the Royal Statistical Society: Series D (The Statistician) 48(2): 257-269.

Nenycz-Thiel, M., V. Beal, G. Ludwichowska and J. Romaniuk (2012). "Investigating the accuracy of self-reports of brand usage behavior." Journal of Business Research 66(2): 224-232.

Nenycz-Thiel, M. and J. Romaniuk (2011). "The nature and incidence of private label rejection." Australasian Marketing Journal 19: 93-99.

Nenycz-Thiel, M. and J. Romaniuk (2014). "The real difference between consumers' perceptions of private labels and national brands." Journal of Consumer Behaviour 13(4): 262-269.

Nenycz-Thiel, M. and J. Romaniuk (2016). "Understanding premium private labels: A consumer categorisation approach." Journal of Retailing and Consumer Services 29: 22-30.

Neslin, S., G. Allenby, A. Ehrenberg, S. Hoch, G. Laurent, R. Leone, J. Little, L. Lodish, R. Shoemaker and D. Wittink (1994). "A research agenda for making scanner data more useful to managers." Marketing Letters 5(4): 395-412.

Nevin, J. R. (1974). "Laboratory Experiments for Estimating Consumer Demand: A Validation Study." Journal of Marketing Research 11(August): 261 -268.

Nichols, A. L. and J. K. Maner (2008). "The good-subject effect: Investigating participant demand characteristics." The Journal of general psychology 135(2): 151-166.

Nielsen. (2009). "Almost half of U.S. supermarket purchases are sold on promotion." Retrieved August 2016, from nielsen.com/us/en/insights/news/2009/almost-half-of-u-s-supermarket-purchases-are-sold-on-promotion.html.

Nielsen. (2014, March 03). "A battle of the sexes plays out in shopping aisles." from http://www.nielsen.com/hk/en/insights/news/2014/a-battle-of-the-sexes-plays-out-in-shopping-aisles.html.

Nielsen (2014). The state of private label around the world - Where it's growing, where it's not, and what the future holds. Online, The Nielsen Company: 1-35.

Nielsen (2015). The future of grocery: E-commerce, digital technology and changing shopping preferences around the world. Online, The Nielsen Company: 1-35.

Nielsen. (2016, 15th December 2017). "Deeper than dollars: global perceptions about premium products " Retrieved 3rd February 2017 from http://www.nielsen.com/us/en/insights/news/2016/deeper-than-dollars-global-perceptions-about-premium-products.html.

Norwood, F. B. and J. L. Lusk (2011). "Social desirability bias in real, hypothetical, and inferred valuation experiments." American Journal of Agricultural Economics 93(2): 528-534.

November, P. (2004). "Seven Reasons Why Marketing Practitioners Should Ignore Marketing Academic Research." Australasian Marketing Journal 12(2): 39-49.

Objective Digital (2014). A review of Australia's mobile grocery apps: What to consider when creating the next generation grocery app. Sydney, NSW, Objective Digital Pty Ltd: 1-26.

Ordabayeva, N. and P. Chandon (2013). "Predicting and managing consumers' package size impressions." Journal of Marketing 77(5): 123-137.

Orquin, J. L., M. Bagger and S. Mueller Loose (2013). "Learning affects top down and bottom up modulation of eye movements in decision making." Judgment and Decision Making 8(6): 700-716.

Orquin, J. L. and S. M. Loose (2013). "Attention and choice: A review on eye movements in decision making." Acta Psychologica 144(1): 190-206.

Orquin, J. L. and J. Scholderer (2011). "Attention to health cues on product packages." Journal of Eyetracking, Visual Cognition and Emotion 1(1): 59-63.

Orquin, J. L., J. Scholderer and H. B. Jeppesen (2012). What you see is what you buy: how saliency and surface size of packaging elements affect attention and choice. Society for Advancement of Behavioural Economics. Granada, Spain.

Palumbo, F. and P. Herbig (2000). "The multicultural context of brand loyalty." European Journal of Innovation Management 3(3): 116-125.

Papagiannidis, S., E. Pantano, E. W. K. See-To and M. Bourlakis (2013). "Modelling the determinants of a simulated experience in a virtual retail store and users' product purchasing intentions." Journal of Marketing Management 29(13-14): 1462-1492.

