Empirical Generalisations of Brand Extension Theory and the Role of General Linear Mixed Models

Damien William Mather

A thesis submitted for the degree of
Doctor of Philosophy
at the University of Otago, Dunedin,
New Zealand

Date: March 1, 2011
Abstract

The objective of this research thesis is to resolve the confusion that exists in the marketing literature around brand extension theory, focusing on models that include original brand quality and four brand extension fit attributes realized as complementarity (of the extension with the original branded product), substitutability (of the extension for the original), transferability (of the firm’s resources and capabilities from the original to the extension) and difficulty (perhaps best explained as the opposite of the triviality of the manufacture of the extension by the firm, given the firm’s perceived level of capability).

The method used to achieve this objective is to pool the data from one original and eight ‘close replication’ experiments, and re-analyze the data using General Linear Mixed Models. Appropriate random effects specifications to model correlation and covariance amongst the predictor variables, as well as random study location and subject effect specifications are employed. A range of models for the fixed effects listed above will be compared, including models with (a) direct effects, (b) indirect or interaction effects amongst the original brand quality and effects of complementarity, substitutability, transferability and difficulty, and finally (c) second-order ‘ideal point’ model effects on extension evaluation as suggested by certain marketing academics.

The implications for brand practitioners are that, contrary to the findings of some of the close replication studies, both direct and indirect effects of quality and fit on brand extension evaluation are supported, with the exception of the difficulty attribute. Natural limits to fixed practitioner brand extension preference strategies are discovered. These limits mean that additional choice modeling or similar is required to correctly explore the implications of proposed brand extensions. The implications for researchers are that seemingly unrelated research streams predicting ‘ideal point’ evaluation responses are shown to be inter-related with brand extension theory, and General Linear Mixed Models are highlighted as a useful inference framework for improving and expanding empirical generalisations in marketing research.
Preface and Personal Acknowledgements

The motivation for me to initiate further research in this topic was stimulated by a friendship and research collaboration with Lorraine Sunde that began in 1988 whilst Lorraine was writing her own Masters’ thesis in this topic area as a student of Marketing at the University of Auckland. Further motivation was built whilst I was working in the Strategic Services Group of Telecom New Zealand Ltd. 1996-1997 when considerable marketing effort within that organization was associated with multiple new product development all of which was brand extension.

Concurrently Lorraine Sunde and her Masters’ thesis supervisor Professor Rod Brodie were publishing a paper highlighting the lack of agreement amongst attempts at empirical generalisations of Brand Extension theory. Subsequently Lorraine took up a position in the Marketing Department at the University of Otago.

Our research partnership strengthened when we jointly researched and presented a paper in this research area at the ANZMAC ’98 conference held at Otago University.

Finally the opportunity and motivation to research and write this thesis arose as I joined the Marketing department at Otago University myself in 2002, several years after Lorraine had left the department and the city to devote her energies to her new roles as a wife to Andrew Miller, mother to Sam and ‘budding’ organic orchardist.

I am indebted to Lorraine and my current supervisors Professor Rob Lawson and Dr. David Holdsworth for their guidance, encouragement and support. I am especially grateful to Professor Rob Lawson, then head of the Department of Marketing, and Professor David Buisson, as then head of the School of Business and Assistant Vice-Chancellor, for equipping me with the opportunity and resources to direct my own learning in the direction this thesis takes. Dr. John Knight also has been of enormous additional help with reviewing this thesis. The University of Otago central library staff provided valuable
additional disability support during the many months I spent recovering from a badly broken leg.

 Lastly, I am indebted to the many academics referenced herein who have published the research associated with this topic, especially to the group who bravely and selflessly shared their close replication study data sets with me and other researchers for further critical analysis at the most detailed and transparent level of accountability possible. It is difficult to find a better example of altruism and adherence to the principles at the cores of both academic scholarship and objective scientific process in marketing literature. This was not a risk-free process for them. During the course of this thesis preparation I have learnt more than I’d ever hoped for, and have thoroughly enjoyed every aspect, to a much greater extent than I had expected.

 Thank you all.
Table of Contents

Abstract ii
Preface and Personal Acknowledgements iii
Table of Contents v
List of Tables ix
List of Abbreviations xi
Chapter 1 Introduction 1
1. The Aim of this Thesis 1
2. The Importance of Brands 1
3. Why is it Important to Understand Brand Extensions? 2
4. Consumer perspectives on brands 3
5. Rarity of empirical generalisations in marketing 4
   The inferential framework for brand extension empirical generalisation is questionable 5
6. Overarching Questions and Related Literature 7
   How do the competing fundamental theories of memory and judgment relate to brand extension theory? 7
   Why has there been so much disagreement about brand extension theory? 8
   How can General Linear Mixed Models help empirical generalisations? 8
   How might Reliability Analysis help Brand Extension Theory? 8
   What informs successful brand extensions? 9
7. Why an Empirical Generalisation context was chosen for this research 9
   Empirical generalisations in brand extension theory: a lack of progress. 9
   Opportunity to generalise inferences from pooled data 10
8. Outline of thesis 10
Chapter 2 Literature 13
1. Empirical Generalisations from Physical to Social Sciences and Marketing 13
   What is an empirical generalisation? 13
   The earliest known empirical generalisations from the Sumerians 14
   The philosophy of science and empirical generalisations of the ancient Greeks 15
   Plato (427 B.C. – 327 B.C.) 16
   Classical Developments of Empirical Generalisations: Renaissance Europe 17
   Copernicus 18
   The increasing rate of empirical generalisations in science 19
   Galileo Galelei (1564-1642) 19
   Issac Newton (1642-1727) 20
   David Hume (1711-1776) 20
2. Empirical Generalisations and theory development 21
   The emergence of social science and Auguste Compte (1798-1857) 21
   Social science positivism and Emile Durkheim (1858-1917) 22
   Causal theories, the social sciences and Norman Campbell (1880-1849) 23
   Inference generation and Fisher 23
   Type I errors and R. A. Fisher 24
   Social science theory development, type II errors and Karl Popper 26
   Relativism, Phenomenology and their impact on the social sciences 27
   The conceptual status of things in language, Wittgenstein (1889-1951) and Heidegger (1889-1976) 28
   Separation of meaning, Hans-Georg Gadamer (1900-2002) and Paul Ricoeur (1913-2005) 30
3. Empirical generalisations in the social sciences and their lack of success 34
   Empirical Generalisations and Marketing 36
   Empirical Generalisations and Marketing Knowledge 38
   Empirical Generalisations and Brand Extensions 38
4. Consumer evaluations of Brands 40
   The importance of Brand Extensions 40
Review of the prominent theories of consumer evaluation of brands

44

Associative Network Memory Model

45

Causal Reasoning Model

50

Categorisation Theories

51

Ideal point evaluation models

52

Cognitive Dissonance Theory

53

Social Judgment Theory

55

5. Brand Extension theory

57

Brand image and knowledge

57

Fit

60

Chapter 3 Research questions and propositions

65

Literature gaps identified in this research

65

Question 1: What is the basis for Brand Extension Theory?

66

Question 2: Is Aaker and Keller’s brand extension theory generalisable?

66

Question 3: What were the problems with previous attempts at Empirical Generalisations in the Brand Extension literature?

67

Question 4: What is the relationship amongst the multiple Brand Extension Theories?

68

Chapter 4 Methodology

70

Increase in certainty and power of statistical tests

70

Parameter estimate bias due to multi-co-linearity

71

Parameter estimate bias due to residual centering

71

Approach Taken

71

General Linear Mixed Model frameworks and inference space

76

Parameter estimate bias due to heteroskedasticity

78

Ideal Point or inverted “U” response

79

Other posited effects on brand extension evaluation

81

Predicted limit outcomes of the associative memory network model

83

1. Lack of measurement models used in existing literature

84

2. Estimation Frameworks, Techniques and Errors

84

Ordinary Least Squares

84

Residual Centering

85

Basic regression modeling errors affect the published inferences

86

6. General Linear Mixed Models

86

7. Proposed Estimation Framework and specification improvement

87

Intra-Fit interaction effects

87

8. Alternative and Complimentary Estimating Frameworks

89

Ridge Regression

89

Random Coefficient Structural Equation Modeling

89

The Jackknife and the Bootstrap

90

Chapter 5 Data

91

Example ratings tasks on a 1-7 scale for a ‘McPhoto’ extension of the ‘McDonalds’ Brand

91

Aaker and Keller

92

Aaker & Keller questions and semantic likert scale anchors

93

Aaker & Keller Brand Extensions

93

Sunde & Brodie

94

Sunde & Brodie questions and semantic likert scale anchors

94

Sunde and Brodie Brand Extensions

95

Alexandre-Bourhis

95

Alexandre-Bourhis questions and semantic likert scale anchors

96

Alexandre-Bourhis Brand Extensions

96

Fu and Saunders

96

Fu and Saunders questions and semantic likert scale anchors

97

Fu and Saunders Brand Extensions

98

Nijssen & Hartman

98

Nijssen & Hartman questions and semantic likert scale anchors

99

Nijssen & Hartman Brand Extensions

99

Bottomley and Doyle

100

Bottomley and Doyle’s questions and semantic likert scale anchors

100

Bottomley and Doyle’s Brand Extensions

101

Holden and Barwise

101
Chapter 6 Results

Analysing the data presented in the above chapter 5 using the methods outlined in chapter 4 yields the following results.

1. Pooled Mixed Effects Model Estimation
2. Pooled Mixed Effects Model Covariance Structure Selection
3. Model Estimation
   - Fixed effects OLS with Difficulty but without Residual Centering or intra-fit interactions
   - OLS Fixed effects with Residual Centering and Difficulty but without intra-fit interactions
   - Note the very high levels of positive estimate bias introduced into the main effect estimates by the erroneous application of the residual centering technique in this context.
   - Fixed effects with Difficulty and intra-fit interactions but without residual centering
   - Fixed effects with Difficulty and intra-fit interactions but without residual centering
   - Mixed effects with Difficulty but without intra-fit interactions and Residual Centering
   - Mixed effects with Difficulty and intra-fit interactions but without residual centering
   - OLS without Residual Centering or intra-fit interactions or Difficulty
   - With Residual Centering but without Difficulty and intra-fit interactions
   - Fixed effects with intra-fit interactions but without Difficulty and residual centering
   - Mixed effects without intra-fit interactions or Residual Centering or Difficulty
   - Mixed effects with intra-fit interactions but without Difficulty or residual centering
   - Pooled GLMM covariance structure model with ideal point specifications
   - Type I and II error and sample size of the GLMM parameter tests
   - Intra-fit interaction effects.
   - Ideal point fit specification effects.

Chapter 7 Analysis of Findings

1. The Indirect role of Quality
2. The Direct and Indirect roles of the three Fit attributes
3. The lack of any generalized role for the Difficulty attribute
4. The problem with residual centering is further demonstrated
5. The advantages of mixed models and intra-fit interactions are further demonstrated
6. Ideal point or inverted “U” model for extrinsic effects was found
7. Optimal extension attribute mixes: an exploration
8. The relationship between power and sample size

Chapter 8 Conclusions, implications and limitations

1. Conclusions
2. Implications
3. Implications for Brand Extension Research
4. Implications for Other Brand Research
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implications for Empirical Generalisations in Marketing Research</td>
<td>149</td>
</tr>
<tr>
<td>Implications for Other Brand Management research</td>
<td>150</td>
</tr>
<tr>
<td>Implications for Marketing Practitioners</td>
<td>151</td>
</tr>
<tr>
<td>3. Limitations</td>
<td>152</td>
</tr>
<tr>
<td>References</td>
<td>154</td>
</tr>
</tbody>
</table>
List of Tables

Table 1 Aaker and Keller items and scale anchors ......................................................... 93
Table 2 Aaker and Keller data brand extensions ............................................................ 93
Table 3 Sunde & Brodie items and scale anchors ........................................................ 94
Table 4 Sunde and Brodie brand extensions ................................................................... 95
Table 5 Alexandre-Bourhis items and scale anchors ...................................................... 96
Table 6 Alexandre-Bourhis brand extensions ................................................................. 96
Table 7 Fu and Saunders items and scale anchors ........................................................ 97
Table 8 Fu and Saunders Brand Extensions ................................................................... 98
Table 9 Nijssen & Hartman items and scale anchors ..................................................... 99
Table 10 Nijssen & Hartman brand extensions ............................................................... 99
Table 11 Bottomley and Doyle’s items and scale anchors ........................................... 100
Table 12 Bottomley and Doyle’s Brand Extensions ....................................................... 101
Table 13 Holden and Barwise’s items and scale anchors .............................................. 102
Table 14 Holden and Barwise brand extensions ............................................................. 102
Table 15 Fractional-factorial reference design for 5 main effects and 4 quality-fit 2-way interactions ........................................................................................................... 103
Table 16 Patro and Jaiswal items and scale anchors ....................................................... 104
Table 17 Patro and Jaiswal brand extensions ................................................................. 104
Table 18 Barratt, Lye and Venkateswarlu’s items and scale anchors ............................ 105
Table 19 Barratt, Lye and Venkateswarlu’s Brand Extensions ...................................... 106
Table 20: Summary of reported significance in replication studies at the 95% C.L......... 106
Table 21 Full-factorial reference design for 5 linear factors .......................................... 108
Table 22 Fractional-factorial reference design for 5 main effects and all 2-way interactions .......................................................... 109
Table 23 Post-hoc factor means for the ESSEC data .................................................... 110
Table 24 Original and close replication post-hoc studies’ design efficiencies ............... 111
Table 25 Pooled original and close replication covariance matrix ................................. 112
Table 26 Pooled original and close replication pearson’s correlation matrix ............... 112
Table 27 Best eight Covariance Structures and associated M-2LL, REML, rAIC, and Pseudo-R-Squared statistics ................................................................. 118
Table 28 Best estimate of Significance of Effects at the 95% C.L. using Kenward-Rogers Density Degrees of Freedom for parameter T-statistics and associated P-Value for Model 2 ................................................................................................................................. 120
Table 29 Significance of Effects at the 95% C.L. using Likelihood Ratio statistics between strictly nested models ........................................................................................ 120
Table 30 Significance of Effects at the 95% C.L. by Study Location using an OLS Fixed Effects Model Estimation Framework without Residual Centering or intra-fit interactions but including Difficulty ............................................................ 122
Table 31 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework with Residual Centering ........................................ 123
Table 32 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework with Difficulty but without Residual Centering ........ 124
Table 33 Significance of Effects at the 95% C.L. by Study Location using a Mixed Effects Model Estimation Framework without Residual Centering ............................ 125
Table 34 Significance of Effects at the 95% C.L. by Study Location using a Mixed Effects Model Estimation Framework with Difficulty but without Residual Centering ............... 126
Table 35 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework without Residual Centering or intra-fit interaction or Difficulty ............................................................................................................. 127
Table 36 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework with Residual Centering .................................... 127
Table 37 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework without Difficulty and Residual Centering .............. 128
Table 38 Significance of Effects at the 95% C.L. by Study Location using a Mixed Effects Model Estimation Framework without intra-fit interactions or residual centering .......... 129
Table 39 Significance of Effects at the 95% C.L. by Study Location using a Mixed Effects Model Estimation Framework without residual centering or Difficulty ................. 130
Table 40 Best eight Ideal point Covariance Structures and associated ML, REML, r-AIC and Pseudo-R-Squared statistics ................................................................. 131
Table 41 Best estimate of Significance of Effects at the 95% C.L. using Kenward-Rogers Density Degrees of Freedom for parameter T-statistics and associated P-Value for ideal point model 1........................................................................................................................................... 133
Table 42 Type II error rates derived from bootstrap estimates......................................................... 135
Table 43 optimised extension evaluation strategies using the best ‘ideal point’ model........ 141
Table 44 Confirmation or disconfirmation of propositions................................................................. 144
List of Abbreviations

C.L.: Confidence Level
REML: REstricted Maximum Likelihood
ML: Maximum Likelihood
M-2LL: Maximum -2 Log Likelihood
DDF: Denominator Degrees of Freedom
AIC: Akaike Information Criteria based on the ML
rAIC: Akaike Information Criteria based on the REML
AICC: AIC (Corrected)
OLS: Ordinary Least Squares
SEM: Structural Equation Model
LRT: Likelihood Ratio Test
Q: Original Brand Quality
C: Complementarity
S: Substitutability
T: Transferability
D: Difficulty
Y: Yes
N: No
N/A: Not Applicable
Chapter 1 Introduction

1. The Aim of this Thesis

Aaker and Keller’s theory of brand extension proposes that original brand quality and the congruence, or ‘fit’ between the original brand and the extension concept, influence brand extension evaluation. However eight subsequent attempts at close replication of the original experiment failed to generalise their original result.

Aaker and Keller’s theory of brand extension is that original brand quality, together with four ‘fit’ attributes of complementarity, substitutability, transferability and difficulty, has a direct effect on the preference for the new brand extension. They also hypothesized that there is an interactive effect between quality and the fit effects, leading to a model where the fit effects moderated (but not completely mediated) the effect of original quality on extension preference.

The aim of this thesis is to resolve the confusion that exists in the literature about which parts of Aaker and Keller’s theory of brand extension can be generalised, and to extend the theory to include ideal point effects as suggested by other researchers. This will be achieved by the analysis of a careful application of a more appropriate quantitative modeling framework to a pooled data set formed from nine close replications studies over thirteen locations.

2. The Importance of Brands

Aaker’s original definition of a brand is ‘a distinguishing name and/or symbol (such as a logo, trademark or package design) intended to identify the goods or services of either one seller or a group of sellers, and to differentiate those goods or services from those of competitors.’ (Aaker 1991 p7). Hence the essence of a brand is a differentiating role in the context of choices amongst alternatives.

Aaker uses the term brand equity to describe “..a set of brand assets and liabilities linked to a brand, its name and symbol, that add to or subtract from the value provided by a product or
service to a firm and/or that firm’s customers” (Aaker 1991 p7). This customer value can be provided in several ways, for example:

1. A saving of customer search time, enabling rapid location of a trusted brand on a supermarket shelf, at a price that the customer knows or expects will be fair or competitive, or

2. Reduction in customer risk, providing assurance to the customer that the product will be reliable and unlikely to fail,

3. Assurance that, if by chance the branded product is found to be faulty, it will be replaced quickly and without further cost.

It is helpful to think of brand equity from a different perspective as the ‘shorthand’ customer value that results from sustained and clearly communicated good performance by a branded product. An ‘acid test’ for this ‘shorthand’ for customer value is whether brand equity will sustain a branded product’s sales even when marketing mix performance or its sales and marketing communication drops temporarily in intensity (when more is better) or alternatively diverges from the optimum level (when there is an ideal intensity).

Using the Interbrand method of estimating the value of global brands, the cumulative value of the top 100 global brands varied between approximately 1 and 1.2 trillion US dollars from 2001 to 2009. This finding is based on aggregating 900 individual brand value estimates from 9 publications (Interbrand 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009). This high level of value has been built by firms investing in promotion, distribution, quality design and customer satisfaction that has resulted in brand awareness, recall and loyalty upon which the valuation is based (Murphy 1987 p3).

3. Why is it Important to Understand Brand Extensions?

Brand extensions capitalize on the above brand value. Extending an established brand into new product categories, re-branding and co-branding continue to enjoy popularity as useful marketing tactics. For example, Caterpillar, ranked 68th. in the top 100 global brands, (Robinson 2005) extended its brand from the original “Caterpillar” branded heavy earth moving machinery to robust boots with the Caterpillar brand. More recent examples include extending well known fashion brands into the fragrance category: ‘Guess by Marciano’
(Jeffries 2008), McDonalds extending their brand into flavoured and iced coffee drinks (York 2010), Donald Trump’s brand into steaks, Huggies diapers into sunscreen and Guinness into whole grains (Hein 2007). By 1980 brand extension had been posited as the most important marketing trend (Ries and Trout 1981 p 127) but was also regarded as complicated and risky (Murphy 1987 p 5).

About 90% of new brands in both U.S. supermarkets and drug stores in 1991 (Rangaswamy, Burke, and Oliva 1993b) and in the European Fast Moving Consumer Goods (FMCGs) markets in 1995 and 1996 (ACNielsen and Ernst&Young 1999) were introduced as brand extensions

Aaker and Keller (1990) posited that a positive attitude towards a brand extension was influenced by the three main-effect brand extension ‘fit’ attributes, namely complementarity, substitutability and transferability, and the indirect effect of original brand quality moderated by the ‘fit’ attributes. Their paper also posited a direct effect of difficulty. Since the Caterpillar boots were a success, Aaker and Keller’s theory suggests we should find a good fit between the concept of heavy earth moving machinery and robust boots (and we do).

4. Consumer perspectives on brands

In addition to a perspective on brands focused around associative memory networks of brand concepts, as typified by Keller (1993), and discussed at length in this thesis, there is a wider perspective to briefly consider, encompassing relationships amongst self, self-concept, identity, and the meaning of owned or possessed objects. This perspective is perhaps best typified by Belk (1988). His compelling argument for the concept of self as a calculus of body, mind, possessions, pets and relatives leads us to view memory constructs of a possessed, and possibly branded, object, from a new perspective. These memory constructs are not just formed by attributes intrinsic to that object, but also possibly by attributes and concepts somewhat extrinsic to that object but intrinsic to the person or self whose associative memory networks we are considering. Ownership is not a one-way street (Belk 1988). Rather, ownership influences both the owner’s perception of the object and, reflexively, the concept of self, being a calculus of things including that owned object, is altered by the very fact of ownership as well. The corollaries associated with this realization assist in the explanation of
much of the phenomenon observed in wider brand studies. These phenomena include identification with brands, the concept of brand personalities and sub-culture adoption of brands in unanticipated ways (Brucks and Das 1999). However, both the negative emotions associated with replacing lost, broken or stolen objects or body parts, and relative dispossession and poverty, and the positive emotions associated with desired possessions form the context within which consumer evaluations of brand extensions are made. From the discussion above it can be seen that these contexts are in general unique to the individual or social group, whose contextual effects on brand evaluations vary between subjects and groups. Thus these effects can be treated as random subject effect in a more specific model of consumer evaluation of brand extensions. Discussion in the following sections will explain how inferences generated from models with random subjects’ effects are explicitly more generalisable.

5. Rarity of empirical generalisations in marketing

There have been relatively few successful generalisations of marketing theory. Leone and Shultz in their book on marketing knowledge concluded that “. . . when the same standards used to define generalisations in other fields were applied to marketing, our scientific foundations appeared to be marsh rather than bedrock.” (Leone and Schultz 1980 pp 10-11).

Even more disheartening are the results and conclusions of a study that randomly sampled 1120 articles from the Journal of Marketing, the Journal of Market Research and the Journal of Consumer Behaviour published during the period 1974 to 1989 inclusive. This sample comprised a substantial 50% of all articles published in these respected journals. Of these 1120 articles, 835 were classed as empirical studies. Only 20 of these were empirical generalisations, and not a single article was found to be a ‘close’ or ‘straight’ replication. All of these 20 studies were extension replications, designed to explore and discover the boundary conditions, i.e. the point at which the broader study fails to replicate the original empirical results. The impact of the breadth of each replication on agreement was not discussed, and may have contributed to the poor agreement found amongst the 20 studies: “15% (three) confirmed earlier results, 25% (five) provided some support, and 60% (12) conflicted with their predecessors.” (Hubbard and Armstrong 1994 p 7).
Close replications based on multiple sets of data have been called for, but few such series have appeared (Ehrenberg 1995). Investigations of empirical generalisations in marketing include studies on multiple data sets covering:

1. the diffusion of new products (Bass 1995; Mahajan, Muller et al. 1995),
2. consumer knowledge and product knowledge (Alba and Hutchinson 1987),
3. market evolution and stationarity (Dekimpe and Hanssens 1995),
4. price elasticity and sales promotion (Blattberg, Briesch, and Fox 1995),
5. firm’s strategic orientation (Boulding and Staelin 1995),
6. market share and distribution (Reibstein and Farris 1995),
7. market entry order effects (Kalyanaram, Robinson, and Urban 1995),
8. brand extensions (Aaker and Keller 1990; Bottomley and Holden 2001)

Six of these eight studies (cited above) were published in a special issue of Marketing Science in 1995. Only a few of the above studies can claim congruent, critically uncontroversial or unchallenged conclusions that have been accepted universally by academics and practitioners (Bass 1995; Blattberg, Briesch et al. 1995).

Controversially, it has even been posited that the relative scarcity of successful replication studies is related to the poor relative performance of academic ‘experts’ at predicting consumer behaviour compared to practitioners, novices or even just chance (Armstrong 1991).

Clearly marketing knowledge (and possibly marketing academic reputation) would benefit from more serious and sustained research in empirical generalisations by marketing researchers, especially research using multiple datasets.

**The inferential framework for brand extension empirical generalisation is questionable**

In addition to the two studies (Aaker and Keller 1990; Bottomley and Holden 2001) employing multiple datasets of close replications for empirical generalisations in brand extension theory, there are six more close replications that employ just one dataset or study

Many of these attempts at empirical generalisation of brand extension employ Residual Centering (Lance 1988) prior to employing Ordinary Least Squares (OLS) regression to deal with model co-linearity. This approach is not usual with respect to mainstream statistics literature and few academic literature streams use this approach uncritically as a method for dealing with co-linearity amongst predictor variables:

1. psychology (Federico 2004; Lance 1988),
2. sociology (Brown, Brody, and Stoneman 2000; Weeks 1992),
3. marketing, mainly in the area of brand extensions (Bottomley and Doyle 1996; Fu and Saunders 2002; Holden and Barwise 1996; Nijssen and Hartman 1994; Völckner and Sattler 2003), but also in the areas of new product development (Stump, Athaide, and Joshi 2002) and
4. management, in the area of new firm (start-up) survival (Sarkar, Echambardi, and Agarwal 2000).

All studies in the above list employ residual centering. The application of residual centering of interaction means in ANOVA studies has also been recommended in the context of marketing research (Ross and Creyer 1993).

One of the authors in the fourth study listed above (Echambardi) contributed to a more critical evaluation of Lance’s residual centering technique with conclusive results showing that it biases the main effects estimates in the presence of interaction effect specifications and therefore should not be used to remove multi-co-linearity amongst main effects and interactions in Ordinary Least Squares (OLS) regression model analysis (Echambadi, Arroniz, Reinartz, and Lee 2004).

Eight key studies involved close replications of Aaker and Keller’s brand extension experiment (Alexandre-Bourhis 1994; Barrat, Lye, and Venkateswarlu 1999; Bottomley and Doyle 1996; Fu and Saunders 2002; Holden and Barwise 1995b; Nijssen and Hartman 1994; Patro and Jaiswal 2003; Sunde 1991). If two-, three- and four-way interaction effects amongst complementarity, substitutability, transferability and difficulty, hereafter described as ‘intra-fit’ interactions, were significant in predicting extension preference, but were omitted from
the model specifications, then the hypothesis inferences would be deemed to be subject to specification bias (Hair, Anderson, Tatham, and Black 1998). However none of these eight studies controlled for intra-fit interactions during the modeling and subsequent inference generation process, despite the exploration of this idea by the authors of the seminal study (Aaker and Keller 1990). This leads to doubts about unaddressed model specification bias in all the close replication studies.

Also, only seven researchers over four studies (Echambadi et al. 2004; Mather 2005; Mather and Lawson 2003; Mather and Sunde 1998) conducting brand extension empirical generalisation studies have employed a General Linear Mixed Model (GLMM) approach and only one (Mather and Sunde 1998) has employed the likelihood ratio test (Self and Liang 1987) based on the Chi-square distribution of the differences in maximum likelihoods between related models widely accepted in mainstream statistics literature to robustly infer individual effect significance in the presence of co-linearity.

Finally there is the assertion (Bottomley and Holden 2001) that some published inferences from the eight key close replication studies were biased by basic procedural errors.

6. Overarching Questions and Related Literature

How do the competing fundamental theories of memory and judgment relate to brand extension theory?

Aaker and Keller’s paper (Aaker and Keller 1990) posited that positive attitude towards a brand extension was influenced by the four main-effect brand extension ‘fit’ attributes, namely complementarity (of the extension with the original branded product), substitutability (of the extension for the original), transferability (of the firm’s resources and capabilities from the original to the extension) and difficulty (perhaps best explained as the opposite of: the triviality of the firm making the extension) and the *indirect* effect of original brand quality moderated by the ‘fit’ attributes. Although widely cited and used in many subsequent replication studies, the Associative Memory Network Model that it is based on is not the only relevant theory for human brand judgments and brand and category similarity (or typicality) judgments. A competing and parallel literature stream of Brand Categorisation comprising two typicality constructs, namely (a) typicality of the extension of the brand and (b) the
typicality of the brand of the extension category, in addition to a (c) parent-extension similarity concept and (d) effects of category hierarchies on evaluations involving prototype and exemplar memory constructs. This literature is logically integrated into an argument for convergence of both main streams of literature into a hybrid theory of extension fit.

**Why has there been so much disagreement about brand extension theory?**

Aaker and Keller’s paper has stimulated numerous attempts at replication with little success (Bottomley and Doyle 1996; Holden and Barwise 1995a). Replication failures have been frequently attributed to (a) excessive covariance or co-linearity amongst model effects, (b) unmeasured important effects and (c) some procedural errors. Only a few studies on the theory of brand extensions have reported investigation into other possible causes such as (a) heteroskedasticity (excessive variation in variances and covariances amongst either effects or study locations), (b) weak effects relative to effect variances, (c) lack of pooled data inferences or (d) bias due to inappropriate residual centering (Echambadi et al. 2004; Mather and Lawson 2003; Mather and Sunde 1998). Whilst encouraging, these studies are limited to addressing only a subset of the full set of problems outlined here.

**How can General Linear Mixed Models help empirical generalisations?**

The applicability of this model framework to pooled data from multiple studies has been highlighted in the marketing literature (Bass and Wittink 1975). In the statistics literature, the ability to correctly analyse data with significant covariance structures (McCullagh and Nelder 1989 pp 437-50; Wolfinger 1993) within an appropriately broad or population-wide inference space with randomly distributed unmeasured effects confounded with the study location (McLean, Saunders, and Stroup 1991; Zeger, Liang, and Albert 1988) has been highlighted.

**How might Reliability Analysis help Brand Extension Theory?**

In the context of brand extension empirical generalisations, it is natural to ask: “Do the ‘fit’ question items share a common dimension?”. The usefulness of Internal Consistency or Reliability Analysis to ensure unidimensional additivity is well documented (Cronbach 1951). Much of the discussion in the existing literature on brand extension theory generalisation blames the lack of agreement, in part, on co-linearity amongst the three measures of the ‘fit’ dimension, namely complementarity, substitutability and transferability. Support for a single
dimension behind the three ‘fit’ items, if achievable, could be employed in subsequent inference frameworks to eliminate this as a potential source of model mis-specification bias. A lack of support for a single fit dimension would indicate support for a more structured theory of brand extension fit evaluation.

What informs successful brand extensions?

Direct (i.e. main) and indirect (i.e. interactive) effects of perceived quality of the original brand and the three ‘fit’ attributes of complementarity, substitutability and transferability were initially posited (Aaker and Keller 1990) to explain most of the preference for the brand extension. Their findings were that only one main or direct effect, transfer, and only two of the three quality-fit interaction effects, specifically quality x complementarity and quality x substitutability, were significant predictors of brand evaluation. Subsequent close replication studies from that time to the present alternately supported and failed to support the original authors’ conclusions on effect significance in almost equal ratios, effectively reducing the perceived reliability of the original authors’ results by the interested research and practitioner community. The ‘ideal point’ fit relationships arising from a parallel stream of relevant categorization theory appear not to have been fully integrated into this line of research. So it seems the answer at this stage of this thesis to the question “What informs successful brand extensions?” is “We are still not sure.”.

7. Why an Empirical Generalisation context was chosen for this research

Empirical generalisations in brand extension theory: a lack of progress.

Empirical generalisations have been the backbone of scientific progress and the subsequent growth of human knowledge for millennia. An early discussion of the empirical generalisation of brand extension theory (Leone and Schultz 1980) appears not to have encouraged many studies on this topic that address the published uncertainties effectively. Empirical generalisation in this literature stream is still elusive despite several subsequent
attempts (Bottomley and Doyle 1996; Bottomley and Holden 2001; Echambadi et al. 2004; Mather and Lawson 2003; Mather and Sunde 1998).

**Opportunity to generalise inferences from pooled data**

Nine data sets from five relevant studies have been made available with permission by the original authors for further research (Holden and Barwise 1995b). In addition, three other close replications of the original study (Fu and Saunders 2002; Nijssen and Hartman 1994; Patro and Jaiswal 2003) have been published and the authors might reasonably be approached for permission to include their data in a larger generalisation study. Thus there is a good opportunity to employ General Linear Mixed Model inference framework on the pooled data from the eight studies.

The contribution that this thesis will make arising from this opportunity to generalise using pooled data is to address the gaps in the existing literature surrounding the disagreements amongst the original Aaker and Keller brand extension findings and those of the close generalizations. This study will explain the reasons for the failure of the close replications to agree, elucidating the brand extension theory and demonstrating the key role of general linear mixed models in empirical generalisations.

**8. Outline of thesis**

Chapter two of this thesis therefore critically examines and discusses five main streams of literature.

The first literature stream introduces the background to empirical generalisations in marketing from an historical context, highlighting its importance in scientific knowledge generation, especially as a precursor, and not simply an adjunct, to theory formation and hypothesis testing.
The second literature stream surrounds both the formation of consumer brand value and the transference of that brand value to new products associated with that brand, i.e. brand extensions based on the Associative Network Memory Model. This stream also includes the majority of attempts at empirical generalisation of brand extension success.

The third literature stream surrounds both the formation of consumer brand value and the transference of that brand value to new products associated with brand extensions based on the Categorisation Model. This stream includes extensions that include category hierarchies, Mandler’s theory of ideal point fit behaviour, prototype and exemplar memory concepts, and fit measure dimensional concepts of congruency, similarity and typicality.

The fourth stream surrounds the conceptual development and practical implementation of inference generating model frameworks that facilitate and enhance the ability to empirically generalise a pattern of cause and effect.

The fifth stream surrounds posited effects on consumer evaluation of brand extensions other than the direct and indirect effects of original brand quality and fit, congruence, similarity or typicality that are components of the first two literature streams. These additional effects can be relevant to consumer evaluation at different parts of the marketing process, such as promotion, or at the point of re-evaluation of the original brand as subsequent to exposure to a brand extension. These additional affects are often cited in the first or second literature stream as confounding effects, and so require more careful study. In addition, our understanding of these additional effects also may be informed by the primary findings of this thesis.

The results of the literature research are synthesised and implemented in the subsequent chapters to improve the quality of inferences generated about brand extension success.

Chapter three outlines research questions derived from the literature and formulates testable propositions.

Chapter four of this thesis develops a rationale for the choice of analysis frameworks and techniques to support the research aims.
Chapter five discusses the data used in detail, and presents the results of some preliminary analysis in order to qualify the data as suitable for further analysis in the subsequent chapter by investigating the reliability of measures of the salient dimensions of ‘fit’ and analysing and discussing the levels of contrast or balance in effects achieved in the original and close replication experiments.

Chapter six presents the results of the main analysis, with several series of general linear mixed models fitted to pooled data, with and without (a) the intra-fit interaction specifications, (b) the ideal point second order effect specifications, (c) with and without the ‘difficulty’ effect item in the close replication design. This implementation is based on more careful and appropriate analysis of multiple existing studies’ data, integrating the General Linear Mixed Model inference-generating framework and additional literature theory to further extend the hypotheses to be tested.

Chapter seven discusses and analyses the results presented in chapter six, interpreting the results into the main findings, primarily on the status of the original brand extension theory and Mandler’s ideal point response theory, and secondarily on the usefulness of empirical generalisations and general linear mixed model analysis frameworks for marketing research. This successfully identifies the empirical generalisation of salient factors affecting brand extension preference formation.

Chapter eight discusses the conclusions that can then be drawn from this research and the implications of these findings on (a) the subsequent interpretation of the existing relevant literature, (b) the implications for marketing practitioners and (c) the directions of future research in this area.
Chapter 2 Literature

This chapter will critically examine and discuss the five main streams of literature relevant to the research questions as outlined in chapter 1.

The first literature stream introduces the background to empirical generalisations in marketing from an historical context, the second and third streams surround the formation of consumer brand value and the transference of that brand value to new products from two differing conceptual bases, and includes the main attempts at empirical generalisation of brand extension success, the fourth addresses the conceptual development and practical implementation of relevant inference generating model frameworks and the fifth discusses other posited effects on consumer evaluation of brand extensions.

The results of the literature research will be synthesised and implemented in the subsequent chapters to improve the quality of inferences generated about brand extension success.

1. Empirical Generalisations from Physical to Social Sciences and Marketing

What is an empirical generalisation?

An empirical generalisation is the formation of a law derived from observed patterns about which hypotheses can be formed. These laws are built primarily to describe how phenomena occur, not why, and the bounds of applicability of these laws are generally explicit. This can be viewed as a desirable first phase of scientific investigation and knowledge formation preceding the formal testing of scientific hypotheses. In other words, empirical generalisation can be conceived as the process of identifying a pattern or law-like behaviour from a group of observations, as a desirable activity preceding the construction and formal testing of competing causal hypotheses that may explain that previously identified, reproducible pattern of behaviour. This is by no means limited to either animate or inanimate objects or any
particular area of science, and should be equally applicable amongst physical, biological and social science disciplines.

In order to justify an approach to addressing the research questions presented in Chapter 3, it is necessary to review aspects of the history of philosophy.

**The earliest known empirical generalisations from the Sumerians**

The oldest existing documented empirical generalisations are in Sumerian cuniform on clay media and concern empirical generalisations in medicine based on careful observations of human illness and effective treatments, dating from about 2000 years BC (Labat c.1600 BC).

The Sumerian medical treatise is organized by general symptom type, and, where that part of the treatise is fully preserved, the observations recorded cover most of the expected human disease symptoms for that type. The texts exhibit the ability of the anonymous authors to observe and report patterns in symptoms and treatments and are perceptive and cogent in their recommendations. The observation and treatment of excessive bleeding is close to current practice for both techniques and medication, as the plants described are readily identified as natural sources of current synthetic drug treatment analogues (Labat c.1600 BC p xxxviii).

We can note, then, that a significant portion of common human illness and treatment follows a pattern, or law of behaviour, that has been documented on permanent, interpretable media, for about 4 millennia.

Not only do these ancient texts contain ancient empirical generalisations, their carefully recorded human medical observations continue to contribute to modern empirical generalisations and furthermore aid in the development of causal hypotheses that naturally follow. For example, current empirical generalisations and associated causal theories concerning viral haemorrhagic fevers are being enhanced by those ancient observations by exhibiting concordance with contemporary epidemiological observations of the symptoms, effective and ineffective treatments (Coleman and Scurlock 1997).

Empirical generalisations have been historically the first stage of the development of what is now known as scientific thinking, which formalizes not only the ‘how’ or pattern recognition
part of the scientific process, but extends also to encompass the ‘why’ or hypothesis testing and confirmation part as well.

**The philosophy of science and empirical generalisations of the ancient Greeks**

Ancient Greek philosophers, especially Plato and Pythagoras, continued the development of what is now known as the philosophy of science.

Pythagorean philosophy developed broadly in line with Platonic ideas. It first explicitly states the importance and role of empirical generalisation with the seven cardinal doctrines of Pythagorean philosophy, reproduced here as a quoted excerpt (Mourelatos 2004):

1. The fundamental realities of the world are structural and mathematical.
2. These structures constitute not only what is more fundamental but also what is normatively better, what is aesthetically more beautiful, and what displays greater simplicity, regularity, and coherence in its mathematical proportions or parameters.
3. Structures in superficially dissimilar contexts can be basically the same. Indeed, there is a pervasive affinity or sympathy between the inanimate and the animate, between man's psyche and the whole cosmos.
4. This cosmic sympathy affords the possibility of moral improvement through a patterning of the individual psyche on the cosmos.
5. Beyond moral improvement, the cosmic sympathy affords the prospect of ascent to a trans-human level of existence, even to immortality, through a process of purification. Correlatively, it also poses the threat of descent into an infra-level of existence.
6. Knowledge or understanding is inherently mystical and can be attained only by the elite.
7. The study of mathematics is the indispensable basis for all intellectual and spiritual progress.

Clearly the Pythagoreans based their beliefs on the importance of recognizing patterns and describing those patterns in mathematical terms. Several significant sets of astronomical observations and subsequent laws of astronomical behaviour were generated by Greek mathematicians and philosophers during the two and a half centuries before and after the birth of Christ.
Plato (427 B.C. – 327 B.C.)

Plato expounded the importance of the link between philosophy and mathematics in ‘The Republic’ by explaining how language facilitates learning and clarifies thinking by possessing persistence in the relationship between physical objects and cognition about those objects that we call thought. He generalised mathematical entities as possessing two recognizable characteristics: they are entities that can be learnt and taught, and they are learned by a reorganization of existing knowledge (Plato, Cary, Davis, and Burges 1850 p 29). This is now known as the separability theorem, more completely defined as the ability to separate abstract meaning from concrete instance.

In the Greek of Plato’s era, ‘ta mathemata’ meant ‘the mathematical’, and ‘mathesis’ meant both ‘learning’ and ‘science’, and mathematics was held as the ideal mechanism to support both.

Plato also demonstrated via the Socratic dialogues that there was difficulty with converting qualitative sociological measures like ‘good’, ‘better’ and ’best’ into the desired mathematical number-measure and ideal form, and was therefore frustrated with the problems of philosophical argument and conclusion in related domains of morals and ideal behaviour. Despite his considerable efforts to bring this to his readers’ attention, little effective progress was made with this problem until the work of Guttman, Rasch and Mokken in the 20th century (McKay, Schofield, and Whiteley 1983 pp 120-22; Shye and Guttman 1978 pp 280-84,88; Wright 1988a). Plato’s philosophical framework of the Socratic dialogues shows also that there were considerable problems at that time, and in that framework, of dealing with significantly stochastic processes, i.e. processes with significant error components confounded with the measures. This issue wasn’t addressed again until Gauss used calculus to develop the ordinary least squares regression approach to data analysis and subsequently gave a real meaning to the ‘mean’ or average measure of an item (Bühler 1981 pp 138-41).

Plato discussed the Greek usage of ‘rational’ and made it explicit that the meaning was closely tied to a ratio. The root meaning of ‘rational’ is the ratio of probabilities of an idea being true versus the probability of the idea being not true, in other words a ‘logit of truth’. This ‘logit of truth’ as the core of rational scientific philosophy appears not to decline in
intensity and not become well re-integrated into the progress of scientific philosophy until the 20th Century.

The intellectual culture of Platonic philosophy described above led Aristarchus of Samos (310-230 BC) to form, from his own careful astronomical planetary observations, the astronomical law that the Earth orbited the Sun. Eratosthenes (276-196 BC) gained more acceptance with his observations concerning a pattern of angles of shadows at the summer solstice, and the distance between Alexandria and Cyrene (now Aswan). The resulting deductive measurement estimate was the radius of the Earth. Ptolemy (85-165 AD) summarized Greek astronomy (Ptolemy 1955 p 465) and his most famous work, ‘The Almagest’, along with Euclid’s ‘Elements’ comprise the scientific texts of longest continuous use. This is likely because it reproduced a large number of objective astronomical measurements that subsequent astronomers found useful for investigating patterns of behaviour.

It is instructive to note the importance of observational measurement accuracy in empirical generalisations. The geocentric model of planetary motion (Ptolemy, Copernicus, and Kepler 1955 p 1) required a more complex and sophisticated level of mathematics to support it than the level of mathematics required to support the heliocentric model (Dibner 1965 pp 19-20). Greek mathematics had developed sufficiently to do so. The competing heliocentric model required greater observational measurement accuracy before it received even limited support.

Classical Developments of Empirical Generalisations: Renaissance Europe

Little permanent record of the further development of scientific process or philosophy survives from the end of the era of Greek science and philosophy around the fifth century AD until the start of the Classical era around 1500 AD. This lack of documented progress in science promoted the misnomer of ‘The Dark Ages’ to describe that intervening period.

---

1. However his heliocentric model of planetary motion was widely rejected due to the less critically evaluated but more widely held beliefs surrounding the Greek gods and planetary bodies. Aristarchus’ observations were reused and replicated and his heliocentric model reinstated in subsequent studies over a millennium later.
2. despite supporting the geocentric model of planetary motion
3. now rejected
4. and eventually successful
5. and this did not evolve until the next millennium after Ptolemy, in the Classical Era
Copernicus

Copernicus’ work was part of a rekindling of scientific interest and thought based on empirical generalisations. Building on the rediscovered, retranslated and newly republished texts of the Greek astronomers, and augmenting his deductions with his own careful observations, Copernicus re-formed, revised, extended and published a set of astronomical laws of planetary motion known as the heliocentric model. This refined model is still the current basis for our laws of planetary motion.

Observations of Copernicus and his method of working by his student Georg Joachim Rheticus highlight the importance of empirical generalisation in his work:

“…my teacher always had before his eyes the observations of all ages together with his own, assembled in order as in catalogues; then when some conclusion must be drawn or contribution made to the science and its principles, he proceeds from the earliest observations to his own, seeking the mutual relationship which harmonizes them all; the results thus obtained by correct inference under the guidance of Urania⁶ he then compares with the hypothesis of Ptolemy and the ancients; and having made a most careful examination of these hypotheses, he finds that astronomical proof requires their rejection; he assumes new hypotheses, not indeed without divine inspiration and the favour of the gods; by applying mathematics, he geometrically establishes the conclusions which can be drawn from them by correct inference; he then harmonizes the ancient observations and his own with the hypotheses which he has adopted; and after performing all these operations he finally writes down the laws of astronomy…”

(Dibner 1965 pp[19]-[20])

Here we can see an early documented example of an astronomer striving towards ‘astronomer-free’ celestial item measurements, which is an extremely important concept as discussed later by Guttman (Shye and Guttman 1978) and Rasch (Wright 1988a).

Copernicus put the geocentric celestial model ‘before’ the data, i.e. he tested a given theoretical model agnostically against the believed, given, observed, empirical data, and

---

⁶ Urania was a mythical figure considered the muse of astronomy
rejected it, in the mode of ‘modus tolens’ or taking away. Popper, as discussed later, would have approved.

Copernicus’ alternative heliocentric theory was a law-only, or an empiricist, theory, not a causal hypothesis, as he did not have available a theory of gravity and the subsequent laws of orbital mechanics, identified later by Newton, to explain why the celestial bodies under observation adhered to their orbits.

The increasing rate of empirical generalisations in science

Galileo Galelei (1564-1642)

Galileo Galelei (1564-1642) re-emphasized the need and methodology for empirical generalisations (Galilei 1963 p 265, 1981 p 5; Galilei, Grandi, Bresciani, and Buonaventura 1718 p 265) and published discussion supporting the heliocentric model of planetary motion. Discourses and mathematical demonstrations concerning the two new sciences documents more fully the important empirical generalisation on acceleration of bodies in free fall:

“The time in which a certain distance is traversed by an object moving under uniform acceleration from rest is equal to the time in which the same distance would be traversed by the same movable object moving at a uniform speed of one half the maximum and final speed of the previous uniformly accelerated motion.”

He continues to state his famous result that the distance that a body moves from rest under uniform acceleration is proportional to the square of the time taken (Galilei 1914).


---

7 Copernicus’ heliocentric celestial model, however, had limited acceptance due to the powerful socio-theological dogma of the Holy Roman Church.
8 .Despite Galileo’s initial wariness and at significant cost to his career, reputation, personal freedom and possibly his health as well. The penalties imposed by the Holy Roman Church due to his published position on the heliocentric model included imprisonment, which was converted to house arrest, and censorship. This forced him to smuggle his last work cited above out of Florence to Holland for publication.
Issac Newton (1642-1727)

Newton (1642-1727) researched and published laws of motion and gravity, invented the reflecting telescope, wrote a book on empirical generalisations in the field of optics, observed the behaviour of visible light spectra and generated more empirical generalisations leading to a theory on the particle nature of light, and invented the branch of mathematics known as calculus to help in work with motion and gravity. He held a general disregard for hypotheses testing⁹ and preferred to focus on accurate measurement, observation and establishing laws of behaviour, a view which is now generally classified as “positivism” (Fisher 1988).

David Hume (1711-1776)

Hume (1711-1776), influenced by John Locke, strenuously argued for an empiricist approach to philosophy of the natural sciences, often citing the benefits to science of the empiricist approach taken by Newton, Gallileo and Copernicus. Hume summarises Newton’s empirical generalizing approach: “It is confessed that the utmost effort in human reason is to reduce the principles, productive of natural phenomena, to a greater simplicity, and to resolve the many particular effects into a few general causes, by means of reasoning by analogy, experience and observation” (Hume 2007 p 32) while citing Newton’s “Rules for Reasoning in Philosophy” (Newton and Huygens 1952 pp 270-71). Hume then underlines the centrality of empiricism to the progress of natural science: “Astronomers had long since contented themselves with proving, from the phenomena, the true motions, orbit and magnitude, of the heavenly bodies, till a philosopher, at last, arose, who seeing, from the happiest of reasonings, to have determined the laws and forces by which the revolutions of the planets are derived.” (Hume 2007 p 77). Here, Hume refers to Newton and his empirical generalisation that is the law of gravity: that the force between two objects is proportional to their masses and inversely proportional to the square root of the distance between them.

⁹ Ironically a causal theory he posited, for all the wrong reasons, and despite his general disregard for causal theories, about the corpuscular nature of light was eventually substantiated by Richard Feynmann’s Quantum Electro-Dynamic theory of light energy as particles in the mid 1970s.
2. Empirical Generalisations and theory development

Empirical generalisations thus have been a very successful mechanism for the development of scientific understanding. Texts in the physical sciences domain consist almost entirely of statements of laws derived from empirical generalisations and their corollaries and examples. An exhaustive list of empirical generalisations in the general scientific domain would be a very long publication indeed, and is posited as the basis of western science (Kline 1954 pp 22-27).

Kline’s work reiterates this theme in the scientific philosophy of theory development: The ‘how’ of phenomena must logically first be established before attempts are made to establish the ‘why’ of them. Even now, science is still lacking a complete model of why gravity exists, but at least physicists are not still in general disagreement as to how gravity works. The clear implication for current research resources in this area is that they are not wasted on trying to explain the ‘why’ of phenomena that may not be a sufficiently well defined ‘law’ of ‘how’ and may not be empirically generalisable.

In contrast to the success of empirical generalisations in the domains of the physical and biological sciences, relatively few empirical generalisations have been successful in the social sciences.

The emergence of social science and Auguste Comte (1798-1857)

Implicit in the evolution of scientific thought and literature until the early 1800s was the axiom that data of science are absolutely objective, in that data exist in an eternally ‘given’ state, and are always factually self-evident. This assumption was explicated as part of a repositioning of scientific philosophy to integrate the social sciences, called ‘positivism’ (Comte and Gillespie 1851). This provided an expanded frame of reference and vocabulary for a more fundamental and rigorous examination of the basis and structure of scientific philosophy from then on. Comte, with his advocacy of positivism, also explicitly argued for a taxonomy or lineage of sciences based on average size of measurement error, observability and a ‘metaphysical’ or ‘spiritual’ level of content amongst the sciences.
Auguste Compte advocated bottom up or disaggregate observational measurement and theory testing for all sciences in his taxonomy up to Biology, but dealt with the presence of significantly stochastic processes with a recommendation for a top-down, or subject aggregate approach to data modeling and theory testing for his newly delineated science of Sociology, within which he placed political science and psychological science. Auguste Compte’s philosophy was linked to the French mathematicians and scientists of his milieu such as Descartes, Fermat, Pascal and Laplace. He originally coined the term ‘social physics’ to reinforce the importance of recognizing objective patterns in the social sciences (Ball 2002).

Compte’s positivist philosophy approach for the social sciences was very much oriented towards how society shapes and influences the behaviour of the individual, and is also known as the ‘macro-sociological’ approach to sociological philosophy.

Multiple hypotheses on the “how” arise in academic literature naturally due to varying competencies in areas such as measurement accuracy and level of mathematical development, as with the laws of planetary motion. Associated with this multiplicity of competing hypotheses, the heightened importance of testing competing, multiple hypotheses against empirical data for robust inference generation has been formally recognized since 1890 (Chamberlin and Raup 1995).

Social science positivism and Emile Durkheim (1858-1917)

Emile Durkheim was a positivist social scientist following August Compte’s philosophy of science, and asserted the ‘brute giveness’ of data he analysed, i.e. he posited that there was no ambivalence of interpretation allowable in the examples he put forward. He mirrored Newton’s disdain for hypothesis formation, recommending a focus on pattern recognition and law formation. He strove for measurement accuracy, and adopted and recommended the ‘Modus Ponens’, or mode of ‘adding to’ theories where there was some statistical support from the diffusion of Ordinary Least Squares regression theory.
He is also identified with the development of “Functionalism”, also known as “consensus” theory, which recommends not addressing conflicts in society but rather projects ideal harmonious relationships (Durkheim 1938).

**Causal theories, the social sciences and Norman Campbell (1880-1849)**

Norman Campbell was a physicist who challenged the Positivist’s satisfaction with non-causal, behavioural ‘laws’ and encouraged causal theories, i.e. theories with causes and explanations of behaviour, to be posited and tested for the advancement of knowledge and learning. He also aided the future development of measurement science and the scientific philosophy of science in general and unwittingly social science in particular by making explicit the fundamental requirement for scientific measurement. This fundamental requirement is that if two measures X and Y are to have the property of additivity then the combination of X and Y has to be greater than X or Y and also that X and Y concatenate to produce an X+Y result. This is the same as a physical addition operation on X and Y. His contribution to social science was unwitting in the sense that he couldn’t conceive, and didn’t believe, that additivity could ever be established for the types of measures typically used in the social sciences (Wright 1988b).

**Inference generation and Fisher**

Ronald A. Fisher (1890-1962) contributed greatly to the experimental design and analysis of agricultural and biological experiments in particular and scientific experiments in general but also contributed to the debate on scientific reasoning and philosophy (Fisher 1956 pp 8-36; Fisher 1988). Like Issac Newton, R.A. Fisher held a general disdain for studies surrounding theories of the “why” and instead held an enthusiasm for observing, delineating and reproducing the “how” of phenomena. He established that the discoveries of the geneticist Gregor Mendel could be shown to support Charles Darwin’s theory of evolution (Fisher 1930 p ix). Whilst primarily interested in supporting inferences in agricultural scientific research with statistics, he also contributed greatly to the scientific process and applied statistics in general, and occasionally offered advice to the social sciences. His recommendations for the social sciences included a call for researchers to pro-actively test multiple hypotheses of
significance of effects and encouraged social scientists to accept those factors whose statistical significance in appropriate regression and ANOVA models surpassed ‘reasonable’ levels of confidence, which he suggested as a 95% confidence limit.

This is consistent with the positivist or ‘modus ponens’ approach, literally in Latin the ‘putting-in mode’, which may explain why he appeared to ignore the relativism debate in scientific philosophy associated with shared understanding and meaning and assumed data ‘giveness’. Fisher recommended all his analysis techniques to the social sciences, apparently ignoring the typically greater precision of biological measurements compared to social science measurements.

Quantitative data of the sort used in Fisher's original examples come from studies in biology or agriculture, and are read off calibrated instruments with an invariant additive structure. Quantitative data of the social sciences are almost never of that sort. The data of social research are usually made up of numbers simply assigned to qualities with weak or no attempts made to justify the assignment through the establishment of structural correspondences between the conceptual measurement apparatus and the phenomenon measured.

**Type I errors and R. A. Fisher**

Classical inferential statistics associated primarily with this positivist approach are identified as “Type I error” statistics (Baird 1981). These are designed to aid in reducing the likelihood of incorrectly failing to reject the null hypothesis.

R.A. Fisher’s genius of innovation in statistical data analysis was the Maximum Likelihood method of modeling. This modeling method changes the model parameter estimates such that the estimates maximize the probability that the observed empirical data comes from the model. This is in contrast to the Ordinary Least Squares (OLS) method which changes the model parameter estimates to minimize the sums of the squared error between the model and the data (Bühler 1981 pp 138-41).
In a sense, Fisher’s Maximum Likelihood (ML) method puts the model ‘ahead’ of the data, rather than Gauss’s Least Squares method which in the same sense puts the data ahead of the model. With this innovation, Fisher discovered, but did not appear to fully realise the significance of, estimation sufficiency, which was to become central to scientific philosophy (Wright 1988a).

Estimation sufficiency may be observed when a response to a (posited) predictor variable is modeled with both a binary logit model using a maximum likelihood estimator (after Fisher) and an ordinary least squares estimator (after Gauss).

Where there is a mismatch between the model and the data, the empirical data might be in a state of, say, a random distribution of 0s and 1’s, unrelated to the predictor variable. In this case, a maximum likelihood estimator will correctly fail to converge due to a ‘flat’ or completely stochastic process likelihood and determine the data ‘insufficient’.

Conversely, the least squares estimator is ‘blind’ to the ‘insufficiency’ of the data and may yield an incorrect, and in this case also inappropriate, parameter estimate for the linear model. This is a Type I error.

Recent work (Nelson and Startz 1990) has shown that the t-statistic for the incorrectly specified Ordinary Least Squares model effect may be biased upward for this ‘weak instrument’ scenario and give a spurious significant P-value. This is another example of a type I error.

This means that Ordinary Least Squares model inferences are more at risk to Type I errors than Maximum Likelihood model inferences. Hence R. A. Fisher’s maximum likelihood estimator is an extremely useful tool when considering the progress of science and social science in particular, and the ‘modus tolens’ approach discussed below.

Estimation sufficiency, whilst appearing primarily of a general statistical nature is much more important and more relevant to the social and biological sciences than to the physical sciences in general due to the much higher average ratio of the data measurement error to the underlying parameter effect coefficient value in the social sciences and, to a lesser extent in general, biological sciences, when compared to the physical sciences. Often, in the physical
sciences, experimental data measurement apparatus can be arbitrarily improved or redesigned such that measurement error becomes vanishingly small. To a large extent, this is simply not possible or practical in the social sciences, hence the increased importance of the statistical inference generating frameworks that aid in reliably identifying and confirming patterns in the data.

**Social science theory development, type II errors and Karl Popper**

Similarly, although Karl Popper was not primarily a social scientist, his work and the conclusions and recommendations that naturally flow from his work have far more importance to the social sciences than to other sciences. This is due to the relative excess of theory generation followed by single study, statistical inference based, theory testing and the low frequency of successful pattern or ‘law-like behaviour’ identification in the social sciences compared to the other sciences (Meehl 1990). This excess of empirically ungeneralised theories demanded a response and Karl Popper provided a very helpful point of view from which to approach this problem.

Karl Popper's philosophy of science promotes ‘modus tollens’, literally the Latin for ‘taking-away mode’, as the method of disconfirming, or falsifying, scientific hypotheses. Scientists are encouraged to start with a current scientific theory and employ deductive reasoning to derive specific conclusions, some of which are predictions. Deductive reasoning is truth preserving, in that it if one starts out with true premises, one can only deduce true conclusions.

Starting with a theory and deducing predictions can be stated as a premise: If the theory is true, then the deductively reasoned prediction is true. Popper showed that one cannot prove a theory is true, but that one can show that a prediction is false. If scientists test one of these predictions and find out that it is not true, allowing for random residual and measurement error, they use ‘modus tolens’ to conclude that the theory cannot be true, or alternatively cannot be true outside some set of boundary conditions (Popper 1959 p. 251).
Classical statistics associated with this approach are identified as “Type II error” statistics, focused on the power of the test, i.e. minimising the likelihood that the test will incorrectly fail to reject the alternative hypothesis (Hair et al. 1998).

If the above approaches are conceptualised as a dichotomy (rather than a continuum) between the extremes of ‘Fisherian’ or Positivist and ‘Popperian’, as is often the case in current literature, then it is still not clear where the ideal philosophical and practical experimenter position lies. However this bipolar conceptualization is flawed and overly limited. It is only with the re-consideration of context, shared understanding and meaning, as developed by Wittengenstein, Heidegger and Hans-Georg Gadamer, that the apparent problem of which is the correct position (Positivist or Popperian) to adopt is resolved.

**Relativism, Phenomenology and their impact on the social sciences**

Aligning with this position led to the development of the ‘Social Action’ branch of sociology as proposed by Max Weber (1864-1921) otherwise known as ‘micro-sociology’. The orientation was from the viewpoint that individuals create society.

Taking this relativistic philosophy further for the social sciences, Shultz in the 1930’s, and later Douglas and Atkinson in the 1960’s and 1970’s, developed the ‘Phenomenology’ branch of sociology, based on the concept that social reality is constructed in the minds of social ‘actors’ or ‘agents’. Phenomenology is therefore philosophically limited to the study of ways in which individuals interpret and create from their own perspectives their own social worlds.

Some of the background to the development of, and argument for, these philosophical approaches to the social sciences carried some unhelpful inductive reasonings and conclusions that would otherwise have retarded the development of other branches of sociology. Included was an overly pessimistic and unhelpful conclusion on the limitations for empirical generalisation concerning both the implication of measurement error and the possibility of confirmed shared meaning and understanding.
These philosophies seem to have fostered beliefs in some readers that could be typified as the (flawed) question: “If an individual’s interpretation of reality is, in general, unique, why should we try to find a shared or common meaning?” The flaw is, of course, that although the assertion that ‘an individual’s perceptions are, in general, unique’ is true, this does not negate the existence of a shared or common meaning.

Widespread adoption of this attitude by social science researchers would have led to a very poor prognosis for social science theory generation and hypothesis testing, indeed!

**The conceptual status of things in language, Wittgenstein (1889-1951) and Heidegger (1889-1976)**

Social research data is often comprised of numbers simply assigned to qualities with no explicit attempt to justify that assignment through the establishment of structural correspondences between the conceptual measurement instrument and the phenomenon measured.

This is contrary to positivist practice in the natural sciences. The shared meaning of the data by the researchers was given at the time of publication. With the changes that have taken place in astronomy with respect to geocentric and heliocentric planetary models in the last millennium, however, the historical nature of what counts as a fact in the natural sciences has been made very evident and has raised the problem of just what facts are if they can change over time.

Philosophy has turned increasingly toward investigations of language in its attempts to resolve this problem. Philosophers such as Ludwig Wittgenstein and Martin Heidegger have changed the philosophy of science by showing that the conceptual status of things in language is of primary interest when the factuality of those things is in question (Heidegger and Anderson 1966). Concepts do not exist in nature, but humans (in addition to some other higher mammalian species) are born into a conceptual world insofar as language is the means through which they learn about life. Thus, scientific researchers in all the sciences become scientific researchers through socialisation and training; the socialization and training are themselves formed historically and conceptually, changing over time, but social scientists’
conceptualisations are more influenced in this way than others. Most social science concepts have no direct tangible instance, and often the concept communication relies completely on language and not on representative geometry, such as is often the case with physical sciences (Heidegger 1968 pp 126-37).

Following this, philosophy has arrived at an explanation of factuality that relies on contextualisation for criteria as to what can count as data. The problem currently is to reconcile this ‘general context-dependency’ with a ‘specific context-independency’. Understandings are situated in general historical, cultural and linguistic contexts, but if these understandings are to ‘mean’ anything, in the sense of another reader understanding a writer’s ‘meaning’, they must be shared in a more specific fashion that is replicable and relatively free from the influence of minor contextual or sample demographic changes.