Park, C. W. and V. P. Lessig (1981). "Familiarity and its impact on consumer decision biases and heuristics." Journal of Consumer Research 8(2): 223-230.

Parkhurst, D., K. Law and E. Niebur (2002). "Modeling the role of salience in the allocation of overt visual attention." Vision Research 42(1): 107-123.

Parkhurst, D. J. and E. Niebur (2003). "Scene content selected by active vision." Spatial Vision 16(2): 125-154.

Paulhus, D. L. (1991). Measurement and control of response bias. Measures of personality and social psychological attitudes. J. P. Robinson, P. R. Shaver and L. S. Wright. San Diego, CA, Academic Press: 17-59.

Perception Research Services (2014) "How does PRS eye-tracking compare to "click on what you saw" & VAS projections?", 1-6.

Peschel, A. O. and J. L. Orquin (2013). "A review of the findings and theories on surface size effects on visual attention." Frontiers in psychology 4(902): 1-10.

Pieters, R. and L. Warlop (1997). Visual attention during brand choice: An eye-fixation analysis. DTEW Research Report 9724: 1-41.

Pieters, R. and L. Warlop (1999). "Visual attention during brand choice: The impact of time pressure and task motivation." International Journal of Research in Marketing 16(1): 1-16.

Porcheddu, D. and A. Venturi (2011). "Choices from identical options in a virtual shopping aisle." The Open Business Journal 4: 36-45.

Precourt, G. (2012). From consumer-focused to shopper-focused packaging: Nestlé realigns its Gerber brand using virtual research. Warc, Event Reports.

Presser, S. and L. Stinson (1998). "Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance." American Sociological Review 63(1): 137-145.

Rao, A. R. and K. B. Monroe (1996). "Causes and consequences of price premiums." The Journal of Business 69(4): 511-535.

Reutskaja, E., R. Nagel, C. F. Camerer and A. Rangel (2011). "Search dynamics in consumer choice under time pressure: An eye-tracking study." The American Economic Review 101(2): 900-926.

Robertson, L. C. and M. R. Lamb (1991). "Neuropsychological contributions to theories of part/whole organization." Cognitive psychology 23(2): 299-330.

Romaniuk, J. (2011). "Are you blinded by the heavy (buyers)...Or are you seeing the light?" Journal of Advertising Research 51(4): 561-563.

Romaniuk, J. (2016). Leveraging distinctive assets. How Brands Grow: Part 2. J. Romaniuk and B. Sharp, Oxford University Press: 87-108.

Romaniuk, J. and J. Dawes (2005). "Loyalty to price tiers in purchases of bottled wine." Journal of Product & Brand Management 14(1): 57-64.

Romaniuk, J. and B. Sharp (2016). How brands grow: Part 2. South Melbourne, Oxford University Press.

Rosenberger, P. and L. Chernatony (1995). "Virtual Reality Techniques in NPD Research." International Journal of Market Research 37(4).

Roupé, M., P. Bosch-Sijtsema and M. Johansson (2014). "Interactive navigation interface for virtual reality using the human body." Computers, Environment and Urban Systems 43: 42-50.

Rungie, C. and G. Laurent (2012). Brand loyalty vs. loyalty to product attributes. Quantitative Marketing and Marketing Management. A. Diamantopoulos, Springer: 423-444.

Russo, J. E. and F. Leclerc (1994). "An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables." Journal of Consumer Research 21(2): 274-290.

Sawyer, A. (1975). "Demand Artifacts in Laboratory Experiments in Consumer Research." Journal of Consumer Research 1(March): 20-30.

Sawyer, A. G. (1982). "Statistical power and effect size in consumer research." Advances in Consumer Research 9(1): 1-7.

Sawyer, A. G., P. M. Worthing and P. E. Sendak (1979). "The role of laboratory experiments to test marketing strategies." Journal of Marketing 43: 60-67.

Scamell-Katz, S. (2012). The Art of Shopping: How We Shop and Why We Buy. UK, LID Publishing.

Scaraboto, D., L. Carter-Schneider and R. Kedzior (2013). "At world's end: Exploring consumer-marketer tensions in the closure of adverworlds." Journal of Marketing Management 29(13-14): 1518-1541.

Schnak, A., J. Holdershaw and M. John Wright (2016). Virtual tracking of consumer behaviour: Pilot-testing of ViCoS 1.0, the Future of Market Research. ANZMAC. Christchurch, New Zealand, ANZMAC.