Study of these issues led Heidegger to re-examine Plato’s reasons for emphasizing the importance of keeping philosophy near to mathematics (Heidegger and Anderson 1966). In his MENO dialogue, Plato makes plain that mathematical entities have two main features: they are those things that can be taught and learned, and they are learned through a reorganization of what is already known (Plato et al. 1850 p 29).

The importance of this adjacency of philosophy and mathematics is highest in the social sciences, where the risk of otherwise failing to apply mathematics appropriately to the important numeric measures and to specify the appropriate problem within the wider context is far greater than in other sciences due to the relatively poor levels of quantitative skill in academic researchers in the social sciences (Meehl 1990).

Heidegger also pointed out that “Most of the propositions and questions of philosophers arise from our failure to understand the logic of our language. They belong to the same class as the question ‘is the good more or less identical to the beautiful?’”. It should not be surprising that “the deepest problems are in fact not problems at all” (Wittgenstein 1922 p 63; Wittgenstein, Anscombe, and Wright 1972). Here he is approaching the cautions of D. R. Hofstadter (Hofstadter 1979 p 470) concerning the limitations of any single logic and reasoning system and admitting the apparent paradoxes they generate. Hofstadter uses Bach’s fugues and Escher’s prints as examples of works that generate apparent paradoxes when analysed by a
single logic or reasoning system but still are intuitively valued and admired as comprising interesting information by significant segments of the population.

**Separation of meaning, Hans-Georg Gadamer (1900-2002) and Paul Ricoeur (1913-2005)**

Hans-Georg Gadamer and Paul Ricoeur raise, review and make explicit important concepts that Wittgenstein and Heidegger left unexamined. They showed that if something is teachable and learnable in the way Plato discussed then the words and signs used to convey meaning must ‘separate’ from that meaning. In other words, in a broad and relative way, words and signs must have a constancy to them that does not depend upon who is using them or upon where or when they are used, as long as everyone has been trained to understand the language involved (Hiley, Bohman, and Shusterman 1991 pp 161-69).

W. P. Fisher points out that this constancy constitutes a specific context-independence that is most evident in mathematics. When a geometrical analysis of a circle is performed, almost anyone is aware that what is being studied is not the property of this particular circle drawn. Rather, it is the general idea of a circle that is invoked, an idea that stands independent of every particular example of it. Thus in the specified context all observers may experience a separation of the symbol, the particular circle drawn with a measurable radius, and the general idea of a circle.

This combination of a general linguistic context-dependence with a specific context-independence is so essential for perceiving or conceiving anything that, with hindsight, it is understandable that it had been overlooked by those who have tried to understand science in the past. Being a fundamental precondition of conceptualisation, it was likely to be ignored until enough incongruent information emerged to highlight the lack of something (Fisher 2003a).

He goes on to highlight that the importance of this conceptualization is highest in the social sciences, where the risk of otherwise failing to separate the mathematics from the numbers, the meaning from the instance and the specific problem from the broader context is far greater.
than in other sciences due to the relatively indirect and abstract nature of entities and attributes in the social sciences (Fisher 2003b).

**Scientific Realism, Constructive Empiricism, Rationalism and Critical Rationalism or Fallibilism**

Much of the debate in the science of philosophy in the last three decades has taken place in a two- or three-cornered argument amongst rationalist dogmatists, skeptical empiricists and scientific realists. More recently scientific realism has emerged as a new ‘ism’ that builds on the strengths of the two former ‘isms’ in a constructive way that enhances the foundations laid by Popper.

Scientific realism demands that the ‘why’ be explained in addition to ‘how’, but insists that basic propositions be the often-debatable ‘self-evident truths’. Alternatively, constructive empiricism accepts the ‘how’ without demanding to know the ‘why’, and does rely on some sort of atomic belief at some basic level. According to this definition, Newton was a constructive empiricist: he explained the motion of the celestial bodies using the law of gravity whilst admitting he could not explain gravity (Musgrave 1982).

“Scientific realism is best defined as an inference to the best explanation of facts about science” (Musgrave 1988 p229).

This leads to a concise statement of what is jokingly labeled by the cited author as “The ultimate argument” for a definition of a philosophy of science:

It is reasonable to accept a satisfactory explanation of any fact, which is also the *best available* explanation of the fact, as true.
F is a fact.
Hypothesis H explains F. More realistically:
Hypothesis H *satisfactorily* explains F.
No available competing hypothesis explains F as well as H does.
Therefore it is *reasonable* to accept H as true.

(Musgrave 1988 p229).
Framing the argument for this position as the only defensible one for scientists to take when evolving and testing hypotheses, Musgrave (1993) explains from a historical perspective that skeptics argue that a belief can only be justified by citing another belief, whereas dogmatists deny this. The two camps are based on the premises that truth or belief arise from either experience (skeptics) or reason (dogmatists). Sense and experience are central to the theory of knowledge and the philosophy of science. This position is also known as empiricism. The dichotomy between these two positions is also classed as empiricism versus rationalism. Rationalists use inductive reasoning until propositions are reduced to self-evident truths. For rationalists, the ‘problem of error’ feeds off dogmatism. Only those dogmatists who think that there is a method for finding the truth are obliged to explain why we do not always find the truth. Otherwise, scientists are free to measure, model and observe error without having to constantly explain why it is so.

The theory of knowledge is a great debate between skeptics and dogmatists. Musgrave (1993) sides with the skeptics mostly, but develops critical rationalism or fallibilism. This leaves scientists in the desirable position of having little or no certain knowledge from a skepticism point-of-view but having plenty of conjectural knowledge, incorporating realism about perception, science and truth.

**Common meaning Spearman, Guttman, Cronbach and Rasch**

Plato’s link between mathematics and philosophy persists in our language. The core concept of rationality refers to the mathematical ratio between what something probably is and what it probably is not, i.e. the logit of truth that goes into every act of recognition and learning. Further, it is common to speak of understanding in terms of ‘getting something straight’, of ‘delineating an idea’, of the ‘dimensions’ of experience, of how things ‘figure out’ or ‘add up’, and of ‘ruling something out’, or ‘putting things in proper order’. The roots of all of these phrases are tied to the platonic concepts of abstract geometry.

Guttman championed the concern for additivity on a scale and shared meaning of item measures, but initially his work was limited in practice to a non-stochastic, or deterministic,
approach. His methods and recommendations required strict adherence of data to common rankings and orderings that became more difficult to achieve in practice as data sets became larger and random errors became more frequent in the data (Shye and Guttman 1978).

Building on Guttman’s ideas, Spearman, as part of his interest in psychological tests of personality traits and abilities, was one of the first social scientists to make a significant contribution towards providing a statistical framework for analyzing item measures in the context of constructing reliable multi-item test scores for the social sciences, given explicit distributional assumptions of the measures (Spearman 1904).

Cronbach (1951) further refined this distributional-assumption-based approach and developed a series of scale item reliability measures, the most commonly used and famous one being what he called the ‘reliability coefficient alpha’, commonly referred to as ‘Cronbach’s alpha’. One of the limitations of these reliability statistics was that they were biased by departures from some data assumptions necessary for valid Ordinary Least Squares regression inferences, namely linearity, homoskedasticity and normality. In this author’s experience these assumptions are not always thoroughly fulfilled nor indeed even documented as having been investigated in contemporary peer reviewed social science journal articles.

Georg Rasch developed and refined a Measurement Model framework that built on Guttman’s (Shye and Guttman 1978) deterministic, or non-stochastic, scalability requirements. Rasch allowed for stochastic departures from the strict rankings and orderings but without assuming or requiring any particular stochastic distribution.

A major benefit of this innovation was that this approach accommodated stochastic or random response errors of any distribution of the type that didn’t ‘make sense’ when interpreting question responses given: (a) the location (or ‘agreeability’) of an individual on an underlying scale or dimension and (b) the relative ‘difficulty of agreement’ of the question item on a multi-item unidimensional scale (Rasch 1966).

---

10 Georg Rasch first developed his measurement theory as part of a consultancy study for the Danish Department of Education, then later independently extended and enhanced his work. Rasch Measurement Models now underpin nearly all school examination grading systems from kindergarten to postgraduate level in the United States.
For instance, an individual subject may rate highly on agreement to the statement ‘I’m disgusted by this concept’ and rate lowly on agreement to the statement ‘I don’t like this concept’. Rasch’s measurement models allow for this to occur in limited, random frequencies with unspecified error distributions and still allow the addition of response measures of these two question items in subsequent analysis to have a valid and useful meaning.

All that is required for additivity and a shared meaning and understanding to be demonstrated is that the overall likelihood of ‘sensible’, or congruent, responses is greater than those that don’t ‘make sense’ in the specific context above.

The importance of these developments is again highest in the social sciences as social science measures frequently contain lower ratios of common or shared meaning to the total variance. This is due in turn to the generally lower levels of direct connectedness between external stated attitudes and internal human cognition processes and generally higher variations amongst individual learnt cognitive and evaluative processing behaviour compared to simpler experimental units typically employed in biological or physical sciences.

Definition of the variable remains incomplete so long as Plato’s sense of the instructiveness of mathematics is left out of our instruments. Data that do not take on a life of their own, separating from the particular respondents to be measured and the particular items used to measure, remain mired in specific context-dependencies that prevent generality from being attained. When Rasch refers to specific objectivity, the separability theorem, sample-free instrumentation, instrument-free measurement or the convergence of items and persons in an measurement model analysis, he is speaking of a realisation of new potentials in the historical possibilities presented to us by science that have remained implicit until now (Wright 1988a).

3. Empirical generalisations in the social sciences and their lack of success
One of the few attempts at exhaustively documenting these laws of social science resulted in a relatively small document that was nevertheless the subject of significant controversy within its domain (Berelson and Steiner 1964 pp 6-7,559-60). The authors noted in passing that although there was not much generally agreed ‘how’ or pattern law, there seemed to be an illogical excess of ‘why’ explanations despite the lack of agreement on the ‘how’ issues, and also a lack of agreement on the appropriate methodologies for conducting empirical generalisations.

Meehl (1990) gives a comprehensive review of how and why this is the norm in social sciences, suggesting ten common masking or confounding factors present in empirical social science studies that are usually sizable, opposed, variable and unknown. Meehl asserts these ten commonly confounding factors are:

1. A loose derivation chain from theories to observations, where interpretations are conditional on prior deductions weakly argued or simply restated.
2. Problematic auxiliary theories, resulting in more confusion as to the state of more theories when hypothesis tests fail.
3. Problematic ceteris paribus (‘other things being equal’) clauses. Empirical generalisations or law-like behaviours do not, and should not, require a controlled, unvarying environment to hold for the behaviour to be reliably observed.
4. Experimenter error, more frequent in published literature than the reputation of many journals would otherwise suggest.
5. Inadequate statistical power, frequently unmeasured in non-experimentally designed studies.
6. Frequent low levels of correlations between all effects in large sample, narrow inference space models, labeled the ‘crud factor’ by Meehl, but possibly ‘clutter factor’ or ‘signal to noise ratio’ are a more meaningful labels from an information processing view.
7. Lack of Pilot studies, defined as identical in design to main studies except for sample size, as opposed to initial studies applied with less rigour than the final study.
8. Selective bias in submitting reports, especially towards either too ‘Fisherian’, reporting all significant effects in a positivist tradition, or too ‘Popperian’ in failing to report significant effects due to perceived prior experimenter biases.

10. Detached validation claims for measurement instruments used, e.g. simply citing use of a measurement instrument in another published article as a justification.

However, not all of Meehl’s confounding factors reflect poorly on the relative skill of social science researchers. They simply imply more focus, skill, effort and other resources are required to overcome the additional obstacles. In particular the high level of measurement error implicit in the analysis of intelligent, complex human systems is rarely encountered in the physical sciences. Also, ethical constraints preclude many designed experimental research approaches more widely applicable in the physical sciences.

It is the last and tenth point above which is possibly the most disturbing as it implies the most basic state of poor scholarship and scientific thinking that appears to be the most prevalent problem with these social science studies. Fortunately both theoretical models found in the relevant literature based on understandings of the underlying processes and the development and study of measurement models, such as common factor or Rasch item scales can be applied to eliminate this problem in an objective manner.

**Empirical Generalisations and Marketing.**

Consistent with the previous discussion, few empirical generalisations in the marketing literature have been universally agreed as successful. One attempt a few decades ago at documenting empirical generalisations in marketing resulted in a conclusion that there were no empirical generalisations to be listed and also that there was a wide disagreement on the appropriate methodologies that should be used for conducting the empirical generalisations (Leone and Schultz 1980). These authors also noted, consistent with the findings in the social science domain as a whole, the many marketing papers explaining the ‘why’ without agreement on the ‘how’. Interest in the state of marketing knowledge and the role of empirical generalisations and theory formation appears to have been rekindled since 1990 with several more authors contributing to the discussion (Armstrong 1991; Ehrenberg 1995; Wright and Kearns 1998). A direct challenge to the dominant paradigm in marketing research of Theory-in-Isolation, or TiL, and a strong call to publish only Empirical-then-Theoretical, or EtT,
research by A. S. C. Ehrenberg stands out as the exemplar of this argument (Ehrenberg 1993). Marketing studies including replications were found to be relatively infrequent in peer-reviewed journals (Hubbard and Armstrong 1994), despite the encouragements for multiple close replication studies proposed in the literature (Barwise 1995).

This renewed focus on empirical generalisations provoked a special issue of the highly regarded Marketing Science journal in which, not withstanding any systematic address to the ten confounding factors cited by Meehl above, marketing empirical generalisations were claimed for the effects of: order of market entry on market share (Kalyanaram et al. 1995), firm strategic actions on firm performance (Boulding and Staelin 1995), distribution on market share (Reibstein and Farris 1995) promotions on sales (Blattberg et al. 1995), marketing mix on short and long run market share (Dekimpe and Hanssens 1995) and time and the proportion of innovators versus adopters on the diffusion of new products (Mahajan, Muller, and Bass 1995). A number of additional examples are provided in the same special edition, together with a proposed summary of desirable properties of empirical generalisations, including multiple sets of independently sampled data, establishing boundary conditions and notably only one example of an empirical generalisation of a ‘high level’ theory (Bass 1995).

More recently a special issue of the Australasian marketing journal has also published several more attempts at empirical generalisations in marketing (Dall'Olmo Riley, Lomax, and Blunden 2004; Mizerski, Miller, Mizerski, and Lam 2004; Romaniuk, Sharp, Paech, and Driesener 2004; Scriven and Ehrenberg 2004) concerning patterns in consumer response to price changes, brand and advertising awareness, luxury brand extensions and lottery purchasing. All of the above are important contributions to the field of knowledge.

However, if this then represents the state of empirical generalisations in the marketing domain a few years ago, then collectively marketing academics have a lot of hard work ahead of them to reach the level of knowledge attained by the physical sciences. To this end a specific and directed shift in marketing research paradigms is required. This shift needs to be away from single sets of data and inferences limited to those single sets towards multiple sets, or sources,
of data and inferences consistent across those multiple data sets (Ehrenberg 1995; Holden and Barwise 1995a).

**Empirical Generalisations and Marketing Knowledge**

One concerned marketing academic (Armstrong 1991) found that the field of marketing knowledge, being closely related to empirical generalisations and summarily defined as the sum of knowledge about the “how” and the “why” of marketing, was so poor that academic marketing experts were no better than practitioners, or worse, relative novices or even worse, chance, at predicting consumer behaviour. Armstrong, Brodie and Parsons (2001) found in a more recent study only very slight improvements in the desirable movement in marketing journal papers away from exploratory hypotheses and single, dominant hypothesis testing to multiple hypothesis testing.

Few studies have explicitly linked the philosophical concepts of marketing knowledge and empirical generalisations. One exception seeks to promote advances in empirical generalisations in marketing by drawing on the enriched philosophical debate in the marketing knowledge literature (Wright and Kearns 1998). They asserted that three main criteria were required for the successful development of marketing knowledge: (a) ensuring falsifiability and theoretical competition, (b) overcoming uncertainty through replication, and (c) using extension to develop generalisations and establish boundary conditions. They conclude their article with a renewed plea that more of the academic research effort in marketing be devoted to replicating and extending existing results, and determining the conditions under which existing theories do, and do not, hold (Wright and Kearns 1998).

**Empirical Generalisations and Brand Extensions.**

One of the areas of marketing knowledge development that has seen some effort expended to produce some of these generalisations is brand extension theory, which includes the concepts of original brand quality and ‘fit’ between the original brand and the extension category and concept.
What is desired is a generalisable theory of how brand extensions are evaluated. This would inform branding and new product development practitioners on how to brand and maximize the acceptance and recognition of new products.

However, a set of apparently conflicting inferences is consistently found in the literature that includes empirical generalisations of the role of extension fit and original brand preference on the evaluation of, or preference for, the brand extension.
4. Consumer evaluations of Brands

The importance of Brand Extensions

An early and popular practitioner text on branding (Ries and Trout 1981 p 127) states that ‘extension is the most significant trend in marketing in the past decade’ (referring to the period 1970-1979). The authors argue extensively against the practice of brand extensions as it can dilute the strength of the brand association, i.e., the extent to which the brand name substitutes for the generic name, such as "Formica" instead of the generic ‘robust decorative laminated thermoplastic sheeting’ or "Band-Aid" instead of the generic ‘small perforated elastic adhesive dressing’. This is consistent with the general theme throughout their book where they advocate a strategic marketing goal of strong original brand positioning in the mind of the consumer.

Note that this is not consistent with the success of the case of the Caterpillar brand extension from heavy earth moving machinery to boots discussed earlier.

They do, however, recommend brand extensions in the following conditions:

1. Small expected volumes from the extension category,
2. low or no competition in the extension category,
3. small advertising budget for the extension,
4. non-significant, commodity extension category products and
5. items sold by sales representatives, i.e. not ‘off the shelf’.

Conditions 1, 3 and 4 above are deduced from the consideration of return on investment, as a completely new brand requires a considerably larger advertising budget as the brand image has to be built from nothing, especially if the target market is a large population, e.g. the U.S. Furthermore, these conditions only yield good returns if the expected revenue is sufficiently
larger than all costs, including marketing communication costs, to be an attractive investment relative to less risky investments such as government bonds and term deposits.

Condition 2 above implies no need to differentiate via branding at all, either as a new brand or as an extension, as there is no alternative and consumer choice is limited.

In condition 5 above, differentiating or branding is less important as the sale does not rely on mass media communications and unassisted consumer purchase at the point of sale. In this condition, the sales representative will target qualifying customers and customize the product design, benefit and alternative comparison presentation to maximize the likelihood of that an individual customer will purchasing the product. In this situation much of the role of the product brand is taken over by the role of the sales person, especially in the modern sales role as a customer relationship builder (Weitz and Bradford 1999). In this condition, the customer depends more on the sales person for the shorthand, cost-and-risk-saving process, and much less on the brand name.

This poses an interesting side question about the corporate brand valuation processes referred to elsewhere in this thesis. Do Ries and Trout differentiate between corporate and commission sales and discount the apparent brand value residing in the sales relationship? High performing sales staff are notoriously employer-agile and are not subject to long periods of legal constraint on which company they may work for, unlike copyright, patents and trade mark protection that can be traded.

Examples of failed brand extension cases abound in Ries and Trout’s book. They end by humorously suggest that brand extension, or BE, products carry the extra warning: “The Marketing General has decided that line extension is dangerous to your profits” (Ries and Trout 1981 p 139). It has been questioned whether most brand managers adequately evaluate the long term costs and benefits of the extensions (Sharp 1993).

Reasons reported by brand managers for the popularity of this activity include both reduced promotional costs otherwise incurred by brand awareness promotional activities and an increased perception of extension success. The former reason seems logical given that new national brand promotional costs were as high as $U.S. 100 million (Aaker 1990), $US 80 million (Tauber 1988) and regularly $U.S. 50 Million (Aaker and Keller 1990). The latter
reason is still controversial notwithstanding this popularity with brand extensions given the relatively higher rate of brand extension failure when compared to new product failure rates as quantified in the following paragraph.

Documented failures of brand extensions are spectacular and plentiful (Aaker 1991; Chen and Chen 2000; Murphy 1987 pp5-6,95). About 88% of European brand extensions failed within 2 years of launch between 1995 and 1996. Really new and innovative products seemed to enjoy a relatively lower, but still high in absolute terms, failure rate of about 80% over the same period (ACNielsen and Ernst&Young 1999), and even lower rates have been reported for new products in earlier studies, such as 50% to 60% (Krum(Jr.), Shepard, and Morrison 1970), 46% to 53% (ACNielsen 1971) and 43% (Anonymous 1974). Between 70% and 80% of all new products launched are brand extensions (ACNielsen and Ernst&Young 1999; Buday 1989; Rangaswamy et al. 1993b; Stern 1992).

Calculating the weighted failure rate for new products of either type using the ACNielsen Ernst&Young 1999 results and an average of 75% launches being brand extensions gives:

\[0.80 \times 0.025 + 0.88 \times 0.75 = 0.86\text{, or 86%}\]

Note the above combined product failure rate of 86% is much closer to the brand extension failure rate of 88% than the new product failure rate of 80%. Thus it can be seen that brand extension failure is the critical component that dominates the high failure rate of all new products, and as such deserves priority for critical marketing management.

The main causes of the brand extension failures are proposed as:

1. An incongruence, or a lack of fit, in the minds of the consumers, between the extension category, product, or the product’s salient attributes and the original category and brand,
2. a lack of advertising support for the brand extension and
3. poor segmentation, targeting, quality or poorly priced extension (Aaker 1991).
Causes 2 and 3 above are marketing failures of the most basic kind, and will not be discussed further in this thesis, as the remedies are widely accepted and accessible, being often taught in introductory marketing courses. Of much more interest is the first postulated cause which is more complex and may benefit from further research and analysis.

A positive example of extension success referred to earlier is that of Caterpillar boots. A possible explanation for this success could be the perceived ‘fit’ or congruence between the concepts usually associated with Caterpillar heavy earth moving machinery and desirable concepts associated with rugged boots. These concepts might reasonably be assumed to be robustness, ruggedness, the ability to operate in dirty and muddy environments, traction and strength.

A negative example of extension success is McPhoto, a researched extension of McDonalds into fast photo processing. Concepts that could reasonably be assumed to be associated with fast food, such as greasiness, ‘hands on’ tactile mode of interaction, moistness, oral stimulation and pleasant or appetizing smells could be seen as having a poor fit or congruence with the concepts reasonably assumed to be associated with fast photo processing such as a ‘hands off’ or machine processing, non-greasy hands and an unpleasant or unappetizing chemical developer smell. Unsurprisingly initial market research uncovered a strong negative evaluation of this extension concept.

Of even more concern is the emerging result that extension failures frequently and seriously harm the parent brand (Aaker 1991; Chen and Chen 2000; John, Loken, and Joiner 1998; Loken and John 1993; Milberg, Whan, and McCarthy 1997; Romeo 1991b; Swaminathan, Fox, and Reddy 2001) particularly when motivation and involvement surrounding the parent brand is high (Gurhan-Canli and Maheswaran 1998).

To add to the risks of brand extension mismanagement, the evolution of the World Wide Web appears to have led to even more opportunities for mismanagement to occur unless web marketing applications are carefully and systematically reanalyzed using relevant brand management theories (Simon 2001; Yoon 2003).

In summary then the question of why, how, and when, if at all, to extend a brand seems simultaneously as important, complex and controversial today as it was in the 1970s.
To understand the process behind the attempts at empirical generalisations of brand extension theory, it is helpful to understand the relevant theories of how consumers evaluate brands.

**Review of the prominent theories of consumer evaluation of brands**

Two main theory streams have emerged in the last two decades to explain consumer evaluations of brands. One is the associative memory network model (Srull and Wyer 1989), formed by combining the elements of the cognitive processing model (Fishbein and Ajzen 1975 pp 21-80) and the affective response model (Fiske and Clark 1982). Another differentiated model arose about the same time and was initially labeled as the Causal Reasoning model (Cheng and Novik 1992) (Waldmann and Holyoak 1992). These two streams were initially competing models. However, later work resulted in a more comprehensive and combined model, which allows a phase of causal reasoning within the associative memory network model during conditions of incongruence (Shanks, Lopez, and Darby 1996).

The second main stream to emerge was the categorisation model, approached initially from a purely semantic generalisation view (Kerby 1967; Mazanec and Schweiger 1981; Roman 1969), and later developed to focus on memory categories and several types of typicality relationships amongst those categories (Collins and Loftus 1975; Rosch, Simpson, and Miller 1976) and the identification of an ‘ideal point’ (inverted ‘U’) or quadratic effect for measures of typicalities, similarities or congruencies on evaluation, attitude and judgment (Bijmolt, Wedel, Pieters, and DeSarbo 1998; Boush and Loken 1991; Hartman, Price, and Duncan 1990; Mandler 1982 p 22; Maoz and Tybout 2002; Meyers-Levy, Louie, and Curren 1994; Meyers-Levy and Tybout 1989; Ozanne, Brucks, and Grewal 1992; Zinkhan and Martin Jr. 1987).

Several relevant studies helpfully categorise effects on brand evaluation into affective (emotional, and intrinsic) and effective (cognitive and extrinsic) types of effects (Derbaix 1995; Hansen and Hem 2004; Mandler 1982 p 22; Martin and Lawson 1998; Mittal 1990b; Osgood 1964; Romeo and Debevec 1992; Sujan 1985). Detailed evaluation of this viewpoint also relates to Social Judgment Theory (Cooksey 1996; Doherty and Kurz 1996; Hammond 1996).
In addition, several studies attempt to explain brand and brand extension evaluations using the theory of cognitive dissonance (Fry 1967; Gedenk and Neslin 1999; Heimbach 1991; Kim 2003; Yoon 2003), and so required some discussion, following.

**Associative Network Memory Model**

Much of the development of consumer brand evaluation theory and generalising experimental research stems from the related discipline of social psychology, and in particular the prominent theories relating the concepts of individual’s memory of other persons and judgment about those other persons, encompassing both cognitive processing (Fishbein and Ajzen 1975) and affective (Fiske and Clark 1982) elements. A prominent paper that integrates a number of associative memory model hypotheses from the psychology and marketing literature is extensively discussed below (Lynch and Srull 1982). This paper is frequently cited in the marketing literature of consumer evaluations of brands (Aaker and Keller 1990; Anderson 1983; Boush and Loken 1991; Pitta and Katsanis 1995; Sheinin and Schmitt 1994; Sujan 1985; Sujan and Bettman 1989; Sunde 1991), and is titled “Person, Memory and Judgment” (Srull and Wyer 1989).

The Associative Memory Network Model theory is both general and specifically accurate in social science applications. The postulates were initially developed from a set of empirical evidence, but when employed individually or in combinations appears capable of predicting a very large number of social science phenomena far exceeding the scope of the original empirical data.

Thus it becomes a candidate for more thorough attempts at falsification in order to test the theory. Nearly all of the derived predictions require a combination of postulates, implying an interlinked network of postulates from which a very large number of hypotheses can be formulated and subsequently tested.

The field of applications covers memory and judgment situations from no to full prior information about an entity to conflicting and reinforcing information and appears to successfully predict and explains evaluation difficulty, speed and sign in many situations.
The original object reference is ‘other persons’ but a more helpful term, consistent with subsequent marketing interpretations cited above, is ‘entity’ which encompasses brands, firms, staff and organizations. Similarly the subject of the theory is ‘a person’ but for marketing applications it is consistently reinterpreted as ‘a consumer’ in the relevant literature. Finally the marketing literature refers to ‘performance’, rather than the original term of ‘behaviour’. Throughout this following section this translation has been made.

The theory has been found to have the following desirable and applicable characteristics:

1. It successfully predicts the amount of information a consumer recalls about an entity, the types of information that are recalled under different conditions, the order of recall of different types of information and the time taken to effect the recall.

2. It prescribes the roles of both evaluative and descriptive factors in both the recall and application of entity information and models the effect of expected entity performance on entity trait information recall.

3. It explains the activity prioritization process associated with high information processing rate situations.

4. It models the effect of information about an entity on entity trait and preference judgments by predicting types of information recalled and also judgments made.

5. It aids conceptualization of (a) the similarities or differences in internal images formed between sets of entities and the individual entities in the set, (b) the effect of received but then discredited information on both information recall and entity judgment; and finally (c) the effect of postprocessing after information has been received.

Srull and Wyer’s model consist of 15 postulates that explain (a) the initial encoding and organization of entity trait and performance during impression formation, (b) the memory storage of formed impressions and (c) the processes that control the subsequent retrieval and use of these encoded representations.

These 15 postulates of Srull and Wyer’s model are organized into five groups, each associated with one of the three categories (a) through (c) above:
Group 1: Attribute encoding postulates.

1. Consumers who acquire information about an entity’s performances to form an impression of that entity will spontaneously interpret these performances in terms of the attribute concepts they exemplify.
   a. This is done by comparing the performance descriptions with features of the attribute concepts that are stored in memory
   b. When more than one attribute concept is relevant to interpreting a performance, the concept applied is usually that which is most easily accessed from memory

2. When attribute concepts of an entity already exist at the time when information about performances of an entity are acquired, only the performances that typify these concepts are encoded in attribute terms. Alternatively, performances that have implications for other attributes are not encoded in attribute terms.

Group 2: Evaluative encoding postulates.

3. Consumers who are asked to form a general impression of an entity will try to construct a general concept of that entity as preferable or not preferable. This concept, though often based on the attribute concepts used to interpret the entity’s performance, is primarily evaluative.

4. The evaluative concept of an entity is formed on the basis of a subset of the information available about that entity, often the first subset that permits a coherent evaluative concept formation.

5. An entity’s performances are interpreted in terms of an evaluative concept once formed.

Group 3: Reactions-to-inconsistency postulates.

6. If a clear evaluative concept of an entity cannot be formed from initial information, either because the initial information elements differ in preference implications or for other reasons, consumers who are uncertain about an entity’s
attributes will review the performances they have encoded in terms of each attribute to ensure they have interpreted these performances correctly. This cognitive activity leads to the formation of associations amongst the performances that have been encoded in terms of each concept.

7. Once an evaluative entity concept has been formed, performances of that entity that are evaluatively inconsistent with this concept are thought about in relation to other performances that have evaluative implications in an attempt to reconcile them. This leads to the formation of associations amongst these performances. Performances that are either evaluatively neutral or consistent with the entity concept do not stimulate this cognitive activity.

8. Consumers who are uncertain of a concept they have formed of an entity will review the entity’s performances that are consistent with this concept to try to confirm validity. This strengthens the association between the performance and the concept.

Group 4: Storage Postulate.

9. Each attribute-performance cluster that is formed as a result of encoding performances in terms of attribute concepts, in addition to the representation finally based on the evaluation, function as separate units of information. All representations are stored in memory independently at locations linked to the relevant entities.

Group 5: Retrieval Postulates.

10. When consumers are tasked to recall an entity’s performance, they will retrieve and report the contents of one representation of the entity at a time. When more than one representation is contained at the memory location pertaining to that entity, the one that is most recently stored has the highest probability of being retrieved.

11. The recall of specific performances contained in a representation is the result of a sequential search process that begins at the central concept node and progresses along various pathways in the network. Performances are reported in the order they are activated in the course of this search.
a. When more than one path connects to a node, the strongest path between nodes is chosen as the next path in the search.

b. When the only path connected to a node is the path taken to reach that node, the search is restarted at the central node.

12. Searching along a path between nodes increases the strength of that path, increasing the posterior likelihood that recalling one of the concepts (performances) will cue the retrieval of the other.

13. When a search activates a predetermined number, say N, successive performances, all of which have been previously retrieved, the search is terminated. N may vary both between individuals and within individuals between judgment risk categories.

14. A consumer who is tasked to judge an attribute performance level of an entity will search for a representation of that entity which has a central concept that specifically pertains to that attribute and the entity’s performance. If such a representation is found, judgments are based on the implications of the concept defining this representation without a review of the stored individual performances.

15. If a representation with a central concept that has direct implications for this judgment cannot be found, the consumer will retrieve and use the general, evaluation-based entity representation as a basis for the judgment. This judgment will be based on both

   a. The evaluative implications of the central entity concept defining the representation and

   b. Any performances contained in it that have direct implications for the judgment.

Please note: the above model description is essentially as described by Srull and Wyer in the context of person memory but has been adapted here to generalize to the context of any entity, including other persons and specifically brands and brand extensions (Srull and Wyer 1989).
Causal Reasoning Model

Another popular and initially competing model of consumer knowledge drawing from psychology literature is frequently referred to as the Causal Reasoning Model (Cheng and Novik 1992; Waldmann and Holyoak 1992). This model ignores affective and other non-cognitive processes and rather proposes that consumers consistently apply scientific reasoning, i.e. seek to form evaluatively consistent judgments at all stages in the memory and reasoning process, not just in the specific situations of evaluative inconsistency or primary evaluative formation described in the Associative Network Memory Model. A corollary of the causal reasoning model of memory is that subsequent brand attribute performance information that is confounded with other previously processed attribute information that initially helped to form an evaluatively consistent brand preference judgment will be just as actively cognitively processed but produce an agnostic attribute evaluation. This is in direct conflict with the associative network memory model. However Waldmann and Holyoak’s initial inferences are particularly speculative, apparently confusing a failure of a type I statistical test with a successful type II test.

The origins of the Causal Reasoning Model theory as a logic of choice rather than a psychology of value were criticized soundly in the marketing literature, establishing the boundary conditions as decision application transparency, not often found in practical choice situations (Tversky and Kahneman 1986). Whole volumes of papers discussing and generalizing on the cognitive biases with respect to causal and deductive scientific reasoning in human decision making have been published with wide acclaim and acceptance (Kahneman, Slovic, and Tversky 1982 pp 101-53,211-68).

More recent works, including some by the original authors, refine these causal models so as to more successfully explain the conditionality of attribute performance concept and overall preference review producing a model more inclusive of the associate network model (Cheng 1997), especially by explaining the conditions leading to evaluative incongruence. In particular, these works successfully describe the conditional cognitive activity surrounding the revised brand schema evaluation as a broadly scientific, causal reasoning process (Busemeyer, McDaniel, and Byun 1996 pp 357-91; Shanks et al. 1996; Waldmann 1996).
Categorisation Theories

An informative yet often differentiated literature stream to the above memory models includes studies surrounding the natural formation of memory categories comprising groupings of discrete concepts, hierarchical category models and their higher order memory concept representations, known as prototypes, as opposed to a lower-level representation used as references in judgments known as exemplars (Posner and Keele 1968). An early scholarly work has integrated the Associated Memory Network Model with hierarchical category memory models of prototypes and exemplars in a semantic processing context to form a Spreading Activation theory of Semantic Processing (Collins and Loftus 1975).

Next, further research moved categorization theory development from the semantic and therefore cognitive to a more holistic perspective that includes affective influences (Cohen 1982; Sujan 1985). Difficulties in designing experiments to successfully separate exemplar- and prototype-based evaluation strategies were noted and somewhat resolved, (Malt 1989), resulting in a realization of a continuous mixture distribution of both exemplar and prototype strategies both within and between subjects in typical judgments.

Thus we see important differences in evaluation tasks arising from a subject employing either an exemplar (a particular instance, e.g. a known branded product, such as a BIC disposable razor) or a prototype (a higher-order belief e.g. about a well-known broad brand, such as General Electric) as the reference concept when performing evaluation tasks. Studies show that when exemplars are employed as referents, extrinsic attributes are more likely to be accessed in an effective, or cognitive evaluation, specific to the product category, e.g. the BIC razor has two blades and has a low price (Barsalou and Hutchinson 1987). Conversely, if a higher-order prototype concept is employed as a referent, it is more likely that an affective evaluation of intrinsic attributes, common to the broad brand and not the category are accessed, e.g. the brand is known for high quality products. This is more likely to be associated with an affective or emotional response to the evaluation task (Boush, Shipp, Loken, Gencturk, Crockett, Kennedy, Minshall, Misurell, and Strobel 1987).

This literature stream then yielded a rich list of additional factors that influenced judgment and evaluation relevant to both the understanding of observed judgment or evaluation...
phenomena and the design of successive, progressive experiments. These studies included those that addressed variations in judgment processes due to memory concept accessibility (Higgins 1989 pp 72 - 81; Ratneshwar, Barsalou, Pechmann, and Moore 2001) and situational context effects (Thompson 1989; Turner, Oakes, Haslam, and McGarty 1994).

**Ideal point evaluation models**

Models have been proposed for inverted “U” or ideal point relationships between congruence and affective evaluation, based on a combination of moderate associated memory network evaluation intensity and a moderately positive evaluation valence. This is sometimes cited as Mandler’s Theory (Mandler 1982 p 22) and has received empirical support in branding experiments (Meyers-Levy and Tybout 1989).

The posited qualitative model for this ideal point relationship is based on the following model: Overall stimulus evaluation is a product of the evaluative valence and the evaluative intensity. Consider an existing, preferred memory concept and number of new concept stimuli ranging along a continuum of congruence from highly congruent to highly incongruent. At the highly congruent end of the scale, although the valence of the evaluation of the new stimulus will in general be also favorable, the intensity of the evaluation will be low or null as little or no affective arousal (nor indeed any effective or cognitive activity) has been stimulated. However, as the congruency of the new stimulus reduces from high to moderate, the increase in cognitive processing results in a moderately positive valence of the new concept evaluation in conjunction with an increased and moderate level of intensity associated with the response. Finally, as the stimulus becomes highly incongruent, the intensity of the evaluative response increases further but the valence of the response changes from positive to negative. Thus we can deduce that if this theory does hold, the maximum evaluation occurs around the moderate range of congruencies, leading to an inverted “U” or ideal point congruence evaluation behavior.

Similarly, ideal point models of effective or cognitive brand evaluation based on depth of attribute search have also been proposed and empirically supported in a brand evaluation context (Maoz and Tybout 2002; Ozanne et al. 1992). Hence these results lead us to
hypothesise ideal point evaluation behavior resulting from either or both affective and/or cognitive or effective evaluation processes.

**Cognitive Dissonance Theory**

Another important and somewhat related model of consumer cognition leading to brand evaluation, drawn from psychology literature and frequently cited, is the theory of cognitive dissonance. Expressed simply, perhaps too simply and too modestly, the original article stated that “inconsistency motivates people to alter their cognitive system in such a way that it will become consistent” (Festinger 1957 pp 8-31). L. Festinger’s original theory consists of dissonance, D, and consonance, C, with intensities held internal to the subject but measured externally on agreement or disagreement attitude scales. Total dissonance was defined as D/(D+C). In Festinger’s original theory, D and C exist and are relevant only to a single, existing cognition G, such that C implies G and D implies not(G). This cognition G was operationalised as an effective behavior, usually a public declaration, in the original and subsequent experiments that supported the theory. The distance between a subject’s pre-existing, well-formed, privately held, internal belief X about a specific concept and the operationalised behavior G defined the intensity of D and C for that subject and cognition.

A major clarification and improvement on the original theory of cognitive dissonance is the theory of radical dissonance, or, more exactly, the radical conception of the theory of cognitive dissonance (Joule 1986). This radical conception removes some logico-deductive inconsistencies left unaddressed in the original theory, such as the introduction of new consistencies lowering the total dissonance. Furthermore, it clarifies the cognitive dissonance theory to emphasise that the salient behavior is the reference datum for consistencies and thereby eliminates any additional dissonance, D’ between the consistency C and dissonance D to a cognition G, from the G-specific measure of total dissonance (Joule 1986).

Regrettably much research has been published that appears to suffer from misinterpretation of the original theory or its misapplication to other related areas of inquiry. A recent example of this is the reinterpretation of the original theory as: “humans are universally motivated to validate their opinions and abilities relative to those held by others” (Pincus 2004 p 379). Whilst this indeed may be true, the carefully designed experiments of the original and subsequent contributory authors (Aronson 1992; Festinger 1957 pp 48-52,98-103,38-55; Joule...
1986; Lempert 2004) were neither designed, nor able, to falsify that particular position, and as such, Pinchus’ position is a mere extrapolation of the original or related theories.

The additional role of cognitive dissonance theory in predicting future behavior is not supported by the core theorists in this area. Neither is the role of novel, hypothetical or otherwise poorly formed cognitions falsifiable in the experiments as they were reported. Therefore this application of the theory is not supported. The core theorists in this area conducted carefully constructed experiments focused strictly on past or present behaviors and present conceptions.

This current boundary of application excludes future predicted behaviors and attitudes to poorly formed or newly formed beliefs, including the predicted behavior associated with newly created brand extension concept stimuli, but cover, or can be logically and rationally extended to cover, some research areas relevant to brand evaluation. These areas include predicted behaviors and associated attitudes surrounding existing brand concepts such as the subsequent effects of disconfirmation messages on either existing brand concepts or previously formed brand extension conceptions.

Interestingly, several researchers have employed this theory or its derivatives to explain the effect of negative information on brand evaluation (Yoon 2003), co-brand evaluation (Heimbach 1991) and communication effects on brand extension evaluation responses (Kim 2003) with mixed levels of scholarship and rigor. The research on co-brand evaluation alone mis-applies the cognitive dissonance theory to predictions of future behavior.

Cognitive dissonance theory is, however, directly applicable to the study of negative or ‘backlash’ effects of individually successful brand extensions when they dilute the desired concept associations of the original parent brands. Relating ‘backlash’ elements to cognitive dissonance theory, the relevant past behaviour in ‘backlash’ studies includes the choice of the relevant brand extension and the present conception includes the (newly reformed, diluted) original brand conception (Aaker 1990).
Social Judgment Theory

Social Judgment Theory, or SJT, is a system of experimental research in psychology rather than a single theory of judgment or cognition, but it addresses many concepts and issues relevant to the investigation of brand extension. SJT was heavily influenced by the early work in experimental psychology of Egon Brunswik (see Doherty and Kurz 1996), and addresses organisms or subjects, their environmental cues, variables, or dimensions, and their perceptions or achievements.

One important element in Brunswik’s approach to the design of experiments was that there should always be a balance of emphasis between the subject and its environment, rather than isolating the subject from its environment as is common in other research traditions. This led in turn to a special and early appreciation of: (a) representative, or ‘covering’ experimental designs, with lower efficiency criteria, that were more representative of a subject’s typical environment, or ecology, rather than the more usual systematically balanced or contrasting designs, with higher efficiency criteria; and also (b) the importance of the role of, and need for quantitative analysis of, environmental covariates and their covariances, in perception and judgment (Doherty and Kurz 1996).

Another important element of Egon Brunswik’s approach was ‘probabilistic functionalism’. This was an early recognition of the stochastic, i.e. probabilistic, and somewhat uncertain, nature of a subject’s environment, and the deduction that adaptations to that environment would therefore likely be based on probabilistic, rather than deterministic, strategies. This in turn implies subjects’ judgments will be based on multiple objects or criterion in the distal environment and associated multiple perceptual cues proximal to the subject, with a characteristic of minimizing the risks of imperfect or ‘noisy’ cue perception, and resulting in a natural perceptual processes that minimize mean errors in judgment at the expense of higher frequencies of (albeit often small) imprecision in that judgment.

Brunswik often used an optical lens analogy to demonstrate this generally correlative relationship amongst environmental criteria and subjective perceptual cues (Cooksey 1996).
The analogy is an apt one and works at many levels. There is a correlation amongst the occurrence and frequency of distal environmental cues in the SJT Lens model as there is amongst light rays emanating from distal objects in the far field of an optical lens. Just as the light rays all converge through the lens itself, becoming co-incident at that point in space, so do the cues converge through a perceptual organ, say, the human eye or ear, becoming co-incident at that point in time and space. Finally, just as the light rays diverge on the proximal side of a lens, so do multiple perceptions form internal to the organism. In both cases, the divergent elements (light rays and multiple perceptions) are again in general correlated.

A more comprehensive theory of human cognition, the cognitive continuum theory, was developed by K. R. Hammond from the conceptual framework of Egon Brunswik’s work in the 1980’s (Hammond 1996). The theory can be presented as four premises:

1. Various modes of cognition can be placed on a continuum anchored at one pole by intuitive cognition and by analytical cognition at the other.
2. The most common sort of cognition includes elements of both intuitive and analytical cognition, and we call this mode *quasirationality*. It is related to and yet different from Simon’s “bounded rationality”.
3. Cognitive activities move along the continuum, alternating back and forth between one pole and the other. Success on the task at hand inhibits movement, failure stimulates movement.
4. Tasks can be ordered on the continuum according to whether they induce analysis, quasirationality, or intuition.
   (Hammond 1996).

This cognitive continuum theory unifies, extends and further explains some of the deductions from cognitive dissonance theory, associative network memory model, the ideal point evaluative response models and the causal reasoning model outlined above. Specifically it links tasks and their evaluative congruency to the various mixtures of rational analysis and intuition in a subject at the time of performing an evaluative task or judgment, and offers an explanation for the mechanism producing the mix.

Social Judgment Theory thus leads us to expect correlations amongst multiple perceived cognitive and affective attributes in any stimulus-response situation, even when attempts are made by social scientists to eliminate inter-factor correlations by employing experimental design. This in turn will increase the likelihood of encountering significant heteroskedasticity when applying quantitative inference generating frameworks, and researchers need to adapt their approach to deal with this problem.
5. Brand Extension theory

Brand image and knowledge

Building on this theoretical framework, important processes surrounding brand image and knowledge have been identified in the marketing literature, often referred to as schema models of brand association (Anderson 1983; Sujan and Bettman 1989).

A prime brand evaluation corollary of Srull and Wyer’s postulates is that of memory accessibility. Frequently associated brand schemas with consistent performances on relevant attributes and positive preferences are those concepts that are most likely to be retrieved first in a new choice situation. This is clearly exemplified by our common understanding of good brand equity: a memory shortcut for concepts associated with consistent high quality, where quality is defined as ‘fit for the purpose’, and value, causing a higher probability of consumer choice of the highest evaluated brand over the lower valued alternatives.

Note, however, that the prime corollary strongly relies on relevance of attributes to the existing evaluative concept and consistent positive performance evaluations to cause a strengthened association between the brand and positive evaluations. The natural question to ask then is what are the brand implications of inconsistent entity performance evaluations, both in attribute relevance and preference?

The 15 postulates successfully predict that either (a) a poor performance of an evaluatively good, preferred brand, or (b) a good performance on an evaluatively poor, preferred brand on an irrelevant attribute, or (c) brand performances on an attribute of any level of relevance consistent with the existing brand preference evaluations will not lead to significant cognitive activity, neither will it affect brand evaluation nor preference and only improve the likelihood of the main brand concept retrieval.
However where attributes are relevant to a brand’s concept evaluation and a newly observed brand performance is inconsistent with the existing stored evaluative concept, then the postulates predict that significant cognitive activity surrounding the search, retrieval, association and reconciliation of all stored past brand attribute performances occurs. If there is enough information retrieved to form a new consistent, but initially weakly retrievable, evaluation of the brand this new concept and its associated evaluation will replace the previous stored concept and evaluation. If there is not enough information available for a new consistent brand concept evaluation, the old evaluation will remain but additional new brand performances will be carefully monitored and processed until a new consistent evaluation can be formed.

This is consistent with the tightrope that successful brands and their managers walk every day with respect to maintaining their good brand image. Regardless of how good the past performance has been, or how relatively infrequently failures occur, it only takes one new, evaluatively inconsistent, failure of a brand to deliver the expected quality or value to comprise a significant attitude-changing event for a formerly satisfied and loyal consumer. Once reformed, this attitude requires considerable resources, time and a similar attitude-changing event comprising inconsistent attribute performances with any new positively evaluated brand’s performance before the original positive brand evaluation may be reformed yet again. Indeed this offers an explanation of the mechanism behind the initially confusing situation of brand usage being both an antecedent and a consequence of satisfaction (Bolton and Lemon 1999).

This is easily explained by a consumer’s preferred branded products being experienced more often by the consumer. At each experience event of consumption or use, the relatively fixed, embedded probability of experiencing a failure of the product to meet expectations exists, and this probability is relatively independent between successive usage or consumption events, so the probability of a consumer experiencing an unsatisfactory event is approximately the product of the embedded failure rate times the frequency of usage. Thus it can be seen that consumers are more likely to encounter an unsatisfactory usage event associated with their frequently used and preferred branded products.

A third corollary of Srull and Wyer’s 15 associative memory postulates for branding is that once an overall brand preference evaluation has been formed from a consistent set of attribute
performance information, other attribute information consistent with the preference and performance judgment will be irrelevant and will not be processed. Note this is not what would be described as causal reasoning or a logical scientific method. This is referred to as the 'blocking’ effect in marketing literature and has been robustly tested and empirically generalized (Osselaer and Alba 2000).

The literature stream of categorisation theories permits several additional insights into consumer brand evaluation, starting with explorations of ‘family branding’, i.e. brand category, phenomena (Fry 1967), including an early and appropriate use of Multidimensional Scaling (MDS) preference mapping technique to the issue (Mazanec and Schweiger 1981). This work provides more solidity and credibility from a theoretical basis for the application of perceptual mapping to consumer brand perceptions that is not often cited. Early success in applying these theories to various brand categories (Ward and Loken 1986) were mixed, with stronger support found with particular categories. Further additional confounding context effects on judgment were also found (Schmitt and Dubé 1992).

There are many other consumer psychology theories relevant to brand and product information evaluation, such as the effects of elaboration (Costley and Brucks 1992), visual vividness (McGill and Anand 1989), the level of involvement (Hawkins and Hoch 1992), gender (Meyers-Levy 1988), mood and framing effects (Martin and Lawson 1998), image beliefs (Mittal 1990a) and many others (Tybout and Artz 1994). In the context of the discussion on brand and brand extension evaluation these can be thought of as (a) independent of the salient process under discussion which is the transfer of brand equity to brand extensions, yet (b) frequently supportive of the general effects of congruency on cognition and elaboration. In other words, these other brand evaluation effects may be equally applicable to both original brands and their brand extensions but there is no theoretical or empirical support for these effects playing a special or differentiated role in the transfer of equity between brands and extensions. Rather they tend to add to the variance of any regression model employed to generate quantitative inferences about effects, generally increasing the level of challenge in experimental design and analysis involved in conducting quantitative analysis in this topic area.
Fit

So how does this rich literature of brand evaluation theory add to our understanding of brand extensions? One of the main theoretical elements in a brand evaluation processes is the concept of congruence. Simply stated, when an attempt at an evaluation of a brand extension, i.e. a new category for an existing brand, then the degree of congruence surrounding both the overall brand preference evaluation and the sufficient set of attribute evaluations to form the aforesaid overall brand evaluation and concept will determine the amount and extent of cognitive processing, attribute performance retrieval and subsequent causal deductive reasoning that will occur.

Consider the scenario where a new category is proposed to a consumer for a brand with an existing positive preference evaluation in an existing category.

The relevant theory predicts that if the concepts at all of the category, brand and attribute levels are congruent, then no additional significant cognitive processing will occur and the overall evaluation of the old brand in the new category, i.e. the evaluation of the brand extension, will be one of a positive preference. There has been a recent generalisation of this result in the broader context of stimulus and conditioning in the related psychology literature (Till and Priluck 2000) with noted applications to both brand extension and private and national label brand management. However it is suggested that although the valence of highly congruent extension concepts are positive, the intensity of the afferent response may in general be low, or even null, due to a generally low level of arousal and cognition (Mandler 1982 p 22).

Conversely, if the overall category and brand concepts, performance judgments or attribute performance evaluations are not congruent between existing and new categories then the theory predicts that: (a) a significant amount of cognitive processing will occur, (b) all relevant stored attribute and brand concept performance evaluations will be retrieved and re-evaluated and, if necessary, (c) new information will be cognitively processed until a consistent evaluation of the brand extension performance concept and preference can be formed.
An important overlay to our understanding of the basic process is offered by categorization theory. In the light of this literature stream, it becomes apparent that we have to ask: “what is the hierarchy level of both the parent, reference memory concept, and the extension, evaluation concept? Are they higher-order, prototype or lower-order, exemplar concepts?”

This in turn influences the nature of the evaluation task, which is more likely to access either effective, extrinsic attributes if exemplar referents are employed or affective, intrinsic attributes of prototype referents are used. Studies have shown that where the extension category is a close enough fit to an existing exemplar in memory, this will more likely be employed as the parent brand referent when evaluating the extension than a higher order, broad brand prototype (Boush and Loken 1991). For narrow parent brands, higher-order prototype concepts may not exist. Extreme examples include the category-dominant brands, where the dominant narrow brand replaces the generic noun and verb in everyday speech, e.g. in some markets, consumers still ‘Hoover’ their floors rather than vacuum them, even though the same firm also sells “Hoover” branded washing machines into the same market. Recent papers explore the effect of enhancing elaboration on extension evaluation via various manipulation techniques such as the interactions between brand personalities, person beliefs and the level of elaboration (Ahluwalia 2008; Monga and John 2010; Yorkston, Nunes, and Matta 2010).

Also, we can reasonably hypothesise an ideal point, or inverted “U” response to parent brand – extension similarity or congruence (Maoz and Tybout 2002), based both on Mandler’s Theory for affective or intrinsic congruence (Mandler 1982 p 22; Meyers-Levy et al. 1994; Meyers-Levy and Tybout 1989) and similar arguments, empirically supported, for effective or cognitive fit (Ozanne et al. 1992) as discussed above. A recent paper explores this relationship from the reverse perspective by investigating how much atypicality an extension can possess and still enjoy a good level of evaluation (Batra, Lenk, and Wedel 2010).

The key to the underlying processes invoked is clearly then the congruence between the original brand associations and the new category concepts. This is indeed the focus of the relevant marketing literature, referred to in this literature area as the dimension of ‘fit’ (Tauber 1988). More helpfully and specifically, the concept of a dimension of congruence or fit has been operationalised as 4 separate items: complementarity, the extent to which the new category compliments the existing one, substitutability, the extent to which the new category
could be viewed as a substitute for the original category, transferability, the extent to which
the firm’s existing competencies in product manufacture and service delivery can be seen as
able to be transferred to that of the new category, and the non-fit attribute difficulty, the
perceived difficulty that the firm would face in producing the extension (Aaker 1990; Keller
and Aaker 1992; Sunde and Brodie 1993). This particular attribute is perhaps more helpfully
conceived as the opposite of triviality, in the sense that subjects might evaluate trivial
extensions poorly.

Several taxonomies have been simultaneously argued for these measures:

Firstly, this author has identified striking parallels between this taxonomy and a much earlier
and general one proposed by the Scottish philosopher David Hume, who proposes: “Three
connecting principles of all ideas are the relations of resemblance, contiguity and causation”
(Hume 2007 p 20). Surely readers will admit close conceptual links between contiguity and
complementarity, resemblance and substitutability, and causation and transferability. If two
concepts are perceived to be contiguous, then are they not likely to be also complimentary,
and vice-versa? Similarly, are not concepts that exhibit resemblance also likely to be
substitutes, and vice-versa? Finally, what else is transferability but the relationship of
causation from parent brand manufacturer’s skills and abilities to the successful production of
the extension? Notably relationships connected to trivial causation do not figure in Hume’s
works.

Additionally, complementarity and substitutability can be seen as both generalisable extrinsic
attributes, after the consumer psychology literature approach, and as generalisable demand-
side attributes from a micro-economic perspective. Similarly, transferability and difficulty can
be viewed as either generalisable intrinsic or supply-side micro-economic attributes (Albion
2006; Smith 1990; Smith and Park 1992). Thus we can link complementarity and
substitutability with the likelihood of an effective, cognitive, lower-order category or
exemplar referent concept accessibility process. Similarly we can link transferability and
difficulty with affective, higher-order category prototype referent concept accessibility.
Whilst there very likely are more detailed intrinsic and extrinsic attribute structures relevant
to particular categories and brands, these measures have a generic character that encompasses
most of those specifics in a measureable way. This in turn allows a reduction of respondent
fatigue and enabling practical experimental design of studies in this area, although this point appears to have eluded some authors (Farquhar 1990).

Thus ideal point or inverted “U” fit relationships might reasonably be expected for all four fit dimension effects as they were theoretically argued and empirically supported in both affective and cognitive judgment contexts. However this point appears to have been overlooked by the original study and all the close replication experiments (Aaker and Keller 1990; Alexandre-Bourhis 1994; Barrat et al. 1999; Bottomley and Doyle 1996; Fu and Saunders 2002; Holden and Barwise 1995b; Nijssen and Hartman 1994; Patro and Jaiswal 2003; Sunde 1991).

It should be noted however that as more interaction and second-order effects are posited in any research problem, attempts to design effective and practical experiments to address the research problems become hampered by increasing numbers of both (a) contrasting stimuli simply as a result of more effects to estimate, even when fractional factorial designs are employed, and (b) sample size requirements. These two problems also interact destructively as studies with more embedded respondent fatigue will suffer from decreased response rates, making the overall problem one of geometrically increasing difficulty with respect to research resources including funding, time, experimental design skills and sample population access. Thus the original authors cited above should not be seen as negligent, rather simply hampered by these practical constraints.

The extent to which extrinsic versus intrinsic judgment processes are employed will depend first on the breadth of the brand and then the existence and accessibility of a higher order prototype concept in the consumer’s associate memory network schema. This has already been described as a mixture distribution in the consumer population, and so can be expected to provide an additional source of variation in any measurement. This, in turn, requires the application of an inference generating framework that is robust to this type of additional error source, as discussed below. By simple inspection of the list of original brands employed in the relevant studies detailed below, we can see that most recognized brands are not only well known (indeed, as stated in most of the relevant articles) but (perhaps not by coincidence) also relatively broad brands.
From a conceptual viewpoint, we can logically deduce that brand extension concepts exhibiting high levels of complementarity between parent and extension brand category concepts will likely show commensurately low levels of substitutability and vice-versa. But other predictions about other relationships amongst transferability, complementarity, substitutability and Difficulty are not so easily made prior to any descriptive analysis. Other authors report similar findings based on equivalent interpretations of extension congruency conditions (Park, Milberg, and Lawson 1991).

This chapter has reviewed and discussed five main literature streams related to this thesis topic which have set the stage for identifying several research questions addressed in the following chapter.
Chapter 3 Research questions and propositions

Review of the literature as outlined in chapter 2 reveals several gaps that can be profitably researched as follows.

Literature gaps identified in this research

Many of the 61 peer-reviewed papers attempting to generalise on brand extension theory using empirical data between 1961 and 2004 referred to in chapter 2 have been centered on the consumer. They embrace models that include measures of perceived brand attribute intensity. Commonly included attributes relating directly to the relevant theoretical literature are quality of the original brand, multiple attributes measuring the fit between original and extension categories and the difficulty of the firm to make the extension. Aaker and Keller (1990) produced a study that is frequently cited in subsequent publications related both to the theory of brand extensions and its generalisability. One of the gaps in the literature is that little agreement exists amongst these studies, leaving little marketing knowledge for the marketing academic or practitioner. Therefore there is a literature gap with respect to successful empirical generalisations.

Another related gap in the literature is that only two research groups have applied general linear models to pooled sets of some of the close replication datasets, but no study has pooled all the qualifying data sets from all the close replications.

Finally, although many studies compare two or more competing theories of brand extension evaluation, and some are fairly comprehensive (Alba and Hutchinson 1987; Hamer 1994), none thoroughly compare and contrast all the relevant literature cited here. A more comprehensive review would produce a deeper understanding of the extension evaluation process and a more informed critical evaluation of both the literature and research findings presented here. Three areas need deserve thorough scholarly investigation: (a) the apparent lack of convergence between the Categorization Theory stream, especially Mandler’s Theory
of ideal point fit behaviour and that of Aaker and Keller’s stream of research, (b) the derivation of Aaker and Keller’s fit dimensions and the relationship of complementarity, substitutability and transferability measures, and (c) their relationship to the measures of similarity, typicality and congruency.

**Question 1: What is the basis for Brand Extension Theory?**

Is there a basis set of elements in a theory of brand extension? Does original brand quality play a direct role in positive brand extension evaluation, or does it only impact via an indirect mediation by the fit attributes? Can an extension be too trivial, and therefore suffer poor evaluation?

This gives rise to the following propositions:

**P1:** Original brand quality plays a direct role in positive brand extension evaluation.

**P2:** Original brand quality impacts brand extension evaluation by an indirect mediation via the fit attributes.

Note the above two propositions are not mutually exclusive.

**P3:** Extensions seen as trivial suffer poor evaluation.

Alternatively:

**P4:** Extension triviality has no impact on extension evaluation.

**Question 2: Is Aaker and Keller’s brand extension theory generalisable?**

Aaker and Keller’s theory on Brand Extension posits that positive attitude towards a brand extension was influenced by the four main-effect brand extension ‘fit’ attributes, namely
complementarity, substitutability, transferability and difficulty and the \textit{indirect} effect of original brand quality moderated by the ‘fit’ attributes. Has this theory, studied empirically in their 1990 paper, been reproducible over many subsequent studies, given an appropriate model framework?

Hence:

\textbf{P5: An appropriate modeling framework will substantiate an empirical generalization of Aaker and Keller’s theory of brand extensions based on all of the data associated with the existing set of close replication studies.}

Alternatively:

\textbf{P6: An appropriate modeling framework will fail to substantiate an empirical generalization of Aaker and Keller’s theory of brand extensions based on all of the data associated with the existing set of close replication studies.}

\textbf{Question 3: What were the problems with previous attempts at Empirical Generalisations in the Brand Extension literature?}

Why has there been so much variation in findings amongst the published studies in this area? How could these studies have been improved? Are there any research steps or processes that have been overlooked, misapplied or have not been allocated sufficient importance?

Many of the close replication studies employed residual centering as a method of dealing with ‘nuisance’ data co-linearity. Given the extant co-linearity and the hypothesized indirect (interaction) effects, study sample size may not have been sufficient to support robust inference generation.

\textbf{P7: Residual centering in models with interaction effects leads to bias in the generated inferences.}

Alternatively:
P8: Residual centering in models with interaction effects does not lead to bias in the generated inferences.

P9: The close replication studies had sufficient sample size to support robust inferences

Alternatively:

P10: The close replication studies did not have sufficient sample size to support robust inferences.

Question 4: What is the relationship amongst the multiple Brand Extension Theories?

Are there deficiencies, contradictions, synergies or similarities amongst the competing Categorisation and Associative Network Memory Model theories? Are these different theories complimentary or competing? Is the Associative Network Memory Model theory so deficient compared to others so as to be unusable? How do the theories surrounding hierarchical category memory constructs influence our view of brand extension evaluation? Specifically: Is there any evidence of an ‘ideal point’ fit relationship, as suggested by Mandler’s Theory, and associated with hierarchical category memory constructs, broader parent and extension brand categories and prototype and exemplar fit models?

The questions above regarding ‘ideal point’ fit relationships give rise to the following three pairs of propositions:

P11: There is an ‘ideal point’ relationship between complementarity and extension evaluation.

Alternatively:

P12: There is no ‘ideal point’ relationship between complementarity and extension evaluation.
P13: There is an ‘ideal point’ relationship between substitutability and extension evaluation

Alternatively:

P14: There is no ‘ideal point’ relationship between substitutability and extension evaluation

P15: There is an ‘ideal point’ relationship between transferability and extension evaluation

Alternatively:

P16: There is no ‘ideal point’ relationship between transferability and extension evaluation

This chapter has identified four main research questions and 16 related research propositions that are investigated in the following two chapters.
Chapter 4 Methodology

The literature review in chapter 2 identified several issues relevant to methodology.