Schuemie, M. J., P. Van Der Straaten, M. Krijn and C. A. P. G. Van Der Mast (2001). "Research on presence in virtual reality: A survey." CyberPsychology & Behavior 4(2): 183-201.

Scriven, J. and A. Ehrenberg (1995). How Consumers Respond to Price. Proceedings of the MRS annual conference, London, Market Research Society.

Sekhon, J. S. and R. Titiunik (2012). "When natural experiments are neither natural nor experiments." American Political Science Review 106(1): 35-57.

Senter, S. M. and D. H. Wedell (1999). "Information presentation constraints and the adaptive decision maker hypothesis." Journal of Experimental Psychology: Learning, Memory, and Cognition 25(2): 428-446.

Sethuraman, R. (2003). "Measuring national brands' equity over store brands." Review of Marketing Science 1(2): 1-25.

Sethuraman, R. and C. Cole (1999). "Factors influencing the price premiums that consumers pay for national brands over store brands." Journal of Product and Brand Management 8(4): 340-351.

Sewell, W. and O. Komogortsev (2010). Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network. CHI 2010 Extended Abstracts on Human Factors in Computing Systems, Atlanta, ACM.

Shams, P., E. Wästlund and L. Witell (2012). Revisiting Russo and Leclerc. Symposium on Eye Tracking Research and Applications, Santa Barbara, California, ACM.

Sharp, A. and J. Romaniuk (2002). Is ignorance bliss? Perceptual response temporal stability and the effect of brand familiarity and market knowledge. European Marketing Academy 31st annual conference. Braga, Portugal.

Sharp, B. (2010). How Brands Grow. South Melbourne, Oxford University Press.

Sharp, B. (2013). Marketing: Theory, evidence, practice. Melbourne, Oxford University Press.

Sharp, B. and S. Bogomolova (2013). Meaningful Marketing Metrics. Marketing: Theory, Evidence, Practice. B. Sharp. Melbourne, Australia, Oxford University Press.

Sharp, B. and J. Romaniuk (2016). How brands grow. How Brands Grow: Part 2. J. Romaniuk and B. Sharp. South Melbourne, Oxford University Press: 1-22.

Sharpe, K. M., R. Staelin and J. Huber (2008). "Using Extremeness Aversion to Fight Obesity: Policy Implications of Context Dependent Demand." Journal of Consumer Research 35(3): 406-422.

Sheridan, T. B. (1992). "Musings on telepresence and virtual presence." Presence: Teleoperators and virtual environments 1(1): 120-125.

Shi, S. W., M. Wedel and R. Pieters (2013). "Information acquisition during online decision making: A model-based exploration using eye-tracking data." Management Science 59(5): 1009-1026.

Shiffrin, R. M. and W. Schneider (1977). "Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory." Psychological review 84(2): 127.

Silayoi, P. and M. Speece (2004). "Packaging and purchase decisions: an exploratory study on the impact of involvement level and time pressure." British Food Journal 106(8): 607-628.

Silayoi, P. and M. Speece (2007). "The importance of packaging attributes: a conjoint analysis approach." European Journal of Marketing 41(11/12): 1495-1517.

Silk, A. J. and G. L. Urban (1978). "Pre-Test-Market Evaluation of New Packaged Goods A Model and Measurement Methodology." Journal of Marketing Research 15: 171-191.

Simonson, I. and R. S. Winer (1992). "The influence of purchase quantity and display format on consumer preference for variety." Journal of Consumer Research 19(1): 133-138.

Singh, J., G. Goodhardt and A. Ehrenberg (2001). Loyalty to product attributes. Report 11 for Corporate Sponsors. Adelaide, Ehrenberg-Bass Institute for Marketing Science: 1-20.

Slater, M. (1999). "Measuring presence: A response to the Witmer and Singer presence questionnaire." Presence: Teleoperators and Virtual Environments 8(5): 560-565.

Slater, M., V. Linakis, M. Usoh, R. Kooper and G. Street (1996). Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess, ACM Press New York.

Slater, M. and S. Wilbur (1997). "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments." Presence: Teleoperators and virtual environments 6(6): 603-616.