The inference space and correlation issues raised by Meehl’s (1990) criticism of a lack of progress in empirical generalizations in marketing are fully addressed in the inference space paragraph in section 1 of this chapter by employing a mixed or hierarchical model inference generating framework.

Similarly, the issue raised by Meehl’s criticism concerning the lack of measurement models is addressed in section 4 of this chapter.

Integrating the related research streams of ideal point evaluation models and brand extension theory is addressed by the ideal point paragraph in section 1 of this chapter.

Increase in certainty and power of statistical tests

By allowing the studies to be pooled into the one hierarchical model and appropriately analyzed as one hierarchical data set, the ratio of degrees of freedom of the data to that of the model greatly increases, thereby decreasing both the type I error and type II error, i.e. increasing the certainty and power of the significance tests (Baird 1981; Fisher 1956 pp 128-38). Once the best variance-covariance specification is found, the accuracy of the non-iterative type I tests for the general linear mixed model can be improved by using the Kenward-Rodger method for more accurately estimating the denominator degrees of freedom, or DDF, for the F-tests for individual fixed effects certainty (Schaalje, McBride, and Fellingham 2001). Since the best general linear mixed models are a much better fit to the data than ordinary least squares regression models, and the necessary assumptions of OLS regression are not met by this data as discussed previously, the existing closed form estimates of test power for OLS regression models do not apply to this data and it is un-instructive to calculate them. However it is instructive to note that there are no closed-form solutions for either exactly calculating or estimating the power of a statistical parameter test in a general
linear mixed model a-priori, and that simulation is recommended both before and after model estimation for test power estimation (Castelloe and O’Brien 2001). Test power in this study is estimated retrospectively as all experiments were designed and conducted before this thesis study was performed.

**Parameter estimate bias due to multi-co-linearity**

The reduction in parameter estimate bias due to selecting the best variance-covariance model specification and non-iterative parameter certainty test bias as discussed above can be further augmented by applying the Likelihood Ratio Test or LRT, to each fixed parameter estimate in the best model. By fitting nested models iteratively to the data with Maximum Likelihood-based estimators, the significance of each parameter effect can be robustly inferred from the change in the minus-2 Log Likelihood statistics of the strictly nested models given the difference in the degrees of model freedom. This Likelihood Ratio Test statistic is distributed as a single Chi-square distribution given that no boundary constraints on the solution to the parameter of interest are encountered during likelihood maximization and any constraints on other parameters are held constant. This test is not biased by predictor variable multi-co-linearity (Self and Liang 1987), but does rely on starting with the best, or nearly best, mixed model specification.

**Parameter estimate bias due to residual centering**

Since the residual centering technique has recently been shown in general to introduce additional bias to the estimates of the main effects and does not improve the parameter estimates of the interactive effects it is therefore completely without merit and potentially misleading and therefore will not be employed as part of this study’s primary inference generating framework. It will only be performed to demonstrate subsequent cautionary illustrative examples.

**Approach Taken**

Data sets from ten empirical studies on brand extension with core attribute perceptions discussed have been made available for further academic research (Holden and Barwise
1995a). Some preliminary measurement model analysis will be conducted to seek support for the additivity and unidimensionality of the ‘fit’ attributes. The data sets are re-analysed in a hierarchical linear model framework using General Linear Mixed Model estimation. The estimates from the GLMM fall into two categories: (a) parameter estimates of fixed main effects and interaction effects, and (b) parameter estimates of random effects and associated variance-covariance structures. Subject, study location and brand extension are modeled as random effects and their associated variances and co-variance structures are estimated and tested. All second-order interaction fixed effects are modeled. The impact of attribute collinearity on parameter estimate significance inference is eliminated by employing Likelihood Ratio Tests distributed as Chi-square on strictly nested models with and without the fixed effect under test. The inferences obtained are therefore robust, least biased of all possible approaches and specifically generalisable to all possible brand extensions, study locations and subjects. The significance of the main effects of consumer perceptions of original brand quality, (Q), and the combined four brand-extension fit attributes of complementarity, (C), substitutability, (S), transferability, (T), and difficulty (D) ratings are inferred at the 95% confidence level.

The results of the above are then compared to those reproduced on individual location study data sets using the OLS and residual centering frameworks used in prior published studies and also using the new proposed modeling frameworks of GLMM and ML. The reliability of the inferences by study location is compared to the inferences made from a GLMM model on the pooled data for each of the three estimating frameworks in the following tables. Note that inferences reported are from this author’s re-analysis on the original data using the frameworks specified, not from any results specifically reported in the primary publications. No outlier removal has been performed on the data, and no imputation of missing data has been performed. All models have an intercept term and this is not shown here and is omitted in the following tables for brevity.

Finally, the effects specification is extended to include the second-order ideal point parameters for all four fit effects, and the eight best covariance models for this effect specification are compared to the best eight models without the ideal point specifications above.
All possible mixed model unstructured variance-covariance structure specifications both at the pooled and at the study location level will be evaluated using an exhaustive grid search, and the best set of eight REML models selected using the AIC criteria.

It is posited that brand extensions with high perceived complementarity, substitutability, transferability or difficulty with respect to the original brand will be more positively evaluated. In addition to the direct main effects noted, brand extensions with a high original brand preference \textit{and also} having perceived high levels of complementarity, substitutability, transferability or difficulty with respect to the original brand will be more positively evaluated.

In an inferential generating framework, the above relationships would be expressed as follows:

Complementarity, substitutability, transferability and difficulty will have significant direct or main effects on brand extension preference and also have significant interactive effects together with original brand preference on brand extension preference.

Casting this in a linear response equation framework suitable for hypotheses testing we have an individual consumer’s brand extension preference model:

\[
y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{10} x_0 x_1 + \beta_{20} x_0 x_2 + \beta_{30} x_0 x_3 + \beta_{40} x_0 x_4 + \epsilon
\]

where

\(y\) is the brand extension preference,

\(\alpha\) is the constant, intercept, or intrinsic preference level for the brand extension that corresponds with unmeasured and therefore extrapolated ‘0’ levels of original brand quality and the three fit attributes on the original 1-to-7 point likert scales. It is therefore not directly interpretable but performs the useful function of absorbing and representing any constant.
level of extension preference in the data that would otherwise bias the remaining parameter estimates,

\( x_0 \) is the original brand preference,

\( \beta_1 \) is the coefficient or part worth of the 1st fit attribute, complementarity, of the extension concept with the original brand concept,

\( x_1 \) is the level or intensity of the 1st fit attribute, complementarity,

\( \beta_2 \) is the coefficient or part worth of the 2nd fit attribute, substitutability, of the extension concept with the original brand concept,

\( x_2 \) is the level or intensity of the 2nd fit attribute, substitutability,

\( \beta_3 \) is the coefficient or part worth of the 3rd fit attribute, transferability, of the extension concept to the original brand concept,

\( x_3 \) is the level or intensity of the 3rd fit attribute, transferability,

\( \beta_4 \) is the coefficient or part worth of the 4th fit attribute, difficulty, of the firm producing the extension concept given its production of the original brand concept,

\( x_4 \) is the level or intensity of the 4th fit attribute, difficulty,

\( \beta_{10} \) is the coefficient or part worth of the 2-way interaction of 1st fit attribute, complementarity, with the original brand concept preference,

\( \beta_{20} \) is the coefficient or part worth of the 2-way interaction of 2nd fit attribute, substitutability, with the original brand concept preference,
\( \beta_{30} \) is the coefficient or part worth of the 2-way interaction of 3rd fit attribute, transferability, with the original brand concept preference,

\( \beta_{40} \) is the coefficient or part worth of the 2-way interaction of non-fit attribute, difficulty, with the original brand concept preference and

\( \varepsilon \) is a vector of random errors, assumed to be homoskedastic and to have a Gaussian independent distribution in the fixed linear model.

A key deduction from the theoretical model developed from Srull and Wyver’s Associative Memory Network Model, categorisation theory and Mandler’s Theory is that the original brand preference evaluation on its own will not directly affect the evaluation of the brand extension, only indirectly via the conditionality of congruence or fit. This may initially seem a counter-intuitive or even shocking deduction to many brand management practitioners, to think that the investment in brand equity generating activities in existing categories is not automatically transferable to their brand in new categories, but this may help explain the high level of brand extension failures of successful original brands noted earlier.

This key deduction is surprisingly in disagreement with the first hypothesis in the seminal empirical paper on brand extension (Aaker and Keller 1990) and similar hypotheses of many of the subsequent related studies, whose initial hypotheses in regression model form was

\[
y = \alpha + \beta_0 x_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{10} x_0 x_1 + \beta_{20} x_0 x_2 + \beta_{30} x_0 x_3 + \beta_{40} x_0 x_4 + \varepsilon
\]

as defined previously on pages 72 and 73 above with the additional definition where:

\( \beta_0 \) is the coefficient or part worth of the original brand concept preference.

Note the difference of an additional second term for the main effect of the original brand quality in this model compared to the previous model equation.
Unsurprisingly, given the theoretical development of the preference model based on Srull and Wyer’s postulates, many of the results of the associated subsequent attempts at empirical generalisation disagreed about the significance of the $\beta_0$ parameter, i.e. the role or part-worth of the original brand preference. Furthermore, disagreements about the significance of the other main and interaction effects were also reported. Causes of the differences amongst the reported significances were frequently posited as co-linearity and additional unmeasured effects.

Surprisingly, whilst many of the ten replication studies suggest correlation or co-linearity as a source of failure to replicate the original results, none discussed or suggested controlling for the possibility that significant covariance structure amongst the interactive effects, i.e. co-linearity amongst the terms $\beta_{10}x_0x_1$, $\beta_{20}x_0x_2$, $\beta_{30}x_0x_3$ and $\beta_{40}x_0x_4$, other than employing the residual centering technique now shown to be inappropriate in this context. Either approach may have biased the Ordinary Least Squares significance tests used mainly as the inference generating framework. Furthermore, misspecification bias (Hair et al. 1998) may have affected the previous studies as many omitted intra-fit interactions and none included second order ideal point effect specifications.

**General Linear Mixed Model frameworks and inference space**

The inference space of the statistical model can be explicitly expanded to cover all study locations and all populations from which subjects were sampled.

Subject and in pooled model, the location, will be specified as random effects and the original variables of interest and all intra-fit effects as either fixed or random-slope, i.e. both fixed and random effects, as determined by model covariance selection (Littell, Milliken, Stroup, and Wolfinger 1996). There are 12 locations associated with the 9 studies analysed in this thesis.

This leads to the model specification to be empirically tested, including the inclusion of random study location effects. The model is then specified as:

$$y = \alpha + \beta_0x_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{10}x_0x_1 + \beta_{20}x_0x_2 + \beta_{30}x_0x_3 + \beta_{40}x_0x_4 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{14}x_1x_4 + \beta_{23}x_2x_3 + \beta_{24}x_2x_4 + \beta_{34}x_3x_4 + \gamma_{i1} + \gamma_{i2} + \gamma_{i3} + \gamma_{i4} + \gamma_{i5} +$$
\[ \gamma_{16} + \gamma_{17} + \gamma_{18} + \gamma_{19} + \gamma_{110} + \gamma_{111} + \gamma_{112} + \gamma_{subj} + \varepsilon \]

Where, in addition to the definitions on pages 72, 73 and 76 above, we have:

- \( \gamma_{11} \) is the random effect for study location 1,
- \( \gamma_{12} \) is the random effect for study location 2,
- \( \gamma_{13} \) is the random effect for study location 3,
- \( \gamma_{14} \) is the random effect for study location 4,
- \( \gamma_{15} \) is the random effect for study location 5,
- \( \gamma_{16} \) is the random effect for study location 6,
- \( \gamma_{17} \) is the random effect for study location 7,
- \( \gamma_{18} \) is the random effect for study location 8,
- \( \gamma_{19} \) is the random effect for study location 9,
- \( \gamma_{110} \) is the random effect for study location 10,
- \( \gamma_{111} \) is the random effect for study location 11,
- \( \gamma_{112} \) is the random effect for study location 12,
- \( \gamma_{subj} \) is the random effect for subjects and
\( \varepsilon \) is a vector of random errors, still assumed to have a Gaussian distribution but now are no longer assumed to be independent and homoskedastic in the general linear mixed model.

**Parameter estimate bias due to heteroskedasticity**

When heteroskedasticity is explicitly modeled in a General Linear Mixed Model, hierarchical and random effects, such as, in this study, locations, brand extensions and subjects, that exhibit less variance can contribute proportionately more information to the parameter estimation process, in the sense of a weighted contribution, for more robust and less biased inference (Diggle 1988). Model covariance structure selection will be performed to identify the best covariance structure, or structures, as there may be more than one best structure. Selection will be performed using the corrected Akaike’s Information Criteria, or AICC with intrinsic penalties for model degrees-of-freedom employed in competing models. This information criteria has been shown to (a) exhibit the best performance in model selection order for linear mixed models over all other efficient information criteria selection methods, included those listed below, and (b) to successfully provide efficient and exact corrections for the asymptotic bias in the AIC in the case of finite samples and general linear mixed models (Hurvich and Tsai 1989).

Other information criteria often used for covariance model selection include Akaike's Information Criterion or AIC (Bozdogan 2000), the Consistent AIC or CAIC (Bozdogan 1987) and the BIC (Schwarz 1978). In the pooled data used in this study the AIC and the AICC are almost identical, differing only by about 0.3 or about 0.001% due to the very large sample size, and so can be used interchangeably when comparing models of the pooled data.

The covariance structure for the empirical model for inference generation then is further extended such that the model specification includes random coefficients, i.e. effects that are specified as both fixed and random effects, otherwise known as random-slope effects. The most general, i.e. unstructured covariance, mixed model specification then becomes:
as defined previously with the additional definitions where:

\[ \gamma_0 \ldots \gamma_4 \] are the five main effect random coefficients,

\[ \gamma_{10} \ldots \gamma_{34} \] are the ten two-way interaction effect random coefficients and

\[ \gamma_{e01} + \ldots + \gamma_{e034} + \ldots + \gamma_{e2434} \] are the 15 x 14 = 210 maximum possible additional covariance parameters.

The above is the most general covariance structure possible which contains many other covariance structures as subsets such as banded or toeplitz, constant, factor analytic and variance component structures, most of which will be tested for significantly improved or more parsimonious fit compared to the unstructured covariance structure outlined above. Unstructured variance-covariance specifications are usually recommended however for pure random coefficient mixed models (Wolfinger 1996).

**Ideal Point or inverted “U” response**

Finally the existence of ideal point or inverted “U” extension evaluation response to the fit dimensions should be investigated, specified as additional fixed and random effects, i.e. random-slope effects. The ideal-point fit response mixed model specification then becomes:
as defined previously with the additional definitions where:

\( \beta_{11} \) is the coefficient or part worth of the polynomial or second-order term for the 1\textsuperscript{st}. fit attributes, complementarity.

\( \beta_{22} \) is the coefficient or part worth of the polynomial or second-order term for the 2\textsuperscript{nd}. fit attributes, substitutability.

\( \beta_{33} \) is the coefficient or part worth of the polynomial or second-order term for the 3\textsuperscript{rd}. fit attributes, transferability.

\( \beta_{44} \) is the coefficient or part worth of the polynomial or second-order term for the 4\textsuperscript{th}. fit attributes, difficulty.

\( \beta_{011} \) is the coefficient or part worth of the second-order term for the 2-way interaction of 1\textsuperscript{st}. fit attribute, complementarity, with the original brand concept preference.

\( \beta_{022} \) is the coefficient or part worth of the second-order term for the 2-way interaction of 2\textsuperscript{nd}. fit attribute, substitutability, with the original brand concept preference.

\( \beta_{033} \) is the coefficient or part worth of the second-order term for the 2-way interaction of 3\textsuperscript{rd}. fit attribute, transferability, with the original brand concept preference.

\( \beta_{044} \) is the coefficient or part worth of the second-order term for the 2-way interaction of 4\textsuperscript{th}. fit attribute, difficulty, with the original brand concept preference.

and

\( \gamma_{011} + \ldots + \gamma_{44044} \) are the additional variance and covariance specifications associated with the unstructured random coefficient specification for the above additional effects.
Other posited effects on brand extension evaluation

There have been numerous other effects posited as important in consumer brand extension evaluation. These include parent brand specific associations (Glynn and Brodie 1998; Pryor and Brodie 1998), the positioning of any previous brand extensions (Dawar and Anderson 1994), brand portfolio specific effects (Dacin and Smith 1994), the level of advertising support (Boush 1993; Pryor and Brodie 1998; Reddy, Holak, and Bhat 1994), price (Taylor and Bearden 2002), and negative information (Romeo 1991a), positive mood (Barone, Miniard, and Romeo 2000), technology hierarchies (Jun, Mazumdar, and Raj 1999), a surprising attempt to generalize the concept of brand credibility in a formal role (Rangaswamy, Burke, and Oliva 1993a) despite the clear contextual variations and even the more surprising antithetical result that “Consumers appeared to construct these inferences on line rather than retrieving them from brand, product-class, subcategory, superordinate category, or exemplar knowledge” (Bristol 1996 p 571). This is not an exhaustive list but merely representative of the activity in this area in the literature. Some recent comprehensive experimental work has recontextualised many of these hypotheses into quantifiably minor or insignificant roles as direct or main effects or interactive effects in brand extension evaluation compared to that of fit or congruence, both in a direct role and also moderated by original brand quality (Volckner and Sattler 2003).

Notwithstanding this generalisation regarding the significance of main and interactive effects, many of these studies support or are broadly consistent with the associative network model of memory and successfully embrace the implications of the extension congruence conditionality and the roles that various additional attribute information presentations (Taylor and Bearden 2002). Many of these studies correctly hypothesise that, consistent with the associative network memory model theory, various salient product attributes will have their effect on the overall evaluation of the brand extension mediated and moderated by extension similarity, fit or congruence. Casting this relationship in an inference generating framework, these product attributes, typically significant as direct or main effects in the context of the original brand products, become indirect, interactive effects in the brand extension evaluation context. Moreover, the main effects that they interact with are the fit or congruence effects. Recall that
the associative network memory model predicts that as the level of incongruence increases, the interactive effect of all these other attributes will increase. This is worth noting that the direction or sign of the interaction is unusual and may initially seem counter-intuitive unless considered in the context of the memory model.

Specifically investigating brand extension with an initial lack of fit to the original brand concept, it has been found that repetition of advertising, and advertising with suitable content that reinforces otherwise weak brand attribute benefit associations can improve initially incongruent brand extension preference evaluation (Boush 1993; Lane 2000; Reddy et al. 1994). This is clearly consistent with the core associative network memory model, as the model predicts that the additional information provided by the advertisements during the phase of high cognitive activity stimulated by the initial highly incongruent evaluation will be stored, retrieved and actively cognitively processed until consistent, ideally positive, concept, attribute and preference evaluations occur. Extension concepts defined or measured as moderately incongruent in this study do not acquire any significant positive evaluative benefit from these treatments in this study. A consistent interpretation is that the moderate level of incongruence is not high enough to invoke the desirable cognitive processing and subsequent concept evaluation reformation processes.

Similar findings consistent with associative network memory model theory as discussed here have been confirmed for the role of price in brand extension evaluation (Taylor and Bearden 2002).

The core associative memory network theory presented here clearly simplifies the understanding of the role of extension product attribute effects as either (a) consistent with the same set of effects of the original brand under conditions of fit or congruence with the brand extension, or (b) interacting with, or moderated by, the original brand effects under conditions of sufficient incongruence or lack of fit between original and extension brand concepts and attributes.

This simplified view is not shared by some who conceptualise the brand affect and product attribute effect paths as remaining separate, competing theoretical literature streams requiring further research (Czellar 2003).
Indeed the positioning of much the research literature on other these other posited effects on brand extensions cited above as apparently unique or distinct from the effects the same attributes have on brands, original, parent or otherwise, can be seen to be less than helpful in some ways for brand extension practitioners, given the existing disagreement about the basic extension evaluation process. Practitioners clearly learn more effectively from as simple and consistent a model as possible, and ideally relating as much as possible to their existing brand management knowledge structure (Vygotskii 1978 pp 86-91).

**Predicted limit outcomes of the associative memory network model**

A helpful thought experiment to conduct with any theoretical model is to use the model to predict extreme limit-approaching scenarios and outcomes and examine the consistency of the predictions with expectations.

Consider the first extreme limit where the brand extension is identical to the original brand. The model proposed here successfully predicts that all of the brand equity will be ‘transferred’, and that the ‘extension’ will be evaluated exactly as the original brand, which is indeed what is expected, as it is the original brand. The ‘perceived’ extent of similarity and substitutability would be extremely high but it is possibly that complementarity would be low or undefined.

The extreme opposite concept consists of a proposed brand extension that is completely incongruous at both the brand affect and attribute evaluation or effect levels, and has no perceived fit on any of the proposed implemented measures or similarity under any condition. In other words, this is not a brand extension, it is in effect a new brand concept that happens to contain a name or logo component of an existing brand. The model proposed here successfully predicts that the new ‘extension’ will inherit no equity from the original brand, and must start life as if it were, and in fact it is in this thought experiment, a new brand, requiring exactly the same brand management for success as any other new brand. The ‘perceived’ extent of similarity and substitutability would be extremely low but possibly that of complementarity would be high by chance.
1. Lack of measurement models used in existing literature

No explicit measurement models, neither Multiple Factor Analysis (Thurstone 1931) nor Rasch Measurement Models (Rasch 1966), were documented in the seminal study of Aaker and Keller nor in any of the nine close replication studies considered hereafter. This means that no analysis has been documented that ensures that the question ratings responses comprised an additive measure and possessed a one-dimensional shared meaning and understanding either within or amongst the populations studied in each case. However pretests were conducted on many of the studies and subject matter experts were consulted so as to increase the likelihood that the attitude and preference measurement questions were meaningful and at least potentially capable of possessing an additive structure. These measures do appear to have some informative taxonomy imposed on them from common theoretical approaches in the disciplines of psychology and economics. Complementarity and substitutability are common demand–sides attribute of goods or services in Economics and are also commonly both classed as generic extrinsic attributes in categorization studies in Psychology. Similarly, transferability (and possibly, to a lesser extent, difficulty) is typically seen as a supply-side firm attribute in Economics and would be classed as intrinsic attributes in categorization studies in Psychology (Ozanne et al. 1992; Smith 1990; Smith and Park 1992).

2. Estimation Frameworks, Techniques and Errors

Ordinary Least Squares

Ordinary Least-Squares regression modeling is used in the seminal study of Aaker and Keller and in all nine close replication studies considered hereafter. None of the studies considered General Linear Mixed Models based on Maximum Likelihood estimation and information criteria. None of the ten studies was able to explicitly expand the inference space to include, and control for, the impact of specific instances of random and repeated effects such as subject effect (there are repeated measures for many brand extension pairs for each subject).
brand-extension pair (these were repeated across study locations amongst the replication) and study location (confounded with cultural norms and sample subpopulations such as business students or general consumers for that location) on the parameter estimates. Parameter significance inferences based on Ordinary Least-Squares regression models have recently been shown to be especially prone to over-estimating the significance of weak instrumental variables in studies with only moderate sample size (Zivot, Startz, and Nelson 1998). This may be contributing to the disagreement amongst the close replications.

**Residual Centering**

Residual centering (Lance 1988) has been suggested as a post-hoc method for eliminating multi-co-linearity in regression models. Say the model of interest is 

\[ y_i = \alpha + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{112}x_1x_2 + \varepsilon, \]

and that there is, naturally, significant multi-co-linearity between the interaction term \( x_1x_2 \) and the two main effects, \( x_1 \) and \( x_2 \). The residual centering procedure involves first using OLS regression to model 

\[ y_2 = \beta_{21}x_1 + \beta_{22}x_2 + \varepsilon_{\text{residual}}, \]

where \( y_2 = x_1x_2 \) and then lastly model

\[ y_1 = \alpha + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{112}x_1x_2 + \varepsilon_{\text{residual}} + \varepsilon. \]

It is easy to see that, in the complete absence of any multi-co-linearity between \( x_1 \) and \( x_2 \), \( \beta_{21} = 0 \), \( \beta_{22} = 0 \) and \( \varepsilon_{\text{residual}} = x_1x_2 \). It is also easy to see that when, as expected, multi-co-linearity amongst the interaction term \( x_1x_2 \) and the two main effects, \( x_1 \) and \( x_2 \) exist, the residual centering process does indeed remove the multi-co-linearities between \( x_1x_2 \) and \( \varepsilon_{\text{residual}} \).

Residual centering has been used in many of the replication studies prior to OLS regression in order to eliminate the multi-co-linearity noted in the data sets (Bottomley and Doyle 1996; Holden and Barwise 1996; Nijssen and Hartman 1994). In a subsequent study this technique has been clearly shown to introduce bias in the estimation and subsequent significance tests of the main effects (Echambadi et al. 2004) and therefore this technique is now shown to be inappropriate for quantitative theory generalisation. The problem of multi-co-linearity that prompted the application of this technique can be dealt with using more robust and validated techniques as described elsewhere in this study. This may also be contributing to the disagreement amongst the close replications.
Basic regression modeling errors affect the published inferences

In addition to differences between sample populations and methods for dealing with multi-co-linearity there were also some basic procedural errors made during the analysis of some of the close replication studies, for example, mis-interpretation of numerical integer codes for missing values as valid response levels (Bottomley and Holden 2001). Although regrettable, it is inevitable that human error occasionally introduces bias into analysis and therefore findings in published results. This study is even more at risk to the introduction of these types of errors as this author’s complete understanding of data sets subject to re-analysis here in turn relies on published documentation and personal communication with the original authors. This risk of mis-interpretation is even greater when it is noted that some of the contributory research was conducted over two decades ago.

6. General Linear Mixed Models

This model framework is also variously known as a hierarchical bayesian model in the statistics literature (Laird and Ware 1982; Rutter and Elashoff 1994; Strenio, Weisberg, and Bryk 1983; Wolfinger 1996), a multi-level linear model (Goldstein 1997 pp 32-45) or an Hierarchical Linear Model (Bryk and Raudenbush 1992 pp 1-4).

The application of this concept to generalizing empirical research findings in the Brand Extension topic area in Marketing has occurred rarely in refereed or peer-reviewed academic publications and is not widely cited in subsequent related works (Bass and Wittink 1975; Glynn and Brodie 1998; Mather and Sunde 1998). The hazards of pooling data across study locations and/or populations without testing for homogeneity and the opportunity to move to a mixed model framework and specify study location as a random effect to facilitate robust pooled inferences despite population heterogeneity have however been clearly highlighted (Bass and Wittink 1975).
7. Proposed Estimation Framework and specification improvement

Intra-Fit interaction effects

Another discussion naturally arises from consideration of the effect of the three fit variables is the impact of modeling any interactions amongst the three main fit effects. As they theoretically share a common underlying dimension it is likely that there are some significant interactions within the group of effects. If there exist significant interactions that are not explicitly included in the regression models used in these replication studies then there are two salient issues that result: (1) any parameter estimates of the originally specified fit effects will absorb or confound with the excluded intra-fit interaction effects and therefore will be biased in the direction of the sign of those effects, as detailed below, and (2) the intra-fit interaction effect parameter estimates will have some useful interpretations for practitioners when planning or generating brand extensions.

Surprisingly, none of the nine close replication studies discussed this natural extension to the theoretical model to be tested, even though the original study did successfully explore this element (Aaker and Keller 1990). If these intra-fit interaction terms are excluded from the theoretical model to be generalized, then the parameter coefficient estimates \( \hat{\beta}_1, \hat{\beta}_2 \) and \( \hat{\beta}_3 \) will be confounded with the true values of \( \beta_{12}, \beta_{13}, \beta_{14}, \beta_{23} \) and \( \beta_{24} \). If these true values are significant, the inferences about the significance of the direct fit effects as part of on attempts at empirical generalisation will be biased.

To guard against the first possibility, all empirical generalisation analysis should include those intra-fit interaction effects in the regression model specification when testing the salient brand extension theory inferences.
Expanding on this second possibility, there may be positive constructive or negative destructive synergies associated with the three types of fit attributes. Any significant positive synergies found can be exploited by designing or selecting extensions that maximise or capitalize on that synergy, and, depending on the mix of signs and intensities of negative intra-fit interaction effect parameter estimates, it may be possible to develop extension strategies to either avoid or otherwise optimally minimize any partial negative contributions to brand extension evaluation.

This leads to a further extension to the formal definition of the model to be tested:

\[
y = \alpha + \beta_{0}x_{0} + \beta_{1}x_{1} + \beta_{2}x_{2} + \beta_{3}x_{3} + \beta_{4}x_{4} + \beta_{10}x_{0}x_{1} + \beta_{20}x_{0}x_{2} + \beta_{30}x_{0}x_{3} + \beta_{40}x_{0}x_{4} + \beta_{12}x_{1}x_{2} + \beta_{13}x_{1}x_{3} + \beta_{14}x_{1}x_{4} + \beta_{23}x_{2}x_{3} + \beta_{24}x_{2}x_{4} + \beta_{34}x_{3}x_{4} + \epsilon
\]

as defined previously in the brand extension fit section above with the additional definitions where:

- \(\beta_{12}\) is the coefficient or part worth of the two-way interaction between the 1st. and 2nd. fit attributes, complementarity and substitutability,

- \(\beta_{13}\) is the coefficient or part worth of the two-way interaction between the 1st. and 3rd. fit attributes, complementarity and transferability,

- \(\beta_{14}\) is the coefficient or part worth of the two-way interaction between the 1st. and the non-fit attributes, complementarity and difficulty,

- \(\beta_{23}\) is the coefficient or part worth of the two-way interaction between the 2nd. and 3rd. fit attributes, substitutability and transferability,

- \(\beta_{24}\) is the coefficient or part worth of the two-way interaction between the 2nd. and the non-fit attributes, substitutability and difficulty and
\( \beta_{34} \) is the coefficient or part worth of the two-way interaction between the 3rd and the nonfit attributes, transferability and difficulty.

8. Alternative and Complimentary Estimating Frameworks

Ridge Regression

Ridge regression (Hoerl and Kennard 1970a, 1970b) is a modeling framework for generating inferences in the presence of ‘troublesome collinearity’ but does not address and is not compatible with the addressing of the other important issues of random effects and heterogeneity for generalisable inferences. It applies only to Ordinary Least-Squares regression, and not to inferences based on Maximum Likelihood information criteria derived from General Linear Mixed Model estimation frameworks. None of the close replication studies discuss or use ridge regression.

Random Coefficient Structural Equation Modeling

Extensions from Structural Equation Modeling, or SEM, to a Random Coefficient SEM (RCSEM) framework (Elrod and Keane 1995) include a somewhat appropriate framework for this context. However non-standard RCSEM, i.e. RCSEM with manifest variables for a two-stage model reduces to a GLMM approach where all effects are specified as random effects, i.e. all effects are either purely random effects as unstructured covariances between manifest variables, or are specified as both fixed and random effects, i.e. random coefficients, otherwise known as random slope effects. These models are a restricted subclass of the models available within the General Linear Mixed Model framework. This restriction is likely to lead to an inability to fully explore and compare desirable model specifications, and also may lead to maximum likelihood estimation stability difficulties where a large proportion of the effects have no significant random effect component estimate. This second shortcoming could, in turn, could be addressed by a large number of a-priori parameter restrictions, but this may be an unnecessarily clumsy, labour-intensive model selection process.
Only one of the five replication studies (Bottomley and Doyle 1996) appears to have considered using Structural Equation Modelling and refers to unpublished results of path analysis on the replication data that supports their published inferences based on an OLS regression framework.

**The Jackknife and the Bootstrap**

Efron (1982) discusses the applicability and necessary conditions of the bootstrap method to regression parameter estimates (p.35), the associated regression model covariance matrix estimate (p36) and the broad applicability and desirability of the properties of the bootstrap parameter bias estimates (p.34) to regression models. These estimating frameworks can be applied to simpler OLS estimation procedures in a process similar to monte-carlo simulation in order to detect and estimate the extent of any bias introduced to the estimation of parameter estimate confidence limits due to breach of necessary assumptions such as multivariate normality or absence of multi-co-linearity. Although generally robust when applied using recommended best practice, these techniques add little value where initial model inferences are generated from Maximum Likelihood based General Linear Mixed Models with appropriate covariance structure specifications that meet the multivariate normality conditions.

Most of these methods will be applied to the data described in the following chapter as appropriate, and the findings arising are discussed in chapter 6.
Chapter 5 Data

The nine studies and their associated data comprising the original brand extension study and eight close replications associated with the data analysed in this thesis were described in the previous literature chapter.

The data for this study consist of nine data sets generously made available for re-analysis by the original authors of a previous study on empirical generalisations (Bottomley and Holden 2001; Holden and Barwise 1995b) and three more recent close replications generously made available by the original authors (Barrat et al. 1999; Fu and Saunders 2002; Patro and Jaiswal 2003). The data are associated with nine studies (Aaker and Keller 1990; Alexandre-Bourhis 1994; Barrat et al. 1999; Bottomley and Doyle 1996; Fu and Saunders 2002; Holden and Barwise 1995b; Nijssen and Hartman 1994; Patro and Jaiswal 2003; Sunde and Brodie 1993) gathered over thirteen locations.

In addition, the final section of this chapter provides some preliminary correlation and related design balance analysis to confirm the need for a mixed or hierarchical model inference framework as motivated and discussed in the previous methodology chapter.

As a representative example, here is how the Sunde and Brodie 1993 study implemented the measurement of the relevant model attributes. The attributes measured are:

1. Original Brand Quality
2. Complementarity
3. Transferability
4. Substitutability
5. Difficulty
6. Brand Extension Preference

Example ratings tasks on a 1-7 scale for a ‘McPhoto’ extension of the ‘McDonalds’ Brand

How do you rate the quality of McDonalds fast food?
Imagine a product called McPhoto:
How does this product complement McDonald's fast food?
How easily can McDonalds transfer their skill and expertise to McPhoto?
How difficult would it be for McDonalds to offer McPhoto service?
To what extent would a McPhoto substitute for a McDonalds fast food item?
How likely are you to buy a McPhoto?

(Sunde and Brodie 1993)

All other studies used for analysis in this thesis implement the items in a similar fashion as they comprise the original publication by Aaker and Keller and a consistent set of admirable attempts to closely replicate the original experiment in many other countries and cultures.

All subsequent studies to the original work are regarded as ‘close replications’ (Lindsay and Ehrenberg 1993) as all authors strived to replicate multiple conditions across studies including approximate sample size, sampling plan, sample population, experimental design, questionnaire length, question framing, measurement scale, original brand saliency and original brand and brand extension evaluation where possible. Editorial license has been applied to the wording of the questions in the tables below as some of the original questions were implemented in an efficient but sometimes disjoint way and possessed a hierarchical structure in the actual questionnaire, rendering exact quoting in table form difficult. Some have further been translated into English from the original language, French (Alexandre-Bourhis 1994; Holden and Barwise 1995b). The following nine sections describe the experimental brand extension data associated with the original study by Aaker and Keller and the eight close replications.

**Aaker and Keller**

This study (Aaker and Keller 1990) employed 6 original brands and 22 brand extensions, and is the original study that stimulated the subsequent attempts at replication. The sample comprised 107 students from Berkeley University in California, USA. The authors chose high quality brands relevant to the student population and formulated extensions with a ‘reasonable’ level of fit on at least one of the three fit dimensions, as determined by prior qualitative and quantitative research involving focus groups and a 100-subject quantitative survey. Note that this condition does not limit the construction of an efficient experimental
The 10-run, nine factor reference design shown on p 103 below meets this criteria and still obtains a 90% A-efficiency. The self-completed questionnaire measured the subjects’ perceptions of original brand quality, fit and difficulty attributes and the overall attitudes towards the extensions. These were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level.

The questions and semantic anchors implemented for each brand and extension are listed below

**Aaker & Keller questions and semantic likert scale anchors**

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor</th>
<th>High anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td>These products are substitutes that I would select between in certain usage situations</td>
<td>Strongly Disagree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>These products are compliments that I would be likely to use together in certain situations</td>
<td>Strongly Disagree</td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>Would the people, facilities and skills used in developing, refining and making the original product be helpful if the manufacturer were to make the extension product?</td>
<td>Not at all helpful</td>
<td>Very Helpful</td>
</tr>
<tr>
<td>What is the overall quality of the original brand</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>What is the likelihood of you trying this extension assuming a purchase was planned in this product class</td>
<td>Not at all Likely</td>
<td>Very Likely</td>
</tr>
<tr>
<td>What do you think the overall quality of the extension would be?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>How difficult would it be, in terms of specialised people, facilities and skills, for a firm to make this extended product</td>
<td>Not at all Difficult</td>
<td>Very Difficult</td>
</tr>
</tbody>
</table>

The original brands and brand extensions used are listed below

**Aaker & Keller Brand Extensions**

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension 1</th>
<th>Extension 2</th>
<th>Extension 3</th>
<th>Extension 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heineken Beer</td>
<td>Light Beer</td>
<td>Wine</td>
<td>Popcorn</td>
<td>N/A</td>
</tr>
<tr>
<td>Vuarnet Sunglasses</td>
<td>Skis</td>
<td>Wallets</td>
<td>Sportswear</td>
<td>Watches</td>
</tr>
<tr>
<td>Haagen-Dazs Ice Cream</td>
<td>Popcorn</td>
<td>Cottage Cheese</td>
<td>Candy Bar</td>
<td>N/A</td>
</tr>
<tr>
<td>Vidal Sassoon Shampoo</td>
<td>Skin Cream</td>
<td>Suntan Lotion</td>
<td>Perfume</td>
<td>Sportswear</td>
</tr>
<tr>
<td>Crest Toothpaste</td>
<td>Mouthwash</td>
<td>Chewing Gum</td>
<td>Shaving Cream</td>
<td>N/A</td>
</tr>
<tr>
<td>McDonalds Meal</td>
<td>Frozen Fries</td>
<td>Theme Park</td>
<td>Photo Processing</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Sunde & Brodie

This study (Sunde and Brodie 1993) closely follows the design of the first study by Aaker and Keller above, employing 6 original brands and 18 brand extensions. This sample comprised 157 students from a university in Auckland, New Zealand. Half of the sample comprised undergraduate business students, the other half were former business students with some employment experience. The authors again chose high quality brands relevant to the student population and formulated extension with a ‘reasonable’ level of fit on at least one of the three fit dimensions, as determined by the authors. The authors also formulated brand extensions that covered a range of product categories and a balance of targeted segments within the student population. These brands and extensions were similar to those in the Aaker and Keller study, allowing for locality-specific branding and distribution differences, and further refined these after a pilot study. The perceptions of original brand quality, fit and difficulty attributes and the overall attitudes towards the extensions were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level. The full design of 6 brands by three extensions was partitioned into two blocks and each respondent only saw one block to keep the average completion time within 15 minutes. The questions and semantic anchors implemented for each brand and extension are listed below

**Sunde & Brodie questions and semantic likert scale anchors**

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor =1</th>
<th>High anchor =7</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely is it that you would select between this original product and this extension in certain usage situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>How likely is it that you would use this original product and this brand extension together in certain situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>Would the people, facilities and skills used in developing, refining and making the original product be helpful if the manufacturer were to make the extension product?</td>
<td>Extremely helpful</td>
<td>Extremely Helpful</td>
</tr>
<tr>
<td>What is the overall quality of the original brand</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>What is the likelihood of you trying this extension assuming a purchase was planned in this product class</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>What do you think the overall quality of the extension would be?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>How difficult would it be, in terms of specialised people, facilities and skills, for a firm to make this extended product class?</td>
<td>Extremely Easy</td>
<td>Extremely Difficult</td>
</tr>
</tbody>
</table>
The original brands and brand extensions used in this study are listed below.

### Sunde and Brodie Brand Extensions

**Table 4 Sunde and Brodie brand extensions**

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension 1</th>
<th>Extension 2</th>
<th>Extension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steinlager Beer</td>
<td>Light Beer</td>
<td>Wine</td>
<td>Popcorn</td>
</tr>
<tr>
<td>Reebok Sports Shoes</td>
<td>Skis</td>
<td>Wallets</td>
<td>Watches</td>
</tr>
<tr>
<td>New American Ice Cream</td>
<td>Popcorn</td>
<td>Cottage Cheese</td>
<td>Chocolate Bar</td>
</tr>
<tr>
<td>Finesse Shampoo</td>
<td>Skin Cream</td>
<td>Sportswear</td>
<td>Perfume</td>
</tr>
<tr>
<td>Twinings Tea</td>
<td>Coffee</td>
<td>Chocolates</td>
<td>Teapots</td>
</tr>
<tr>
<td>McDonalds Meals</td>
<td>Frozen Fries</td>
<td>Theme Park</td>
<td>Photo Processing</td>
</tr>
</tbody>
</table>

### Alexandre-Bourhis

Alexandre-Bourhis (1994) also closely follows the design of the first study by Aaker and Keller above, employing 6 original brands and 18 brand extensions. This sample comprised 126 students from a university in IAC-Caen, France. The sample was blocked into two strata to reduce respondent fatigue. The first strata of 74 respondents evaluated 9 extensions to the brands Heineken, Garnier and Mac Donald’s, and the second strata of 54 respondents evaluated the other 9 extensions to the brands Reebok, Häagen Dazs and Fluocaril. The author chose high quality brands relevant to the sample population and formulated extension with a ‘reasonable’ level of fit on at least one of the three fit dimensions. The author also formulated brand extensions that covered a range of product categories and a balance of targeted segments within the student population whilst also maintaining a balance of global and national brands. These brands and extensions were similar to those in the Aaker and Keller study, allowing for locality-specific branding and distribution differences. The perceptions of original brand quality, fit and difficulty attributes and the overall attitudes towards the extensions were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level.

The questions and semantic anchors implemented for each brand and extension are listed below. These are translated from the original French language and reworded to give the sense of the question without exactly quoting the original, which mentions each original brand and extension explicitly. Neither are the questions listed in the original order used.
Alexandre-Bourhis questions and semantic likert scale anchors

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor</th>
<th>High anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely is it that you would select between this original product and this extension in certain usage situations</td>
<td>Far from probable</td>
<td>Very probable</td>
</tr>
<tr>
<td>Do the original brand product and the extension product appear to you to be products that you could use in a complimentary way?</td>
<td>Far from probable</td>
<td>Very probable</td>
</tr>
<tr>
<td>What is the overall quality of the original brand</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>Would you try this product with this brand by purchasing it?</td>
<td>Far from probable</td>
<td>Very probable</td>
</tr>
<tr>
<td>Do you judge that the overall quality of the extension would be inferior or superior?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>Do you think that the skills and knowledge used in making the original type of product be helpful if the manufacturer were to suddenly make the extension product?</td>
<td>Far from helpful</td>
<td>Very Helpful</td>
</tr>
<tr>
<td>In a general way, do you think the above is a difficult thing?</td>
<td>Far from Difficult</td>
<td>Very Difficult</td>
</tr>
</tbody>
</table>

The original brands and brand extensions used in this study are listed below

Alexandre-Bourhis Brand Extensions

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension 1</th>
<th>Extension 2</th>
<th>Extension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluocaril Toothpaste</td>
<td>Face Wash</td>
<td>Chewing Gum</td>
<td>Shaving Cream</td>
</tr>
<tr>
<td>Garnier Shampoo</td>
<td>Moisturiser</td>
<td>Perfume</td>
<td>Tanning Lotion</td>
</tr>
<tr>
<td>Haagen-Dazs Ice Cream</td>
<td>Popcorn</td>
<td>Cottage Cheese</td>
<td>Candy Bar</td>
</tr>
<tr>
<td>Reebok Sports Shoes</td>
<td>Skis</td>
<td>Wallets</td>
<td>Watches</td>
</tr>
<tr>
<td>Heineken Beer</td>
<td>Light Beer</td>
<td>Wine</td>
<td>Popcorn</td>
</tr>
<tr>
<td>McDonalds Meals</td>
<td>Frozen Fries</td>
<td>Theme Park</td>
<td>Photo Processing</td>
</tr>
</tbody>
</table>

Fu and Saunders

Fu and Saunders (2002) also closely follows the design of the first study by Aaker and Keller above, employing 6 original brands and 18 brand extensions. This sample comprised 470 students from a university in Beijing, China. The sample was blocked into three strata comprising a 50% random of two separate business classes each. Each stratum of respondents evaluated 6 extensions. The author chose high quality brands relevant to the sample population and formulated extension with a 'reasonable' level of fit on at least one of the
three fit dimensions. The author also formulated brand extensions that covered a range of product categories and a balance of targeted segments within the student population. These brands and extensions were similar to those in the Aaker and Keller study, allowing for locality-specific branding and distribution differences. The perceptions of original brand quality, fit and difficulty attributes and the overall attitudes towards the extensions were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level.

The questions and semantic anchors implemented for each brand and extension are listed below

**Fu and Saunders questions and semantic likert scale anchors**

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor =1</th>
<th>High anchor =7</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely is it that you would select between this original product and this extension in certain usage situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>How likely is it that you would use this original product and this brand extension together in certain situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>Would the people, facilities and skills used in developing, refining and making the original product be helpful if the manufacturer were to make the extension product?</td>
<td>Extremely helpful</td>
<td>Extremely Helpful</td>
</tr>
<tr>
<td>What is the overall quality of the original brand</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>What is the likelihood of you trying this extension assuming a purchase was planned in this product class</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>What do you think the overall quality of the extension would be?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>How difficult would it be, in terms of specialised people, facilities and skills, for a firm to make this extended product class?</td>
<td>Extremely Easy</td>
<td>Extremely Difficult</td>
</tr>
</tbody>
</table>
The original brands and brand extensions used in this study are listed below

**Fu and Saunders Brand Extensions**

Table 8 Fu and Saunders Brand Extensions

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension 1</th>
<th>Extension 2</th>
<th>Extension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liang Mian Zheng Toothpaste</td>
<td>Face Wash</td>
<td>Chewing Gum</td>
<td>Shaving Cream</td>
</tr>
<tr>
<td>Head &amp; Shoulders Shampoo</td>
<td>Sports Wear</td>
<td>Perfume</td>
<td>Sun Lotion</td>
</tr>
<tr>
<td>Jie Libo Orange Juice</td>
<td>Mineral Water</td>
<td>Biscuits</td>
<td>Candy Bar</td>
</tr>
<tr>
<td>Seiko Watch</td>
<td>Sunglasses</td>
<td>Popcorn</td>
<td>Walkman</td>
</tr>
<tr>
<td>Qing Dao Beer</td>
<td>Mineral Water</td>
<td>Wine</td>
<td>Popcorn</td>
</tr>
<tr>
<td>McDonalds Meals</td>
<td>Frozen Fries</td>
<td>Theme Park</td>
<td>Photo Processing</td>
</tr>
</tbody>
</table>

**Nijssen & Hartman**

Nijssen and Hartman (1994) also closely follows the design of the first study by Aaker and Keller above. The pre-test employed 40 marketing students resulted in 7 brands and their hypothetical extensions chosen to represent a reasonable balance of original brand quality and fit perceptions. The main study employed those 7 original brands and 24 brand extensions, and the responses were collected from 151 business school subjects at Erasmus University, comprising a sample derived by a scheme weighted by strata to match population demographics of life-cycle stage and home city size of the Netherlands. The stimuli were partitioned into several blocks to reduce respondent fatigue. The perceptions of original brand quality, and three fit attributes, but excluding difficulty, and the overall attitudes towards the extensions were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level.
The questions and semantic anchors implemented for each brand and extension are listed below

**Nijssen & Hartman questions and semantic likert scale anchors**

Table 9 Nijssen & Hartman items and scale anchors

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor =1</th>
<th>High anchor =7</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely is it that you would select between this original product and this extension in certain usage situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>How likely is it that you would use this original product and this brand extension together in certain situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>Would the people, facilities and skills used in developing, refining and making the original product be helpful if the manufacturer were to make the extension product?</td>
<td>Extremely helpful</td>
<td>Extremely Helpful</td>
</tr>
<tr>
<td>What is the overall quality of the original brand</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>What is the likelihood of you trying this extension assuming a purchase was planned in this product class</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>What do you think the overall quality of the extension would be?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
</tbody>
</table>

The original brands and brand extensions used in this study are listed below

**Nijssen & Hartman Brand Extensions**

Table 10 Nijssen & Hartman brand extensions

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension1</th>
<th>Extension 2</th>
<th>Extension 3</th>
<th>Extension 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ariël washing powder</td>
<td>Carpet Cleaner</td>
<td>Disinfecting Soap</td>
<td>Washing Machine</td>
<td>Woolen Sweater</td>
</tr>
<tr>
<td>Kip paté</td>
<td>Canned Vegetables</td>
<td>Cheese</td>
<td>Chicken Curry Salad</td>
<td>N/A</td>
</tr>
<tr>
<td>Kramers Dictionary</td>
<td>Crossword Magazine</td>
<td>Fountain Pen</td>
<td>Opinion Magazine</td>
<td>Translation Computer</td>
</tr>
<tr>
<td>Lätta butter</td>
<td>Bread</td>
<td>Cheese Cake</td>
<td>Licorice</td>
<td>N/A</td>
</tr>
<tr>
<td>Remia mayonnaise</td>
<td>Biscuits</td>
<td>Chips</td>
<td>Orange Juice</td>
<td>N/A</td>
</tr>
<tr>
<td>Van Dale Dictionary</td>
<td>Crossword Magazine</td>
<td>Fountain Pen</td>
<td>Opinion Magazine</td>
<td>Translation Computer</td>
</tr>
<tr>
<td>Zeeuws Meisje butter</td>
<td>Bread</td>
<td>Cheese Cake</td>
<td>Licorice</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Bottomley and Doyle

Bottomley and Doyle (1996) also closely follows the design of the first study by Aaker and Keller above, employing 6 original brands and 18 brand extensions. The data were collected from a convenience sample of 156 student subjects at three locations: 91 subjects at Bath University, 47 subjects at Glasgow University and 18 subjects at Cranfield. The design was partitioned into two blocks of stimuli to reduce respondent fatigue, similar to the Sunde and Brodie study. The perceptions of original brand quality, fit and difficulty attributes and the overall attitudes towards the extensions were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level.

The questions and semantic anchors implemented for each brand and extension are listed below

Bottomley and Doyle’s questions and semantic likert scale anchors

Table 11 Bottomley and Doyle’s items and scale anchors

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor =1</th>
<th>High anchor =7</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely is it that you would select between this original product and this extension in certain usage situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>How likely is it that you would use this original product and this brand extension together in certain situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>Would the people, facilities and skills used in developing, refining and making the original product be helpful if the manufacturer were to make the extension product?</td>
<td>Extremely helpful</td>
<td>Extremely Helpful</td>
</tr>
<tr>
<td>What is the overall quality of the original brand</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>What is the likelihood of you trying this extension assuming a purchase was planned in this product class</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>What do you think the overall quality of the extension would be?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>How difficult would it be, in terms of specialised people, facilities and skills, for a firm to make this extended product class?</td>
<td>Extremely Easy</td>
<td>Extremely Difficult</td>
</tr>
</tbody>
</table>
The original brands and brand extensions used in this study are listed below

**Bottomley and Doyle’s Brand Extensions**

Table 12 Bottomley and Doyle’s Brand Extensions

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension 1</th>
<th>Extension 2</th>
<th>Extension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>McDonalds Meal</td>
<td>Frozen Fries</td>
<td>Theme Park</td>
<td>Photo Processing</td>
</tr>
<tr>
<td>Stella Artois Beer</td>
<td>Light Beer</td>
<td>Wine</td>
<td>Popcorn</td>
</tr>
<tr>
<td>Reebok Sports Shoes</td>
<td>Skis</td>
<td>Wallets</td>
<td>Watches</td>
</tr>
<tr>
<td>Haagen-Dazs Ice Cream</td>
<td>Popcorn</td>
<td>Cottage Cheese</td>
<td>Candy Bar</td>
</tr>
<tr>
<td>Vidal Sassoon Shampoo</td>
<td>Sportswear</td>
<td>Suntan Lotion</td>
<td>Perfume</td>
</tr>
<tr>
<td>Crest Toothpaste</td>
<td>Mouthwash</td>
<td>Chewing Gum</td>
<td>Shaving Cream</td>
</tr>
</tbody>
</table>

**Holden and Barwise**

This study (Holden and Barwise 1995b) also closely follows the design of the first study by Aaker and Keller above, employing 6 original brands and 18 brand extensions. The data were collected from a convenience sample of 289 subjects at three locations: 98 subjects at ESSEC, France, 131 subjects at the London Business School, UK, and 60 subjects at University of Illinois in Chicago, U.S.A.. The design was partitioned into two blocks of stimuli to reduce respondent fatigue, similar to the Sunde and Brodie study. The perceptions of original brand quality, and the four fit attributes and the overall attitudes towards the extensions were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level.
The questions and semantic anchors implemented for each brand and extension are listed below

**Holden and Barwise’s questions and semantic likert scale anchors**

Table 13 Holden and Barwise’s items and scale anchors

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor =1</th>
<th>High anchor =7</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely is it that you would select between this original product and this extension in certain usage situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>How likely is it that you would use this original product and this brand extension together in certain situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>Would the people, facilities and skills used in developing, refining and making the original product be helpful if the manufacturer were to make the extension product?</td>
<td>Extremely helpful</td>
<td>Extremely Helpful</td>
</tr>
<tr>
<td>What is the overall quality of the original brand</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>What is the likelihood of you trying this extension assuming a purchase was planned in this product class</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>What do you think the overall quality of the extension would be?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>How difficult would it be, in terms of specialised people, facilities and skills, for a firm to make this extended product class?</td>
<td>Extremely Easy</td>
<td>Extremely Difficult</td>
</tr>
</tbody>
</table>

The original brands and brand extensions used in this study are listed below

**Holden and Barwise’s Brand Extensions**

Table 14 Holden and Barwise brand extensions

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension 1</th>
<th>Extension 2</th>
<th>Extension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>McDonalds Meal</td>
<td>Frozen Fries</td>
<td>Theme Park</td>
<td>Photo Processing</td>
</tr>
<tr>
<td>Stella Artois Beer</td>
<td>Light Beer</td>
<td>Wine</td>
<td>Popcorn</td>
</tr>
<tr>
<td>Reebok Sports Shoes</td>
<td>Skis</td>
<td>Wallets</td>
<td>Watches</td>
</tr>
<tr>
<td>Haagen-Dazs Ice Cream</td>
<td>Popcorn</td>
<td>Cottage Cheese</td>
<td>Candy Bar</td>
</tr>
<tr>
<td>Vidal Sassoon Shampoo</td>
<td>Sportswear</td>
<td>Suntan Lotion</td>
<td>Perfume</td>
</tr>
<tr>
<td>Crest Toothpaste</td>
<td>Mouthwash</td>
<td>Chewing Gum</td>
<td>Shaving Cream</td>
</tr>
</tbody>
</table>
Patro and Jaiswal

Patro and Jaiswal (2003) also closely follows the design of the first study by Aaker and Keller above, employing 5 original brands and 10 brand extensions. The data were collected from a convenience sample of 106 subjects at the Xavier Labour Relations Institute (XLRI) in Jamshedpur, India, but data from only 88 subjects covering 4 original brands and 8 brand extensions could be located and made available for re-analysis in this study, with apologies from the original authors. Despite the loss of 18 subjects and two brand extensions, being those associated with the Baskins Robins Ice Cream original brand, the reduced size of this data set is still larger than the data sets from Cranfield, Glasgow and Illinois and therefore is still appropriate for inclusion in this study. The authors reduced the brand extension design levels from the original three brand extensions per original brand to two brand extensions per original brand to reduce respondent fatigue. Although the concern for respondent fatigue is admirable, this change leads to a significant decrease in efficiency compared to the original experimental design with respect to maximising the contrasting original brand quality, extension fit and interaction information.

A highly efficient design for estimating all 5 main effects and all 10 2-way interactions is theoretically possible with 16 carefully pretested and contrasting brand extensions as shown previously, but the 10 brand extensions implemented by Patro and Jaiswal can, in theory, at best, only support efficient estimation of the 5 main effects and only the 4 quality-fit 2-way interactions without further limiting the efficiency of the design. A saturated design with 10 runs and an A-efficiency of 90% can easily be found for those requirements.

Table 15 Fractional-factorial reference design for 5 main effects and 4 quality-fit 2-way interactions

<table>
<thead>
<tr>
<th>Run (brand extension index)</th>
<th>Quality</th>
<th>Complement</th>
<th>Substitute</th>
<th>Transfer</th>
<th>Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>
The perceptions of original brand quality, fit and difficulty attributes and the overall attitudes towards the extensions were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level.

The questions and semantic anchors implemented for each brand and extension are listed below

**Patro and Jaiswal questions and semantic likert scale anchors**

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor =1</th>
<th>High anchor =7</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely is it that you would select between this original product and this extension in certain usage situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>How likely is it that you would use this original product and this brand extension together in certain situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>Would the people, facilities and skills used in developing, refining and making the original product be helpful if the manufacturer were to make the extension product?</td>
<td>Extremely helpful</td>
<td>Extremely Helpful</td>
</tr>
<tr>
<td>What is the overall quality of the original brand?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>What is the likelihood of you trying this extension assuming a purchase was planned in this product class</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>What do you think the overall quality of the extension would be?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>How difficult would it be, in terms of specialized people, facilities and skills, for a firm to make this extended product class?</td>
<td>Extremely Easy</td>
<td>Extremely Difficult</td>
</tr>
</tbody>
</table>

The original brands and brand extensions used in this study are listed below

**Patro and Jaiswal Brand Extensions**

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension1</th>
<th>Extension 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ray-Ban Sunglasses</td>
<td>Wallets</td>
<td>Sportswear</td>
</tr>
<tr>
<td>Barista Coffee</td>
<td>Fast Food</td>
<td>Theme Park</td>
</tr>
<tr>
<td>Baskin Robbins Ice Cream</td>
<td>Popcorn</td>
<td>Cottage Cheese</td>
</tr>
<tr>
<td>Clinic-Plus Shampoo</td>
<td>Suntan Lotion</td>
<td>Perfume</td>
</tr>
<tr>
<td>Colgate Toothpaste</td>
<td>Mouthwash</td>
<td>Chewing Gum</td>
</tr>
</tbody>
</table>
Barratt, Lye and Venkateswarlu

Barratt, Lye and Venkateswarlu (1999) also closely follows the design of the first study by Aaker and Keller above, employing 6 original brands and 16 brand extensions. The brand extension were pre-tested on 69 students and business professionals to meet the same criteria as in the original Aaker and Keller study. The data were collected from a systematic sample of 1000 potential respondents obtained by selecting every 2400th. resident on the New Zealand 1996 Electoral Roll. The response rate was nearly 32%, with 319 surveys returned. The sample was relatively representative of the 1996 New Zealand Census population, with only slightly higher proportions of higher educated and female respondents than the national average. The design was partitioned into two blocks of stimuli to reduce respondent fatigue, similar to the Sunde and Brodie study. The perceptions of original brand quality, and the three fit attributes, the difficulty of the firm making the extension and the overall attitudes towards the extensions were all measured on 7-point likert scales where 1 was semantically anchored as a low level and 7 as a high level. The original mailing of the questionnaire with covering letter was followed up ten days later with another letter to initial non-respondents.

Barratt, Lye and Venkateswarlu’s questions and semantic likert scale anchors

<table>
<thead>
<tr>
<th>Question</th>
<th>Low anchor =1</th>
<th>High anchor =7</th>
</tr>
</thead>
<tbody>
<tr>
<td>How likely is it that you would select between this original product and this extension in certain usage situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>How likely is it that you would use this original product and this brand extension together in certain situations</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>Would the people, facilities and skills used in developing, refining and making the original product be helpful if the manufacturer were to make the extension product?</td>
<td>Extremely helpful</td>
<td>Extremely Helpful</td>
</tr>
<tr>
<td>What is the overall quality of the original brand</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>What is the likelihood of you trying this extension assuming a purchase was planned in this product class</td>
<td>Extremely Unlikely</td>
<td>Extremely Likely</td>
</tr>
<tr>
<td>What do you think the overall quality of the extension would be?</td>
<td>Inferior</td>
<td>Superior</td>
</tr>
<tr>
<td>How difficult would it be, in terms of specialised people, facilities and skills, for a firm to make this extended product class?</td>
<td>Extremely Easy</td>
<td>Extremely Difficult</td>
</tr>
</tbody>
</table>
The original brands and brand extensions used in this study are listed below

**Barratt, Lye and Venkateswarlu’s Brand Extensions**

Table 19 Barratt, Lye and Venkateswarlu’s Brand Extensions

<table>
<thead>
<tr>
<th>Original brand</th>
<th>Extension 1</th>
<th>Extension 2</th>
<th>Extension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marque Vue Sparkling Wine</td>
<td>Chocolate Truffles</td>
<td>Still Wine</td>
<td>Orange Juice</td>
</tr>
<tr>
<td>Rolex Watches</td>
<td>Pocket Watch</td>
<td>Wallets</td>
<td>Jet Skis</td>
</tr>
<tr>
<td>Swatch Watches</td>
<td>Pocket Watch</td>
<td>Wallets</td>
<td>Jet Skis</td>
</tr>
<tr>
<td>Dom Perignom Champagne</td>
<td>Chocolate Truffles</td>
<td>Still Wine</td>
<td>Orange Juice</td>
</tr>
<tr>
<td>Harley Davidson Motorcycles</td>
<td>Cars</td>
<td>Bicycles</td>
<td>N/A</td>
</tr>
<tr>
<td>Vespa Scooters</td>
<td>Cars</td>
<td>Bicycles</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Summary of reported fit significance amongst original and close replication studies**

This table below summarises the main findings of the studies and demonstrates the lack of agreement amongst the studies, and hence a failure to empirically generalise to a ‘law of brand extension’. ‘Y’ indicates reported significance and N indicates no reported significance.

Table 20: Summary of reported significance in replication studies at the 95% C.L.

<table>
<thead>
<tr>
<th>Study x effect -&gt;</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>D</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A &amp; K 1990</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>S &amp; B 1993</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>N &amp; H 1994</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>A</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>A-B 1994</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>B &amp; D 1995</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>B, L &amp; V 1999</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>H &amp; B 1995</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>F &amp; S 2002</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>P &amp; J 2003</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td>55%</td>
<td>66%</td>
<td>77%</td>
<td>77%</td>
<td>75%</td>
<td>66%</td>
<td>66%</td>
<td>55%</td>
</tr>
</tbody>
</table>

The pooled agreement is poor. Wherever available, results not biased by the flawed application of the residual centering technique have been reported here, but this has not been possible in all cases. This further reinforces the motivation to improve the consistency and quality of inferences in this research area.
Experimental design efficiencies of the brand extension studies.

An efficient way to characterize the experimental data is to inspect the important qualities of the experimental designs used in the studies.