Smolentzov, L., A. Olsen and Tommy Strandvall (2010). Comparing different eye tracking cues when using theretrospective think aloud method in usability testing, Tobii Technology, Clemsen University: 45-53.

Soderman, M. (2005). "Virtual reality in product evaluations with potential customers: An exploratory study comparing virtual reality with conventional product representations." Journal of Engineering Design 16(3): 311-328.

Sommer, R. and S. Aitkens (1982). "Mental mapping of two supermarkets." Journal of Consumer Research: 211-215.

Sorensen, H. (2016). Inside the mind of the shopper: The science of retailing. New Jersey, United States, Pearson Education.

Sorensen, H., S. Bogomolova, K. Anderson, G. Trinh, A. Sharp, R. Kennedy, B. Page and M. Wright (2017). "Fundamental patterns of in-store shopper behavior." Journal of Retailing and Consumer Services 37: 182-194.

Sorensen, H. and J. Suher (2010). "The Power of Atlas: Why In-Store Shopping Behavior Matters." Journal of Advertising Research 50(1): 21-29.

Sorensen, H. and J. Suher (2010). Vision: The Doorway to the Shopper's Brain, TNS Sorensen.

Sørensen, H. S., J. Clement and G. Gabrielsen (2012). "Food labels–an exploratory study into label information and what consumers see and understand." The International Review of Retail, Distribution and Consumer Research 22(1): 101-114.

Sparto, P. J., S. L. Whitney, L. F. Hodges, J. M. Furman and M. S. Redfern (2004). "Simulator sickness when performing gaze shifts within a wide field of view optic flow environment: Preliminary evidence for using virtual reality in vestibular rehabilitation." Journal of Neuroengineering and Rehabilitation 1(14): 1-10.

Steenkamp, J., M. de Jong and H. Baumgartner (2010). "Socially Desirable Response Tendencies in Survey Research." Journal of Marketing Research 47(2): 199-214.

Steenkamp, J.-B. E. M., H. J. van Heerede and I. Geyskens (2010). "What Makes Consumers Willing to Pay a Price Premium for National Brands over Private Labels?" Journal of Marketing Research 47(6).

Steuer, J. (1992). "Defining virtual reality: Dimensions determining telepresence." Journal of Communication 42(4): 73-93.

Stilinovic, M. (2016, June 16). "Premiumization: The most affluent retail trend of 2016." Retrieved October 2016, from forbes.com/sites/millystilinovic/2016/06/16/premiumisation-the-most-affluent-retail-trend-of-2016/#63d877a55d07.

Stone, D. and B. Congdon (2007). Perceived Realism and Eye Tracking Performance within a Virtual Shopping Environment. South Carolina, USA, Clemson University: 1-8.

Sudman, S. and N. M. Bradburn (1974). Response effects in surveys. Chicago, Aldine Publishing Company.

Sutcliffe, A., B. Gault and J.-E. Shin (2005). "Presence, memory and interaction in virtual environments." International Journal of Human-Computer Studies 62(3): 307-327.

Szabo, M., S. Bogomolova, R. Kennedy and M. Nenycz-Thiel (2012). How manufacturers and retailers make decisions about price promotions. ANZMAC, Adelaide, Australian and New Zealand Marketing Academy.

Tal, A. and B. Wansink (2011). "Turning Virtual Reality into Reality: A Checklist to Ensure Virtual Reality Studies of Eating Behavior and Physical Activity Parallel the Real World." Journal of Diabetes Science and Technology 5(2): 239-244.

Tan, D., D. Gergle, P. Scupelli and R. Pausch (2004). Physically large displays improve path integration in 3D virtual navigation tasks. SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, ACM.

Tanusondjaja, A., M. Nenycz-Thiel and R. Kennedy (2016). "Understanding shopper transaction data: how to identify cross- category purchasing patterns using the duplication coefficient." International Journal of Market Research 58(3): 1-12.

Tatler, B. W. (2007). "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions." Journal of Vision 7(4): 1-17.

Tellis, G. J. (1988). "The Price Elasticity of Selective Demand: A Meta Analysis of Econometric Models of Sales." Journal of Marketing Research 25(November): 331-341.

Tjostheim and Saether-Larse (2005). How to validate a new MR tool? A case study in FMCG. ESOMAR, Paris, France.

Tjostheim, I. and B. Haugland (2005). What do you plan to buy in the store today? Impulse buying in grocery stores - the challenge for market research. ESOMAR. Budapest, WARC.