The experimental design procedures employed by Aaker and Keller and the six other close-replication teams did not result in a completely balanced and contrasting full-factorial design of predictor effects to be empirically tested, and some effects in the design are relatively unbalanced. Authors contrast, from their own perceptions, or from a small-sample pilot study: ‘near’ and ‘far’ fit brand extensions, but explicitly did not select contrasting high and low quality original brands or brand extensions with contrasting levels of within-fit interactions. Only original brands with widespread knowledge and recall amongst the sample populations were included in the studies. Since the designs were not directly balanced, preliminary analysis of the design efficiency of the studies is desirable to assess the extent of imbalance and the associated design efficiency. Note that this assessment of design balance and efficiency will be made on the full 15-effects model as described in the Brand Extension Fit section below, comprising 5 main effects, 4 quality-fit interaction effects and 6 intra-fit interactions effects, not on the 9-effects model with fewer two-way interaction effects explicitly assumed in the original studies. Assuming initially linear and not polynomial or non-linear effects on response, only two levels of response to each effect are required, ideally contrasting brand extensions with means of 1 and 7 on the 7-point rating scales.
A full-factorial design capable of estimating all 5 main effects and all interactions consists of 32 runs or brand extension evaluations per respondent, and is shown for reference purposes below.

Table 21 Full-factorial reference design for 5 linear factors

<table>
<thead>
<tr>
<th>Run (brand extension index)</th>
<th>quality</th>
<th>complement</th>
<th>substitute</th>
<th>transfer</th>
<th>difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>21</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>23</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>25</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>27</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>29</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>31</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>32</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

This design if implemented would have most likely introduced unacceptable levels of respondent fatigue, so some type of fractional factorial or partial experimental design is required.

The original and close replication study authors would have improved the quality and efficiency of the response data if these design issues had been considered. The appropriate summary statistic for the experimental design efficiency for ratio or interval design variables is the A-efficiency statistic, as it is sensitive to the type of linear coding employed. This was
the first type of experimental design efficiency statistic in general use, and is calculated as the ratio of total variances, standardized by the number of replications, between the proposed design and the ideal full factorial design (Cochran and Cox 1950). The ideal A-efficiency of a fractional factorial design is 100%, and several equivalent 100% A-efficient fractional factorial designs can easily be found with 16 runs or contrasting brand extension pairs per respondent.

Table 22 Fractional-factorial reference design for 5 main effects and all 2-way interactions

<table>
<thead>
<tr>
<th>Run (brand extension index)</th>
<th>quality</th>
<th>complement</th>
<th>substitute</th>
<th>transfer</th>
<th>difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Since the design variables are individual’s perceptions and therefore in general heterogeneous amongst subjects, a considerable level of pre-testing and brand-extension pair selection would be required to identify a sufficient number of brand extension pairs that are both homogeneous with respect to the perception ratings of the 15 effects and well balanced with respect to the fractional factorial design. This may also be an impossible goal if a high level of respondent heterogeneity exists with respect to brand extension fit evaluations across these 5 main effects, so the design efficiencies obtained by the efforts of the contributory authors tabled below should not be judged harshly and could well be an artifact of the extant population heterogeneity.

Also concerns about the ecological validity of the design also should be addressed (Doherty and Kurz 1996) which naturally lead to a lowering of design balance and efficiency, moving away from balanced designs to covering designs. This is not a negative statistic, rather a helpful, explicit acknowledgement that there exists a trade-off between sample size for a
given type I and type II error limits and ecological validity in designed experiments, consistent with the themes associated with Social Judgment Theory.

The most efficient and balanced design was found with post-hoc efficiency analysis in the ESSEC data (Holden and Barwise 1995b), and the post-hoc study attribute means for this 19-run, 96 respondent design with a 15.8% A-efficiency are shown below for comparison with the 100% A-efficient fractional-factorial design tabled on the previous page.

Table 23 Post-hoc factor means for the ESSEC data

<table>
<thead>
<tr>
<th>Run (brand extension index)</th>
<th>quality</th>
<th>complement</th>
<th>substitute</th>
<th>transfer</th>
<th>difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMW bike</td>
<td>6.4</td>
<td>3.2</td>
<td>3.5</td>
<td>3.5</td>
<td>4.2</td>
</tr>
<tr>
<td>BMW hifi</td>
<td>6.4</td>
<td>4.8</td>
<td>1.9</td>
<td>3.2</td>
<td>5.7</td>
</tr>
<tr>
<td>BMW outboard motor</td>
<td>6.4</td>
<td>3.0</td>
<td>2.5</td>
<td>5.7</td>
<td>6.1</td>
</tr>
<tr>
<td>Moet truffles</td>
<td>6.3</td>
<td>5.9</td>
<td>3.2</td>
<td>2.5</td>
<td>3.6</td>
</tr>
<tr>
<td>Moet cookbook</td>
<td>6.3</td>
<td>5.1</td>
<td>2.4</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Moet liquor</td>
<td>6.3</td>
<td>4.1</td>
<td>4.3</td>
<td>4.8</td>
<td>4.7</td>
</tr>
<tr>
<td>Haagen-Dazs candy bar</td>
<td>6.2</td>
<td>4.9</td>
<td>5.1</td>
<td>4.8</td>
<td>2.9</td>
</tr>
<tr>
<td>Haagen-Dazs popcorn</td>
<td>6.2</td>
<td>4.1</td>
<td>3.9</td>
<td>2.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Haagen-Dazs cottage cheese</td>
<td>6.2</td>
<td>3.8</td>
<td>4.4</td>
<td>4.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Vidal-Sassoon wallet</td>
<td>5.7</td>
<td>4.9</td>
<td>2.1</td>
<td>2.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Vidal-Sassoon sportswear</td>
<td>5.6</td>
<td>5.8</td>
<td>2.7</td>
<td>2.6</td>
<td>3.4</td>
</tr>
<tr>
<td>Vidal-Sassoon skis</td>
<td>5.6</td>
<td>6.4</td>
<td>1.9</td>
<td>2.3</td>
<td>5.3</td>
</tr>
<tr>
<td>Vidal-Sassoon watches</td>
<td>5.6</td>
<td>5.6</td>
<td>2.6</td>
<td>2.9</td>
<td>4.8</td>
</tr>
<tr>
<td>Heineken lite beer</td>
<td>4.3</td>
<td>4.3</td>
<td>4.2</td>
<td>6.5</td>
<td>3.6</td>
</tr>
<tr>
<td>Heineken popcorn</td>
<td>4.3</td>
<td>4.1</td>
<td>1.8</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Heineken wine</td>
<td>4.3</td>
<td>3.7</td>
<td>2.4</td>
<td>3.3</td>
<td>4.7</td>
</tr>
<tr>
<td>McDonalds frozen fries</td>
<td>3.7</td>
<td>4.4</td>
<td>4.7</td>
<td>5.8</td>
<td>1.8</td>
</tr>
<tr>
<td>McDonalds theme park</td>
<td>3.7</td>
<td>4.7</td>
<td>2.6</td>
<td>3.5</td>
<td>5.3</td>
</tr>
<tr>
<td>McDonalds photo processing</td>
<td>3.7</td>
<td>2.3</td>
<td>1.6</td>
<td>2.0</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Note, for example, that the McPhoto extension exhibited the lowest overall complementarity, substitutability with, and the second lowest transferability from, the original parent McDonalds fast food brand. It also exhibited the (equal) lowest average original brand quality level. Hence its inclusion in the study added positively to the level of contrast in the design and helped to reduce colinearity in the response data more than many of the other extensions, as explained below.

**Absence of co-linearity: a necessary OLS model framework assumption.**

The low efficiencies of the studies’ designs tabled below indicate that the data will have a lack of balance and therefore prone to significant colinearities and covariances that require management during the inference generation process.
Table 24 Original and close replication post-hoc studies’ design efficiencies

<table>
<thead>
<tr>
<th>Data Location</th>
<th>respondents</th>
<th>Brand extensions evaluations per respondent</th>
<th>A-efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Ideal’ design</td>
<td>N/A</td>
<td>16</td>
<td>100</td>
</tr>
<tr>
<td>NZ2</td>
<td>319</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>ESSEC</td>
<td>96</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>LBS</td>
<td>127</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Berkeley</td>
<td>106</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>Illinois</td>
<td>58</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>128</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>Glasgow</td>
<td>46</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>NZ1</td>
<td>183</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>China</td>
<td>470</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>Bath</td>
<td>91</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>Cranfield</td>
<td>18</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>XLRI</td>
<td>88</td>
<td>18</td>
<td>3</td>
</tr>
</tbody>
</table>

These covariances and correlations calculated from the pooled data shown in the two tables following are of such a magnitude that the data does not satisfy the necessary conditions for both homoskedasticity and independence of predictor variables for an OLS regression model and consequently the models have significant parameter estimate and fit statistic bias (Hair et al. 1998). Some correlations are as high as 0.9 and a significant number are greater than 0.7. The covariances exhibit a strong pattern where the main effect variances and covariances are many orders of magnitude smaller than those for the interaction effects. This is to be expected as there has been no scaling of the interactive effects, which are of the order of magnitude of the square of the main effects, and the covariances will be another square order of magnitude greater again, realizing a quartic, of fourth-order, difference in general between the two sub tables of covariances. Therefore an OLS regression framework is a poor inference generating framework choice for the analysis of this data.
ions are congruent with the original brand evaluation attribute evaluation and overall concept preference evaluation. If a brand is evaluated overall as a high preference, or if it is evaluated overall as good for the extension, then this high perceived levels of these 3 extension attributes can only generate positive attribute and overall concept preferences in their own right, i.e. good fit is all directly good for the extension.

It is informative to note here that high perceived levels of these 3 extension attributes can only generate positive attribute and overall concept preferences in their own right, i.e. good fit is all directly good for the extension.

The other key theoretical element relevant to the brand extension evaluation is the original brand concept preference evaluation. If a brand is evaluated overall as a high preference, or equivalently high quality, as in ‘fit for purpose’, and the extension concept and all relevant attribute evaluations are congruent with the original brand evaluation then this high
preference evaluation will transfer to the brand extension. Similarly if a brand is evaluated overall as a low preference, or equivalently low quality, as in ‘not fit for purpose’, and the extension concept is congruent with the original brand then this low preference will transfer to the brand extension.

This chapter has described in detail the 13 data sets from 9 studies comprising the original brand extension study and 8 close replications, and in addition provides the results of preliminary correlation and design balance analysis that confirms the need to employ a general linear mixed or hierarchical linear model estimation framework for the generation of robust inferences.
Chapter 6 Results

Analysing the data presented in the above chapter 5 using the methods outlined in chapter 4 yields the following results.

**Pooled data fit attribute additivity.**

Preliminary common factor analysis is useful to help us judge if the postulated fit attributes of complementarity, substitutability, transferability and difficulty possess additivity and therefore some shared, common meaning along a single ‘fit’ dimension. Exploratory factor analysis was conducted on the three fit items and difficulty using the pooled data and using both the scree slope and eigenvalue unity cutoff methods two underlying dimensions or factors are supported. Various rotation methods were employed yielding a consistent result where the three fit attributes loaded heavily (loadings > 0.6) on one factor and difficulty alone loaded heavily on the other factor. The *standardized* coefficient alpha statistic is appropriate here as all items are measured on the same seven-point scale (Yu 2001). The standardized Cronbach’s Alpha statistic for the pooled nine studies’ data for the four attributes is 0.46 to 2 s.d., and this statistic drops in value if any attribute is excluded except for difficulty. When difficulty is excluded from the common factor analysis on the standardized attributes, the Cronbach’s Alpha statistic increases to 0.56 to 2 s.d. Common factors with Cronbach’s Alpha statistics in excess of 0.50 exhibit a dominance of shared or common meaning over unique meaning and are therefore potentially useful in further analysis that assumes or relies on that common meaning and additivity, as does the empirical implementation of brand extension evaluation based on the theoretical associate memory network model proposed in this thesis.

This inference is reproduced at the study location level of analysis in ten of the eleven study locations, with Cronbach’s Alpha statistics ranging from 0.50 to 0.79. Only the data collected at ESSEC failed to reach the 0.50 level for the Cronbach’s Alpha statistic for the standardized attributes, although even for the data from this location, the exclusion of the difficulty attribute from the common factor specification still increased the statistic, in this case from 0.27 to 0.37. Despite the strong evidence confirming that the difficulty attribute measures a
different underlying construct to fit, most of the analysis will note the result but continue to include and test effects associated with difficulty, as the analysis framework is robust to the inclusion of weak or poor instruments, with and without relatively high and varying levels of variance. Also, the lack of commonality of the difficulty attribute with the ‘fit’ dimension does not, a-priori, preclude the attribute from exhibiting a significant relationship with the brand extension preference attribute, nor does it preclude a useful interpretation for practitioners if this is the case. Similarly, the data collected from ESSEC will continue to be included in the analysis as the General Linear Mixed Model is robust to variations in error variance across strata given the strata effect is specified in the model covariance structure.

1. Pooled Mixed Effects Model Estimation

The primary model of interest is the general linear mixed model fitted to data pooled from twelve of the thirteen close replication study locations, excluding the Nijssen and Hartman Erasmus data, as it did not include measures of the non-‘fit’ attribute, difficulty. The model fixed effects specification includes an intercept and all direct effects as main effects, the indirect effects of quality moderated by the three fit attributes as two-way or second order interaction effects, and lastly the intra-fit interactions as two-way or second order interactions amongst the four ‘fit’ attributes.

The model specification comprises eleven parameters for the eleven fixed effects, up to forty-five parameters for a second order factor-analytic structure of fifteen random between-subject effects and between one and twelve parameters for the within-subject repeated effect, either grouped by data location or ungrouped as specified below. Initially, since the individual location GLMM and OLS models need to be compared on the fairest basis, the ideal point fit effects are omitted from all but the final model selection tables towards the end of this chapter, as all hypothesized effects could not be fully explored in a single fixed effect model framework due to degrees-of-freedom in information constraints. Common inferences are consistent, however, between the best model sets with and without the ideal point specification, with the exception of the Q x T effect significance, which disappears, due to model mis-specification error, in the final model selection.

Some data locations were merged during the model covariance selection process if the estimated location random effect sizes were not significantly different at the 95% confidence
level and the resulting nested, merged covariance structure specification improved the REstricted Maximum Likelihood (REML) statistic significantly at the 95% confidence level using the Likelihood Ratio (LR) statistic. Where seven merged data locations are referred to in the table below, the following data locations have been merged: Bath, ESSEC and NZ1 and also: Cranfield, IAE-Caen and XLRI; Berkeley and LBS.

Many possible covariance structures were specified, and models were successfully estimated from a range of between subject variance-covariance structures including variance component and second-order factor analytic structures. Within-subject residual variances were successfully specified both as a single overall residual within-subject structure and also as residual structures grouped by data location. No models were successfully fitted with unstructured covariance or complimentary symmetric variance structure specifications to the 12 pooled location data.

A subsequent confirmatory model will be fitted to data pooled from all thirteen close replication study locations excluding difficulty effect specifications if difficulty effects are not significant in the primary model. This checks that the inclusion of the Erasmus data is consistent with, and does not alter, the overall inferences from the primary re-analysis of the other 12 data locations.

2. Pooled Mixed Effects Model Covariance Structure Selection

The best eight model random effects specifications all include general or unstructured variance-covariance structures including at least three of the four main or direct fit effects, between one and two of the four quality-fit 2-way interactions and two or four of the six intra-fit 2-way interactions. The best eight covariance models all had an additional within-subject repeated measures variance component specified within each of the twelve data locations. Thus the refined model specification space comprises all of the sixteen fixed effects, up to fifty-six random between-subject attribute covariances, twelve variance component between-subject variances and twelve within-subject residual variances.

The single best model covariance specification comprised a general, or unstructured, variance-covariance structure based on all three of the four main or direct fit effects,
excluding complementarity, one four quality-fit interactions, namely quality*complementarity, three of the six intra-fit second-order interactions, comprising all those formed with complementarity, and variance components for each of the twelve data locations, as well as within-subject residual variance components grouped by those same twelve data locations. This model had the best, i.e. lowest, corrected and uncorrected Akaike Information Criteria, or t-AIC and AICC.

All of the best eight covariance specification model’s restricted likelihoods shown below were arrived at with an optimized likelihood that is not computed on a structural boundary, i.e. with no variance-covariance element estimate set to a structural zero during model estimation. This condition is strictly necessary to calculate valid Likelihood Ratio (LR) statistics as discussed in the following section. Non-iterative, adjusted fixed parameter significance inferences from all models agree, lending a high degree of confidence with the model specification and parameter significance inference even without the application of the final confirmatory LRT.
Pooled GLMM covariance structure model structure

Table 27 Best eight Covariance Structures and associated M-2LL, REML, rAIC, and Pseudo-R-Squared statistics

<table>
<thead>
<tr>
<th>Model index</th>
<th>Covariance Structure</th>
<th>M-2LL</th>
<th>REML</th>
<th>rAIC</th>
<th>Pseudo-R-Squared (M-2LL-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Random: Unstructured Q, S, T, CxS, CxT, CxD, QxS, QxT; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64815.04</td>
<td>64951.3</td>
<td>65316.6</td>
<td>0.113</td>
</tr>
<tr>
<td>2</td>
<td><strong>Random: Unstructured Q, C, S, T, CxT, CxD, TxS, QxT; Repeated: Variance Components 12 data location; subject grouped by data location</strong> *</td>
<td>64962.2</td>
<td>65060.5</td>
<td>65327.5</td>
<td>0.114</td>
</tr>
<tr>
<td>3</td>
<td>Random Unstructured Q, C, S, T, CxT, CxD, CxT, TxS, QxT; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64966.6</td>
<td>65064.8</td>
<td>65331.9</td>
<td>0.114</td>
</tr>
<tr>
<td>4</td>
<td>Random: Unstructured Q, S, T, CxT, CxD, TxS, QxT; Repeated: 12 data location; subject grouped by data location</td>
<td>64993.5</td>
<td>65075.7</td>
<td>65299.2</td>
<td>0.115</td>
</tr>
<tr>
<td>5</td>
<td>Random: Unstructured Q, C, S, T, CxS, CxT, CxD, QxT; Repeated: 12 data location; subject grouped by data location</td>
<td>64994.1</td>
<td>65076.3</td>
<td>65299.8</td>
<td>0.114</td>
</tr>
<tr>
<td>6</td>
<td>Random: Unstructured Q, S, T, CxT, CxD, QxS, QxT; Repeated: 12 data location; subject grouped by data location</td>
<td>64996.6</td>
<td>65078.8</td>
<td>65302.3</td>
<td>0.113</td>
</tr>
<tr>
<td>7</td>
<td>Random: Unstructured Q, C, S, T, CxS, CxD, QxS, QxT; Repeated: 12 data location; subject grouped by data location</td>
<td>64984.4</td>
<td>65082.7</td>
<td>65349.7</td>
<td>0.113</td>
</tr>
<tr>
<td>8</td>
<td>Random: Unstructured Q, S, T, CxT, CxD, TxS, QxS, QxT; Repeated: 12 data location; subject grouped by data location</td>
<td>65005.3</td>
<td>65087.5</td>
<td>65311.0</td>
<td>0.113</td>
</tr>
</tbody>
</table>

Note: * the preferred model highlighted in bold both here and on page 131
Following the discussion in Chapter 3, the significance of each fixed effect parameter coefficient estimate as indicated by the initial T-statistics and associated P-values was *robustly* confirmed in the presence of *any* multi-co-linearity and heteroskedasticity using the Likelihood-Ratio statistics between strictly ‘nested’ models, distributed as a typical double-sided Chi-squared statistic with a single degree of freedom for each parameter.

It is important to recall that the Likelihood Ratio Test is only valid when sufficient conditions exist such that the difference is distributed as a single, i.e. unmixed, Chi-squared distribution. The necessary and sufficient conditions are that the models are strictly nested, i.e. one model specification is completely contained within the other, and that none of the estimates are produced at a likelihood boundary, where fixed or random parameter estimates are further constrained to a ‘structural’ zero (Self and Liang 1987). All variance-covariance model specifications meet this necessary condition of unrestricted likelihood except for the best model, model 1. Also, all of the eight best models agree with respect to the significance or otherwise of all of the fixed effects at the 95% confidence level. The reference or upper level model, which estimates all components contained in the other, nested subsidiary models, and therefore meets the necessary conditions for the reference of the LR tests is model 2. Therefore the parameter estimate significances can be stated with a high degree of confidence for all models and furthermore all of the models except model 1 meet the single chi-square nested model distribution criteria and can be used as the basis of the conclusive Likelihood Ratio Test (LRT). Here is shown the results of the LRT for Model 2.
Pooled GLMM fixed effect T and Likelihood Ratio statistics

Table 28 Best estimate of Significance of Effects at the 95% C.L. using Kenward-Rogers Density Degrees of Freedom for parameter T-statistics and associated P-Value for Model 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter coefficient estimate</th>
<th>Kenward-Rogers method approximation for Density Degrees of Freedom</th>
<th>Full model T-Statistic</th>
<th>Full model Probability difference by chance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>+1.0320</td>
<td>5423</td>
<td>+7.25</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>Q</td>
<td>+0.01868</td>
<td>7986</td>
<td>+0.72</td>
<td>0.4730</td>
</tr>
<tr>
<td>C</td>
<td>+0.1229</td>
<td>10000</td>
<td>+4.40</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>S</td>
<td>+0.2812</td>
<td>2613</td>
<td>+7.28</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>T</td>
<td>+0.07437</td>
<td>3353</td>
<td>+2.23</td>
<td>0.0259</td>
</tr>
<tr>
<td>D</td>
<td>+0.03302</td>
<td>9973</td>
<td>+1.22</td>
<td>0.2211</td>
</tr>
<tr>
<td>CxS</td>
<td>-0.02274</td>
<td>1128</td>
<td>-5.83</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>CxT</td>
<td>+0.006500</td>
<td>1744</td>
<td>+1.66</td>
<td>0.0968</td>
</tr>
<tr>
<td>CxD</td>
<td>+0.006217</td>
<td>3756</td>
<td>+1.69</td>
<td>0.0903</td>
</tr>
<tr>
<td>SxT</td>
<td>-0.01486</td>
<td>1234</td>
<td>-3.50</td>
<td>0.0005</td>
</tr>
<tr>
<td>SxD</td>
<td>-0.00284</td>
<td>8605</td>
<td>-0.66</td>
<td>0.5075</td>
</tr>
<tr>
<td>TxD</td>
<td>-0.00211</td>
<td>11000</td>
<td>-0.53</td>
<td>0.5962</td>
</tr>
<tr>
<td>QxC</td>
<td>+0.02280</td>
<td>10000</td>
<td>+5.27</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>QxS</td>
<td>+0.01406</td>
<td>5561</td>
<td>+2.63</td>
<td>0.0085</td>
</tr>
<tr>
<td>QxT</td>
<td>+0.02356</td>
<td>3197</td>
<td>+4.50</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>QxD</td>
<td>-0.00146</td>
<td>12000</td>
<td>-0.32</td>
<td>0.7493</td>
</tr>
</tbody>
</table>

The direct effects parameter estimates for quality and difficulty fail to achieve significance at the 95% confidence level using the best-practice mixed-model parameter T-tests, as do all interaction effects involving the difficulty construct.

Table 29 Significance of Effects at the 95% C.L. using Likelihood Ratio statistics between strictly nested models

<table>
<thead>
<tr>
<th>Parameter omitted in nested model</th>
<th>-2LL full model</th>
<th>-2LL nested model</th>
<th>Likelihood Ratio</th>
<th>Difference in Degrees of Freedom</th>
<th>Probability of difference by chance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>64825.9</td>
<td>64826.4</td>
<td>0.5</td>
<td>1</td>
<td>0.480</td>
</tr>
<tr>
<td>C</td>
<td>64825.9</td>
<td>64845.2</td>
<td>19.3</td>
<td>1</td>
<td>&lt; 0.0005</td>
</tr>
<tr>
<td>S</td>
<td>64825.9</td>
<td>64878.9</td>
<td>53.0</td>
<td>1</td>
<td>&lt; 0.0005</td>
</tr>
<tr>
<td>T</td>
<td>64825.9</td>
<td>64830.9</td>
<td>5.0</td>
<td>1</td>
<td>0.025</td>
</tr>
<tr>
<td>D</td>
<td>64825.9</td>
<td>64827.4</td>
<td>1.5</td>
<td>1</td>
<td>0.221</td>
</tr>
<tr>
<td>QxC</td>
<td>64825.9</td>
<td>64853.7</td>
<td>27.8</td>
<td>1</td>
<td>&lt; 0.0005</td>
</tr>
<tr>
<td>QxS</td>
<td>64825.9</td>
<td>64832.9</td>
<td>7.0</td>
<td>1</td>
<td>0.008</td>
</tr>
<tr>
<td>QxT</td>
<td>64825.9</td>
<td>64846.1</td>
<td>20.2</td>
<td>1</td>
<td>&lt; 0.0005</td>
</tr>
<tr>
<td>QxD</td>
<td>64825.9</td>
<td>64826.0</td>
<td>0.1</td>
<td>1</td>
<td>0.752</td>
</tr>
</tbody>
</table>
The inferences generated by the robust Likelihood-Ratio Chi-Squared tests are completely congruent with those generated by the best-practice T-test statistics in the best full model, therefore both LR- and T- statistic-based inferences will be used as the reference inferences for the comparisons amongst study location inference frameworks in the following tables. Intrafit interaction effects are not shown as the LRT is a sequential test and unnecessary repeated application of the test is not recommended in order to avoid inflation of the type I error rate.

The P-values for the Likelihood Ratio statistic significance of the transferability and difficulty attributes are in close agreement, within a 0.001 difference, to the P-values for the T-statistics using the Kenward-Rogers density degrees of freedom approximation, further confirming that minimal bias has been introduced into the critical statistics for these effects due to multi-collinearity or unspecified heteroskedasticity.

The proportion of agreement amongst inferences by location with the GLMM inferences from the pooled data is shown in the next 10 tables. The significance or otherwise and agreement proportions marked with a ‘+’ or a ‘−’ indicates positive (desirable) or negative (undesirable) changes in agreement amongst estimating frameworks with respect to table 39, the reference table using the ‘best’ inference framework, albeit still individually with insufficient sample size to achieve reliability as discussed later in this thesis. Note as the inference frameworks improves in applicability, the pooled agreement levels, although not perfect, also improve.
3. Model Estimation

Fixed effects OLS with Difficulty but without Residual Centering or intra-fit interactions

Table 30 Significance of Effects at the 95% C.L. by Study Location using an OLS Fixed Effects Model Estimation Framework without Residual Centering or intra-fit interactions but including Difficulty

<table>
<thead>
<tr>
<th>Location x effect -&gt;</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>D</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
<th>QxD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated pooled effect size, 2 s.d.</td>
<td>0.02</td>
<td>+0.12</td>
<td>+0.28</td>
<td>+0.07</td>
<td>+0.03</td>
<td>+0.02</td>
<td>+0.01</td>
<td>+0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Bath</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Berkley</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Cranfield</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>ESSEC</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Glasgow</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LBS</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NZ</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NZ2</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>XLRI</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td>83%</td>
<td>42%</td>
<td>33%</td>
<td>17%</td>
<td>83%</td>
<td>25%</td>
<td>0%</td>
<td>42%</td>
<td>100%</td>
</tr>
</tbody>
</table>
## OLS Fixed effects with Residual Centering and Difficulty but without intra-fit interactions

Table 31 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework with Residual Centering

<table>
<thead>
<tr>
<th>Location x effect</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>D</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
<th>QxD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bath</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>Y+</td>
<td>Y-</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Berkley</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>Y+</td>
<td>Y-</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Cranfield</td>
<td>Y-</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>ESSEC</td>
<td>Y</td>
<td>Y+</td>
<td>Y+</td>
<td>Y+</td>
<td>Y-</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Glasgow</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>N</td>
<td>N+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>N</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LBS</td>
<td>Y-</td>
<td>Y+</td>
<td>Y+</td>
<td>Y+</td>
<td>Y-</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NZ</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NZ2</td>
<td>Y-</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>XLRI</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td>58%-</td>
<td>100%+</td>
<td>100%+</td>
<td>83%+</td>
<td>58%-</td>
<td>25%</td>
<td>0%</td>
<td>42%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Note the very high levels of positive estimate bias introduced into the main effect estimates by the erroneous application of the residual centering technique in this context.
Fixed effects with Difficulty and intra-fit interactions but without residual centering

Table 32 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework with Difficulty but without Residual Centering

<table>
<thead>
<tr>
<th>Location x effect</th>
<th>Subjects</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>D</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
<th>QxD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated pooled effect size, 2 s.d.</td>
<td></td>
<td>0.02</td>
<td>+0.12</td>
<td>+0.28</td>
<td>+0.07</td>
<td>+0.03</td>
<td>+0.02</td>
<td>+0.01</td>
<td>+0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Bath</td>
<td>91</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Berkley</td>
<td>106</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>470</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y-</td>
</tr>
<tr>
<td>Cranfield</td>
<td>18</td>
<td>N</td>
<td>N</td>
<td>N-</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>ESSEC</td>
<td>98</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>Glasgow</td>
<td>47</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>N</td>
<td>N+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>128</td>
<td>Y</td>
<td>N-</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y-</td>
<td></td>
</tr>
<tr>
<td>Illinois</td>
<td>60</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y-</td>
<td></td>
</tr>
<tr>
<td>LBS</td>
<td>131</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>NZ</td>
<td>186</td>
<td>N</td>
<td>Y</td>
<td>N-</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>NZ2</td>
<td>319</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XILR</td>
<td>88</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled all locations</td>
<td>1729</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td></td>
<td>83%</td>
<td>42%</td>
<td>33%+</td>
<td>17%</td>
<td>17%+</td>
<td>25%</td>
<td>8%+</td>
<td>42%</td>
<td>75%-</td>
</tr>
</tbody>
</table>
### Mixed effects with Difficulty but without intra-fit interactions and Residual Centering

Table 33 Significance of Effects at the 95% C.L. by Study Location using a Mixed Effects Model Estimation Framework without Residual Centering

<table>
<thead>
<tr>
<th>Location x effect -&gt;</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>D</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
<th>QxD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimated pooled effect size, 2 s.d.</strong></td>
<td>0.02</td>
<td>+0.12</td>
<td>+0.28</td>
<td>+0.07</td>
<td>+0.03</td>
<td>+0.02</td>
<td>+0.01</td>
<td>+0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Bath</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Berkley</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Cranfield</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>ESSEC</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Glasgow</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N+</td>
<td>N-</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>N+</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LBS</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>N-</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NZ</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NZ2</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>XLRI</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td><strong>Pooled Agreement</strong></td>
<td>92%+</td>
<td>42%</td>
<td>33%</td>
<td>17%</td>
<td>92%+</td>
<td>25%</td>
<td>17%+</td>
<td>25%-</td>
<td>100%</td>
</tr>
</tbody>
</table>
Mixed effects with Difficulty and intra-fit interactions but without residual centering

Table 34 Significance of Effects at the 95% C.L. by Study Location using a Mixed Effects Model Estimation Framework with Difficulty but without Residual Centering

<table>
<thead>
<tr>
<th>Location x effect</th>
<th>Subjects</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>D</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
<th>QxD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated pooled effect size, 2 s.d.</td>
<td></td>
<td>0.02</td>
<td>+0.12</td>
<td>+0.28</td>
<td>+0.07</td>
<td>+0.03</td>
<td>+0.02</td>
<td>+0.01</td>
<td>+0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Bath</td>
<td>91</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Berkley</td>
<td>106</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>470</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y-</td>
</tr>
<tr>
<td>Cranfield</td>
<td>18</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>N-</td>
<td>N</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>ESSEC</td>
<td>98</td>
<td>N+</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>N</td>
<td>Y+</td>
<td>N</td>
</tr>
<tr>
<td>Glasgow</td>
<td>47</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>N+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>128</td>
<td>N+</td>
<td>N-</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>60</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y-</td>
</tr>
<tr>
<td>LBS</td>
<td>131</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NZ</td>
<td>186</td>
<td>N</td>
<td>Y</td>
<td>N-</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NZ2</td>
<td>319</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>XILR</td>
<td>88</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Pooled all locations</td>
<td>1729</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td></td>
<td>100%+</td>
<td>42%</td>
<td>42%+</td>
<td>33%+</td>
<td>92%+</td>
<td>50%+</td>
<td>17%+</td>
<td>50%+</td>
<td>83%+</td>
</tr>
</tbody>
</table>
### OLS without Residual Centering or intra-fit interactions or Difficulty