Tobii (2012). Accuracy and Precision Test Report: Tobii X120 Eye Tracker: 27.

Tobii Technology (2010). Tobii Technology: Tobii TX300 eye tracker. online, Tobii Technology: 18.

TobiiPro. (2018). "Tobii Pro VR Integration: Specifications." Retrieved August 2018, from https://www.tobiipro.com/product-listing/vr-integration/#Specifications.

Toh, R. S., E. Lee and M. Y. Hu (2006). "Social desirability bias in diary panels is evident in panelists' behavioral frequency." Psychological Reports 99(2): 322-334.

Tonkin, C., A. D. Ouzts and A. T. Duchowski (2011). Eye tracking within the packaging design workflow: interaction with physical and virtual shelves. 1st Conference on Novel Gaze-Controlled Applications, ACM.

Tourangeau, R. (2000). Remembering what happened: memory errors and survey reports. The science of self-report: Implications for research and practice. A. A. Stone, J. S. Turkkan, C. A. Bachrach et al. Mahwah, Lawrence Erlbaum Associates: 29-47.

Tourangeau, R. and K. Rasinski (1988). "Cognitive Processes Underlying Context Effects in Attitude Measurement." Psychological Bulletin 103(3): 299-314.

Treiber, B. and S. Needel (2000). New Ways to Explore Consumer Point-of-Purchase Dynamics: Combining virtual reality and choice modeling in consumer purchase analyses: 261-275.

Tyndiuk, F., V. Lespinet-Najib, G. Thomas and C. Schlick (2004). Impact of large displays on virtual reality task performance. Proceedings of the 3rd international conference on Computer graphics, South Africa, ACM.

Uncles, M. (2011). "Viewpoint: Publishing replications in marketing." International Journal of Market Research 53(5): 579-582.

Uncles, M. D. and S. Kwok (2013). "Designing research with in-built differentiated replication." Journal of Business Research 66(9): 1398-1405.

Uncles, M. D. and M. Wright (2004). "Editorial: empirical generalisation in marketing." Australasian Marketing Journal 12(3): 5-12.

Underhill, P. (1999). Why We Buy: The Science of Shopping. London, Obat, Inc.

Underwood, R. L. (2003). "The communicative power of product packaging: creating brand identity via lived and mediated experience." Journal of Marketing Theory & Practice 11(No. 1): 62-76.

Underwood, R. L., N. M. Klein and R. R. Burke (2001). "Packaging communication: attentional effects of product imagery." Journal of Product & Brand Management 10(7): 403-422.

Usoh, M., E. Catena, S. Arman and M. Slater (2000). "Using presence questionnaires in reality." Presence: Teleoperators and Virtual Environments 9(5): 497-503.

Van Der Lans, R., R. Pieters and M. Wedel (2008). "Competitive brand salience." Marketing Science 27(5): 922-931.

Van Der Lans, R., R. Pieters and M. Wedel (2008). "Eye-Movement Analysis of Search Effectiveness." Journal of the American Statistical Association.

van Herpen, E., E. van den Broek, H. C. M. van Trijp and T. Yu (2016). "Can a virtual supermarket bring realism into the lab? Comparing shopping behavior using virtual and pictorial store representations to behavior in a physical store." Appetite 107: 196-207.

Van Ittersum, K., B. Wansink, J. M. E. Pennings and D. Sheehan (2013). "Smart shopping carts: how real-time feedback influences spending." Journal of Marketing 77(6): 21-36.

van Nierop, E., D. Fok and P. H. Franses (2008). "Interaction Between Shelf Layout and Marketing Effectiveness and Its Impact on Optimizing Shelf Arrangements." Marketing Science 27(6): 1065-1082.

Vanhuele, M. and X. Dreze (2002). "Measuring the price knowledge shoppers bring to the store." The Journal of Marketing 66(4): 72-85.

Voleti, S. and R. Sethuraman (2015). Are national brands more promotion elastic than store brands? Advances in national brand and private label marketing. F. J. Martínez-López, J. C. Gázquez-Abad and R. Sethuraman, Springer International Publishing: 63-70.

Vrechopoulos, A. P., R. M. O'Keefe, G. I. Doukidis and G. J. Siomkos (2004). "Virtual store layout: an experimental comparison in the context of grocery retail." Journal of Retailing 80(1): 13-22.

Wakefield, K. L. and J. J. Inman (2003). "Situational price sensitivity: the role of consumption occasion, social context and income." Journal of Retailing 79: 199-212.

Walker, P. (2015). "Depicting visual motion in still images: Forward leaning and a left to right bias for lateral movement." Perception 44(2): 111-128.

Walker, R., R. Pettit and J. Rubinson (2009). "The foundations of quality initiative: a five-part immersion into the quality of online research." Journal of Advertising Research 49(4): 464-485.

Wansink, B. (1996). "Can package size accelerate usage volume?" The Journal of Marketing 60(3): 1-14.

Wansink, B. and K. van Ittersum (2003). "Bottoms Up! The Influence of Elongation on Pouring and Consumption Volume." Journal of Consumer Research 30(December): 455-464.

WARC (2016a). What is virtual reality? - IAB VR Report. United Kingdom, WARC: 1-18.

WARC (2016b) "What we know about virtual reality." 1-3.

WARC (2017) "Brand experience in 2017: how brands can use virtual and augmented reality." 1-6.

Wasserman, T. (2009). Virtual Shopping, Real Results. AdWeek. Online.

Waterlander, W. E., Y. Jiang, I. H. M. Steenhuis and C. N. Mhurchu (2015). "Using a 3D virtual supermarket to measure food purchase behavior: A validation study." Journal of Medical Internet Research 17(4): e107.

Waterlander, W. E., M. Scarpa, D. Lentz and I. H. Steenhuis (2011). The virtual supermarket: An innovative research tool to study consumer food purchasing behaviour. BMC Public Health. online, BioMed Central.

Wedel, M. and R. Pieters (2000). "Eye Fixations on advertisements and memory for brands: A model and findings." Marketing Science 19(4): 297-312.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." Econometrica: Journal of the Econometric Society 48(4): 817-838.

Wierenga, B. (2012). Understanding marketing decision-making. RSM Insight. Rotterdam, Netherlands, Rotterdam School of Management, Erasmus University. 11: 12-14.

Wilson, C. J. and A. Soranzo (2015). "The use of virtual reality in psychology: a case study in visual perception." Computational and Mathematical Methods in Medicine 2015: 7.

Wind, Y. and D. Lerner (1979). "On the measurement of purchase data: surveys versus purchase diaries." Journal of Marketing Research 16(February): 39-47.

Witmer, B. G. and M. J. Singer (1998). "Measuring presence in virtual environments: A presence questionnaire." Presence: Teleoperators and virtual environments 7(3): 225-240.

Wright, M., A. Schnack, J. Holdershaw and J. Elms (2017). Virtual tracking of consumer behaviour: Pilot-testing of ViCoS 2.1. New Zealand, EMAC 1-20.

Wright, M. and L. Stocchi (2010). The temporal stability of a stochastic model. ANZMAC. Christchurch.

Yao, J. and H. Oppewal (2016). "Unit pricing matters more when consumers are under time pressure." European Journal of Marketing 50(5/6): 1094-1114.

Yeager, D. S., J. A. Krosnick, L. C. Chang, H. S. Javitz, M. S. Levendusky, A. Simpser and R. Wang (2009). "Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples." Public Opinion Quarterly 75(4): 1-41.

Young, S. (2004). "Breaking Down the Barriers to Packaging Innovation." Design Management Review 15(1): 68-73.

Young, S. (2005). Breaking through the clutter: Insights for measuring and improving shelf visibility Package Design. Skokie, Illinois, Package Design. July/August: 58-61.

Zhen, C., J. L. Taylor, M. K. Muth and E. Leibtag (2009). "Understanding differences in self-reported expenditures between household scanner data and diary survey data: A comparison of homescan and consumer expenditure survey." Review of Agricultural Economics 31(3): 470-492.

Zygourisa, S., D. Giakoumisb, K. Votisb, S. Doumpoulakisb, K. Ntovasa, S. Segkoulic, C. Karagiannidisd, D. Tzovarasb and M. Tsolakia (2014). "Can a virtual reality cognitive training application fulfill a dual role? Using the virtual supermarket cognitive training application as a screening tool for mild cognitive impairment." Journal of Alzheimer's Disease 40: 1-10.