Table 35 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework without Residual Centering or intra-fit interaction or Difficulty

<table>
<thead>
<tr>
<th>Location</th>
<th>x</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bath</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Berkley</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Cranfield</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Erasmus</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>ESSEC</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Glasgow</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LBS</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NZ</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>NZ2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>XLRI</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td>69%</td>
<td>38%</td>
<td>38%</td>
<td>23%</td>
<td>23%</td>
<td>15%</td>
<td>54%</td>
<td></td>
</tr>
</tbody>
</table>

### With Residual Centering but without Difficulty and intra-fit interactions

Table 36 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework with Residual Centering

<table>
<thead>
<tr>
<th>Location x effect</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bath</td>
<td>N+</td>
<td>Y</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Berkley</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Cranfield</td>
<td>Y-</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Erasmus</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>ESSEC</td>
<td>Y</td>
<td>Y+</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Glasgow</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>Y-</td>
<td>Y+</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>N</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LBS</td>
<td>Y-</td>
<td>Y+</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NZ</td>
<td>Y-</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>NZ2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>XLRI</td>
<td>Y-</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td>38%-</td>
<td>92%+</td>
<td>100%+</td>
<td>85%+</td>
<td>25%</td>
<td>17%</td>
<td>54%</td>
</tr>
</tbody>
</table>
Fixed effects with intra-fit interactions but without Difficulty and residual centering

Table 37 Significance of Effects at the 95% C.L. by Study Location using a Fixed Effects Model Estimation Framework without Difficulty and Residual Centering

<table>
<thead>
<tr>
<th>Location x effect</th>
<th>Subjects</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated pooled effect size, 2 s.d.</td>
<td>0.02</td>
<td>+0.12</td>
<td>+0.28</td>
<td>+.11</td>
<td>+.02</td>
<td>+.01</td>
<td>+.02</td>
<td></td>
</tr>
<tr>
<td>Bath</td>
<td>91</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Berkley</td>
<td>106</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>470</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Cranfield</td>
<td>18</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Erasmus</td>
<td>119</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>ESSEC</td>
<td>98</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Glasgow</td>
<td>47</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>128</td>
<td>N</td>
<td>N-</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>60</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LBS</td>
<td>131</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>NZ</td>
<td>186</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>NZ2</td>
<td>319</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>XILR</td>
<td>88</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Pooled locations</td>
<td>1848</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td>69%</td>
<td>38%+</td>
<td>38%</td>
<td>31%+</td>
<td>23%</td>
<td>8%-</td>
<td>54%</td>
<td></td>
</tr>
</tbody>
</table>
Mixed effects without intra-fit interactions or Residual Centering or Difficulty

Table 38 Significance of Effects at the 95% C.L. by Study Location using a Mixed Effects Model Estimation Framework without intra-fit interactions or residual centering

<table>
<thead>
<tr>
<th>Location x effect -&gt;</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bath</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Berkeley</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Cranfield</td>
<td>Y-</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Erasmus</td>
<td>Y</td>
<td>N</td>
<td>N-</td>
<td>Y</td>
<td>N</td>
<td>N-</td>
<td>Y</td>
</tr>
<tr>
<td>ESSEC</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>Y</td>
</tr>
<tr>
<td>Glasgow</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>Y+</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>LBS</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>Y</td>
</tr>
<tr>
<td>NZ</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NZ2</td>
<td>N+</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>XLRI</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td>69%</td>
<td>38%</td>
<td>38%</td>
<td>31%+</td>
<td>23%</td>
<td>15%</td>
<td>54%</td>
</tr>
</tbody>
</table>
Mixed effects with intra-fit interactions but without Difficulty or residual centering

Table 39 Significance of Effects at the 95% C.L. by Study Location using a Mixed Effects Model Estimation Framework without residual centering or Difficulty

<table>
<thead>
<tr>
<th>Location x effect</th>
<th>Subjects</th>
<th>Q</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>QxC</th>
<th>QxS</th>
<th>QxT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated pooled effect size, 2 s.d.</td>
<td>0.02</td>
<td>+0.12</td>
<td>+0.28</td>
<td>+.11</td>
<td>+.02</td>
<td>+.01</td>
<td>+.02</td>
<td></td>
</tr>
<tr>
<td>Bath</td>
<td>91</td>
<td>N+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Berkley</td>
<td>106</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>China</td>
<td>470</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Cranfield</td>
<td>18</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Erasmus</td>
<td>119</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N-</td>
<td>Y+</td>
</tr>
<tr>
<td>ESSEC</td>
<td>98</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Glasgow</td>
<td>47</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y+</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>IAE-Caen</td>
<td>128</td>
<td>N</td>
<td>N-</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Illinois</td>
<td>60</td>
<td>N</td>
<td>N</td>
<td>N-</td>
<td>N</td>
<td>Y+</td>
<td>Y+</td>
<td>Y+</td>
</tr>
<tr>
<td>LBS</td>
<td>131</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>Y</td>
<td>Y+</td>
</tr>
<tr>
<td>NZ</td>
<td>186</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NZ2</td>
<td>319</td>
<td>N+</td>
<td>Y</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>XILR</td>
<td>88</td>
<td>N</td>
<td>Y</td>
<td>Y+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Pooled locations</td>
<td>1848</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Pooled Agreement</td>
<td>85%+</td>
<td>33%-</td>
<td>38%</td>
<td>38%+</td>
<td>38%+</td>
<td>33%+</td>
<td>62%+</td>
<td></td>
</tr>
</tbody>
</table>
Pooled GLMM covariance structure model with ideal point specifications

Table 40 Best eight Ideal point Covariance Structures and associated ML, REML, r-AIC and Pseudo-R-Squared statistics

<table>
<thead>
<tr>
<th>Model index</th>
<th>Covariance Structure</th>
<th>M-2LL</th>
<th>REML</th>
<th>r-AIC</th>
<th>Pseudo-R-Squared (M-2LL-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Random: Unstructured Q, S, T, CxS, CxT, CxD, QxS, QxT QxSxS; Repeated: Variance Components 12 data location; subject grouped by data location *</td>
<td>64730.7</td>
<td>64876.3</td>
<td>64992.3</td>
<td>0.115</td>
</tr>
<tr>
<td>2</td>
<td>Random: Unstructured Q, S, T, CxS, CxT, CxD, QxS, QxT SxS; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64731.5</td>
<td>64877.0</td>
<td>64993.0</td>
<td>0.115</td>
</tr>
<tr>
<td>3</td>
<td>Random Unstructured Q, S, T, CxT, CxD, QxS, QxT QxSxS; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64768.0</td>
<td>64913.8</td>
<td>65011.8</td>
<td>0.113</td>
</tr>
<tr>
<td>4</td>
<td>Random Unstructured Q, S, T, CxT, CxD, QxS, QxT SxS; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64770.2</td>
<td>64915.9</td>
<td>65013.9</td>
<td>0.113</td>
</tr>
<tr>
<td>5</td>
<td>Random Unstructured Q, S, T, CxT, CxD, QxS, QxT QxTxT; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64753.0</td>
<td>64898.6</td>
<td>65014.6</td>
<td>0.114</td>
</tr>
<tr>
<td>6</td>
<td>Random Unstructured Q, S, T, CxT, CxD, QxS, QxT SxS, TtT; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64754.5</td>
<td>64900.1</td>
<td>65016.1</td>
<td>0.114</td>
</tr>
<tr>
<td>7</td>
<td>Random Unstructured Q, S, T, CxS, CxT, CxD, QxS, QxT; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64777.6</td>
<td>64923.3</td>
<td>65021.3</td>
<td>0.113</td>
</tr>
<tr>
<td>8</td>
<td>Random Unstructured Q, S, T, CxS, CxT, CxD, QxS, QxT TtT; Repeated: Variance Components 12 data location; subject grouped by data location</td>
<td>64761.3</td>
<td>64915.2</td>
<td>65031.2</td>
<td>0.114</td>
</tr>
</tbody>
</table>
All these best eight models are better than the equivalent best eight models tabled initially in table 27 on all information criteria, including the M-2LL, r-AIC and pseudo-R-squared model fit statistics. The only difference between the common or overlapping effect inferences is for that of the Q x T parameter coefficient estimate, which is significant in the models without the ideal point specifications, and which disappears in the presence of the additional second-order ideal point fit effect specifications that further and significantly improve model fit and information criteria. The estimated type II error rate of the Q x T estimate in the best model is relatively low. This leads us to firmly deduce that the Q x T significance is a uniform artifact of model mis-specification in the earlier models.
Table 41 Best estimate of Significance of Effects at the 95% C.L. using Kenward-Rogers Density Degrees of Freedom for parameter T-statistics and associated P-Value for ideal point model 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter coefficient estimate</th>
<th>Kenward-Rogers method approximation for Density Degrees of Freedom</th>
<th>Full model T-Statistic</th>
<th>Full model Probability difference by chance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>+0.6549</td>
<td>5678</td>
<td>+3.62</td>
<td>0.0003</td>
</tr>
<tr>
<td>Q</td>
<td>+0.04663</td>
<td>8928</td>
<td>+1.36</td>
<td>0.1752</td>
</tr>
<tr>
<td>C</td>
<td>+0.1640</td>
<td>13000</td>
<td>+4.42</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>S</td>
<td>+0.2970</td>
<td>1357</td>
<td>+6.46</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>T</td>
<td>+0.3400</td>
<td>9033</td>
<td>+3.37</td>
<td>0.0008</td>
</tr>
<tr>
<td>D</td>
<td>+0.02898</td>
<td>9339</td>
<td>+1.19</td>
<td>0.2346</td>
</tr>
<tr>
<td>CxS</td>
<td>-0.02055</td>
<td>1054</td>
<td>-5.07</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>CxT</td>
<td>+0.008167</td>
<td>1860</td>
<td>-2.12</td>
<td>0.0345</td>
</tr>
<tr>
<td>SxS</td>
<td>-0.01354</td>
<td>1721</td>
<td>-2.62</td>
<td>0.0090</td>
</tr>
<tr>
<td>SxT</td>
<td>-0.00090</td>
<td>5219</td>
<td>-0.21</td>
<td>0.8374</td>
</tr>
<tr>
<td>TxT</td>
<td>-0.04314</td>
<td>9450</td>
<td>-3.25</td>
<td>0.0012</td>
</tr>
<tr>
<td>CxC</td>
<td>-0.00492</td>
<td>14000</td>
<td>-1.21</td>
<td>0.2268</td>
</tr>
<tr>
<td>QxC</td>
<td>+0.02375</td>
<td>9508</td>
<td>+5.48</td>
<td>Less than 0.0001</td>
</tr>
<tr>
<td>QxS</td>
<td>+0.01308</td>
<td>955</td>
<td>+2.23</td>
<td>0.0262</td>
</tr>
<tr>
<td>QxD</td>
<td>-0.00512</td>
<td>11000</td>
<td>-0.27</td>
<td>0.7857</td>
</tr>
<tr>
<td>QxTxT</td>
<td>+0.004167</td>
<td>11000</td>
<td>+1.72</td>
<td>0.0861</td>
</tr>
</tbody>
</table>

This table shows the parameter estimates and associated statistics for the best “inverted U” ideal point response specification model from table 40 (model 1, in bold). This model supports the best and most empirically generalisable inferences this thesis can make given the available pooled data. These inferences are discussed in detail in the following chapters.
4. Type I and II error and sample size of the GLMM parameter tests

The parameter estimate results shown in Table 41 were obtained from a pooled sample size of 1729 subjects drawn from 12 different location sample frames with up to 20 extension evaluation observations each. The Erasmus data was excluded as difficulty was not included in that replication. 17041 observations were used for the pooled reference model estimation. Sample size was then systematically reduced by observation within subject within location, and the model was re-estimated until the first effect significance inference at the 95% confidence level failed to agree with the reference model inferences. The first error at the 95% confidence level identified was a type I error, i.e. a decision to reject the null hypothesis when the null hypothesis was true. This occurred at all sample sizes less than 10 location sample frames, 1180 subjects and 12885 observations. At that boundary the difficulty parameter effect estimate, whose true value can be confidently assumed to be 0 based from all models estimated with more data, was associated with a probability of a non-zero estimate by chance of 5.4%, i.e. less than the usually acceptable 95% confidence level for this inference using the kenward-roger denominator-degrees-of-freedom for the T-test. Similarly, sample size was further reduced until an opportunity to reject a null hypothesis correctly was lost, i.e. a type II error occurred, at the 95% confidence level. This occurred at all sample sizes less than 9 location sample frames, 1085 subjects and 11567 observations. At that boundary, the quality*substitutability parameter estimate, whose true value can be confidently estimated to be 0.01 from all models estimated with more data, was associated with a probability of a non-zero estimate by chance of 5.16%, i.e. less than the usual 95% confidence level, using the kenward-roger estimated denominator-degrees-of-freedom for the T-test.

Additionally, the bootstrap estimation method (Efron 1982; Efron 1993 pp 220-32; Shao and Tu 1995 pp 376-78) was applied to conservatively and robustly estimate the type II error rate, and the power of the combined alternative hypotheses inferences.
Table 42 Type II error rates derived from bootstrap estimates

<table>
<thead>
<tr>
<th>Effect label</th>
<th>Observed Type II error Statistic in original model at 95% C.L.</th>
<th>Bootstrap Estimated Type II Error rate</th>
<th>Minimum Resampled estimate</th>
<th>Maximum Resampled Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>C x C</td>
<td>0</td>
<td>0.35</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0.27</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Q x C</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Q x S</td>
<td>0</td>
<td>0.91</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Q x T</td>
<td>0</td>
<td>0.12</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S x S</td>
<td>0</td>
<td>0.76</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T x T</td>
<td>0</td>
<td>0.15</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>0.10</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

These estimates of type II error rate are biased as all 113 bootstrap iterations employ resamples with replacement of 90% of the original data set, made up to the original size (17014 observations used), which we know will reduce the information content and therefore the effective power of the tests as demonstrated above. However the bias is in the conservative direction, i.e. we know that the actual type II error rates will be at least as good as or lower than those estimated in the table above. Power is the 1’s complement of the type II error rate, and statistical powers of 0.8 or 0.7 are generally acceptable (Hair et al. 1998), corresponding to type II error rates of 0.20 or 0.30 respectively. Even by making a small allowance for the conservative bias inherent in these estimates, there are two inferences in the final model which require further discussion. These are the second-order ideal point coefficient for substitutability and the moderating effect of substitutability on original brand quality. This is not surprising as Q x S effect has the lowest ratio of significant effect estimate to covariance, followed by the second order ideal point effect estimate for substitutability. This result is consistent with the random sample-size reduction findings reported above.

One of the corollaries of Claude E. Shannon’s Information Theory (Shannon 1948) is that the maximum amount of information is gained by a system of tests when the alternative outcomes are equi-probable. Thus we can see that if the risk of type I and type II errors are (as close as is practicable to being) equal then we can deduce that we are (as close as is practical to) gaining the most information from the inference generating system that comprises the GLMMs employed here in this study. Although the type II error rates are not equal to the type
I error rates (the probability of being different by chance) shown in table 42 above on an effect-by-effect comparison basis (and cannot in general be controlled to be so), they cover approximately the same range of values and so represent approximately the best model realizable in term of (inference) information extraction from the data.

3. Intra-fit interaction effects.

Note that the parameter estimate results shown in Tables 28 and 41 above for complementarity*substitutability and substitutability*transferability are both negative and significant at the 95% confidence level. Although not directly relevant to the main theory-centered research questions, these results will have implications for practitioners as discussed below.

4. Ideal point fit specification effects.

Note that the parameter estimate results shown in Table 41 above for complementarity*complementarity and transferability*transferability are both negative and significant at the 95% confidence level. These results are relevant to the main research questions, and thus will have implications for market researchers and brand management practitioners as discussed below.

This chapter presents the findings of the analysis of the data described in chapter 5, using the analysis chosen in chapter 4. Conclusions arising from these findings are discussed in the following chapter 7. Implications are found for both academics and practitioners both for brand extension theory and empirical generalizations.
Chapter 7 Analysis of Findings

1. The Indirect role of Quality

Original brand quality (Q) is shown not to have a significant direct effect on brand extension preference in the generalized pooled models, although it does contribute positively via interactions with the three fit attributes, namely complementarity, substitutability and transferability. This means that the effect of the Original Brand Quality on extension preference is completely mediated by the main effects of complementarity, substitutability and transferability. The generalized inferences generated in the GLMM framework about the role of Original Brand Quality (Q) is consistently in agreement with both the underlying theory and the empirical generalisation published by the seminal (Aaker and Keller 1990) paper both at individual study and pooled studies levels.

Original brand quality does appear to have some direct effect on brand preference in two of the thirteen locations sampled in the GLMMs results, notably Erasmus and ESSEC, but this inference is localized to these studies and is likely produced by high type II error rates associated with inadequate sample size compared to SEM recommendations given the known heteroskedasticity problems with these experimental data.

2. The Direct and Indirect roles of the three Fit attributes

Both extrinsic effects of complementarity (C) and substitutability (S) have significant positive direct effects on brand extension preference. All three significant fit effects, complementarity (C), substitutability (S) and transferability (T) all mediate the effect of original brand quality (Q) on extension preference.

3. The lack of any generalized role for the Difficulty attribute

No significant effect, either direct or indirect, was found for the difficulty attribute in the generalized pooled GLMM analysis framework. Only two of the twelve sample locations that included difficulty, China and Illinois, indicated a localized significance for this effect. Again, this inference is localized to these studies and is likely produced by high type II error rates.
associated with inadequate sample size compared to SEM recommendations given the known heteroskedasticity problems with these experimental data.

4. The problem with residual centering is further demonstrated

As demonstrated in a previous study (Echambadi et al. 2004), Tables 30 and 35 above show that the residual centering technique in the context of model specifications with interactions introduces considerable positive bias to the significance inferences for the direct effects, regardless of their true significance in the pooled GLMM, and fails to improve the agreement of individual study location inferences with the generalized inferences from the pooled model.

5. The advantages of mixed models and intra-fit interactions are further demonstrated

Tables 0 through 39 above show that with or without difficulty, the general linear mixed model analysis framework with intra-fit interactions specified is superior to the OLS framework, or a mixed model specification without intra-fit interactions, for this type of model data, where significant heteroskedasticity exists amongst predictor and dependent variables. The improvements in inference are evenly distributed over both direct and indirect effects types and over 58% and 62% of the 12 and 13 study locations respectively. Indeed, given the relative true size of the effects and the observed data variance, there is a good chance that some or all of the OLS-based positive inferences that were not supported by the mixed model at the study location level were false positives for that location, generated by the problems with OLS spuriously inflating the significance of weak instruments (Zivot et al. 1998).

6. Ideal point or inverted “U” model for extrinsic effects was found

The M-2LL statistics for the all eight best models in Table 40 above shows that the ideal point fit effect specifications are significantly better than any of the eight best models without the inclusion of second-order ideal point fit effect. Furthermore, the odds that the best two models
in table 40 are better than all other models in tables 40 and all models shown in table 27 above are greater than 0.99 using the permutation approach to GLMM selection (Esler 1998). Table 40 above shows that both the intrinsic effect substitutability and the extrinsic effect transferability, have significant positive linear, or first order parameter estimates and significant negative second order parameter estimates. This is then potentially significant inverted “U” or ideal point behaviour, where there exists some moderate level of both that yields a global maximum contribution to extension evaluation.

By setting the first partial derivative of the estimated function for extension evaluation as a function of substitutability to 0 and solving for substitutability, we estimate the optimal levels as a function of substitutability’s significant interactions:

\[ Y \equiv \left( +0.297 - 0.0297 \times C + 0.0301 \times Q \right) \times S - 0.01354 \times S^2 + \kappa \]

For the purposes of the partial derivative to follow all other terms not involving S at any level of measurement are grouped in here with the constant, \( \kappa \). Differentiating the evaluative response \( Y \) with respect to substitutability, \( S \), and setting this to 0 at the maximum of \( Y \) we get:

\[ \frac{\partial Y}{\partial S} \equiv 0 = 0.297 - 0.02055 \times C + 0.031 \times Q - 2 \times 0.01354 \times S \]

at the optimal level of \( S \).

Moving the first two RHS terms over to the LHS and dividing by \( S \)'s coefficient, we get:

\[ \Rightarrow S_{opt} \equiv -\frac{\left(0.297 - 0.02055 \times C + 0.0301 \times Q\right)}{-\left(2 \times 0.01354\right)} \]

\[ \Rightarrow S_{opt} \equiv 10.97 - 0.76 \times C + 1.11 \times Q \]

The difference in extension response evaluation between high and low levels of complementarity and original brand quality are not influential, as the implied optimal point
for substitutability, in the range 6.76 to 17.98 is still approximately equal to, or higher than, the highest realizable scale measure of 7. Thus we can further clarify the nature of the extrinsic brand effect substitutability as not effectively exhibiting an ideal point relationship, but rather as a monotonically increasing but still curvilinear relationship over the realizable measurement range. The practical interpretation is then very similar to a linear effect, in the sense that simply more is increasingly better over the realizable scale. The only difference is that the maximum rate of increase in extension evaluation occurs at low levels of substitutability, and the marginal impact of additional substitutability on extension evaluation is least at the high end of the scale, where the response curve slope is least. It is unlikely, however, that this subtle difference is of any practical or theoretical use to practitioners or academics alike.

A similar analysis approach for transferability yields the following:

\[ Y \equiv (+0.34 + 0.008167 \times C) \times T - 0.04314 \times S^2 + \kappa \]

\[ \frac{\partial Y}{\partial T} = 0 \equiv +0.34 + 0.008167 \times C - 2 \times 0.04314 \times T \quad \text{at the optimal level of } T \]

\[ \Rightarrow T_{opt} = -\frac{(+0.34 + 0.008167 \times C)}{-2 \times 0.04314} \]

\[ \Rightarrow T_{opt} = +3.9 + 0.095 \times C \]

Unlike substitutability, this finding for the extrinsic transferability effect is exactly the type of ideal point evaluation response effect that was posited for affective intrinsic fit effects (Bijmolt et al. 1998; Mandler 1982 p 22; Meyers-Levy et al. 1994). The optimal range of transferability for the maximum extension evaluation is between 4.0 and 4.6 on the original 7-point scale, depending on the level of complementarity (between 1 and 7 respectively).
7. Optimal extension attribute mixes: an exploration

Table 43 optimised extension evaluation strategies using the best ‘ideal point’ model

<table>
<thead>
<tr>
<th>attribute</th>
<th>coefficient</th>
<th>attribute level</th>
<th>partworth</th>
<th>attribute level</th>
<th>partworth</th>
<th>attribute level</th>
<th>partworth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.6549</td>
<td>1</td>
<td>0.65</td>
<td>1</td>
<td>0.65</td>
<td>1</td>
<td>0.65</td>
</tr>
<tr>
<td>Q</td>
<td>0.04663</td>
<td>7</td>
<td>0.00</td>
<td>7</td>
<td>0.00</td>
<td>7</td>
<td>0.00</td>
</tr>
<tr>
<td>C</td>
<td>0.164</td>
<td>1</td>
<td>0.16</td>
<td>7</td>
<td>1.15</td>
<td>7</td>
<td>1.15</td>
</tr>
<tr>
<td>S</td>
<td>0.297</td>
<td>7</td>
<td>2.08</td>
<td>1</td>
<td>0.30</td>
<td>7</td>
<td>2.08</td>
</tr>
<tr>
<td>T</td>
<td>0.34</td>
<td>4.04</td>
<td>1.37</td>
<td>4.60</td>
<td>1.56</td>
<td>4.60</td>
<td>1.57</td>
</tr>
<tr>
<td>CxS</td>
<td>-0.02055</td>
<td>7</td>
<td>-0.14</td>
<td>7</td>
<td>-0.14</td>
<td>49</td>
<td>-1.01</td>
</tr>
<tr>
<td>CxT</td>
<td>0.008167</td>
<td>4.04</td>
<td>0.03</td>
<td>32.20</td>
<td>0.26</td>
<td>32.22</td>
<td>0.26</td>
</tr>
<tr>
<td>SxS</td>
<td>-0.01354</td>
<td>49</td>
<td>-0.66</td>
<td>1</td>
<td>-0.01</td>
<td>49</td>
<td>-0.66</td>
</tr>
<tr>
<td>TxT</td>
<td>-0.04314</td>
<td>16.28</td>
<td>-0.70</td>
<td>21.16</td>
<td>-0.91</td>
<td>21.19</td>
<td>-0.91</td>
</tr>
<tr>
<td>QxC</td>
<td>0.02375</td>
<td>7</td>
<td>0.17</td>
<td>49</td>
<td>1.16</td>
<td>49</td>
<td>1.16</td>
</tr>
<tr>
<td>QxS</td>
<td>0.01308</td>
<td>7</td>
<td>0.09</td>
<td>7</td>
<td>0.09</td>
<td>49</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>max S</td>
<td>3.1</td>
<td>max C</td>
<td>4.1</td>
<td>max C, S</td>
<td>4.9</td>
<td></td>
</tr>
</tbody>
</table>

This table above shows the average extension evaluation of the 3 possible optimum extension brand design and communication management strategies. These have been calculated using only significant parameter coefficient estimates from the best model in table 41 and optimized using Excel solver, restricting attribute levels to realizable ranges, and fixing some to differentiate the three possible optimal strategies.

The “Max C” maximum evaluation of an extension resulting from an extension design strategy of high levels (7) of quality and complementarity, and optimal level of transferability of 4.6 yields a predicted extension valuation of 4.1 on a 7 point scale.

This compares favourably to the maximum evaluation of an extension resulting from a “Max S” extension design strategy of high levels of quality and substitutability, and optimal level of transferability of 4.04 yields a predicted extension valuation of 3.1 on a 7 point scale. Furthermore, extensions with high substitutability ratings will also logically suffer from cannibalization, and therefore it is recommended discrete choice models (McFadden 1972 pp 3-11; McFadden and Train 2000; Walker and Ben-Akiva 2002) are included in the evaluation using this type of extension design strategy.
Since the negative part-worths arising from the intra-fit interaction coefficients are always smaller than all other part-worths, the optimal extension evaluations from a “Max C,S” high complementarity and substitutability and an optimum level of transferability of 4.6 are estimated to yield an estimation of 4.9. However, not only does the same caveat recommending choice models to further evaluate original brand cannibalization as for the high substitutability extension design strategy apply to this strategy, but also there is the additional creative difficulty (but admittedly not the impossibility) of conceptualizing an extension that simultaneously rates highly on both complementarity and substitutability.

8. The relationship between power and sample size

As discussed earlier, there are no closed form solutions to the problem of determining the type I or II errors for mixed models, and simulation is recommended (Castelloe and O’Brien 2001). Since the mixed model is the only appropriate and practical model for this data, other calculations are irrelevant. In other words, for adequate type I and II error control for the given data, variance-covariance structure and true effect sizes, a minimum of 10 location sample frames, 1180 subjects and 12885 observations were necessary. Whilst this might initially seem unusually large, it is in agreement within an order of magnitude to the recommended sample size associated with the closest alternative estimation framework mentioned earlier: the random coefficient non-standard structural equation model. For structural equation models, five times the number of parameters to be estimated is a common guideline, and there are 210 covariances, 15 variances, 15 fixed effect parameters, and for the random coefficient part, as many additional parameters as there are extracted principle components of locations and subjects, unknown but say, conservatively, 4, multiplied by the 15 exogenous manifest variables. 300 times 5 or about 1500 experimental unit observations, or 1500 subjects, are therefore indicated. With the type of ecologically valid covering experimental design used here, there are at least 16 repeated observations per subject, about 24,000 stimulus responses is indicated. This type of sample size requirement calculation assumes that all fixed and random parameter estimates are significant, which is rarely the case, and so possibly explains why the empirical limit to sample size based on the other type I and II error investigations reported above is somewhat lower.
Taken together the results of the random sample size reduction and the conservatively bias bootstrap estimates of type II error rates does not imply that the significances of the Q x S and S x S effect coefficients are to be questioned, rather caution needs to be taken when estimating interaction and higher-order effects from semi-experimentally designed or observational data sources to ensure sufficient additional sample size is practical and available when planning to test these hypotheses. Additionally, the S x S second order ideal point effect of response to complementarity is such that it makes little practical difference to the overall interpretation and implications for the linear response alternative model as the maximum evaluation response occurs at the high end of the evaluation scale anyway. Rather than an ideal point response, substitutability is merely exhibiting a saturation response effect at the high end of its evaluative scale.

The above findings are summarised with specific reference to the original 16 propositions in the table below.
## Table 44 Confirmation or disconfirmation of propositions

<table>
<thead>
<tr>
<th>Proposition</th>
<th>Confirmed Y/N</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1: Original brand quality plays a direct role in positive brand extension evaluation</td>
<td>Y</td>
</tr>
<tr>
<td>P2: Original brand quality impacts brand extension evaluation by an indirect mediation via the fit attributes</td>
<td>Y</td>
</tr>
<tr>
<td>P3: Extensions seen as trivial suffer poor evaluation</td>
<td>N</td>
</tr>
<tr>
<td>P4: Extension triviality has no impact on extension evaluation</td>
<td>Y</td>
</tr>
<tr>
<td>P5: An appropriate modeling framework will substantiate an empirical generalization of Aaker and Keller’s theory of brand extensions based on all of the data associated with the existing set of close replication studies</td>
<td>Y</td>
</tr>
<tr>
<td>P6: An appropriate modeling framework will fail to substantiate an empirical generalization of Aaker and Keller’s theory of brand extensions based on all of the data associated with the existing set of close replication studies</td>
<td>N</td>
</tr>
<tr>
<td>P7: Residual centering in models with interaction effects leads to bias in the generated inferences</td>
<td>Y</td>
</tr>
<tr>
<td>P8: Residual centering in models with interaction effects does not lead to bias in the generated inferences</td>
<td>N</td>
</tr>
<tr>
<td>P9: The close replication studies had sufficient sample size to support robust inferences</td>
<td>N</td>
</tr>
<tr>
<td>P10: The close replication studies did not have sufficient sample size to support robust inferences</td>
<td>Y</td>
</tr>
<tr>
<td>P11: There is an ‘ideal point’ relationship between complementarity and extension evaluation</td>
<td>N</td>
</tr>
<tr>
<td>P12: There is no ‘ideal point’ relationship between complementarity and extension evaluation</td>
<td>Y</td>
</tr>
<tr>
<td>P13: There is an ‘ideal point’ relationship between substitutability and extension evaluation</td>
<td>N</td>
</tr>
<tr>
<td>P14: There is no ‘ideal point’ relationship between substitutability and extension evaluation</td>
<td>Y</td>
</tr>
<tr>
<td>P15: There is an ‘ideal point’ relationship between transferability and extension evaluation</td>
<td>Y</td>
</tr>
<tr>
<td>P16: There is no ‘ideal point’ relationship between transferability and extension evaluation</td>
<td>N</td>
</tr>
</tbody>
</table>
Chapter 8 Conclusions, implications and limitations

1. Conclusions

The primary conclusions from this study are those generalized conclusions that can be drawn about brand extension preference formation.

The GLMM framework with intrafit interactions specified improves inferences by study location in a more consistent way amongst all effects compared to residual centering with respect to the GLMM pooled result. The pattern of improvement in agreement for residual centering is biased in that it improves agreement mainly with respect to the main effects of complementarity (C), substitutability (S) and transferability (T). Both intra-fit interaction specification with OLS regression and the GLMM framework without intra-fit interactions perform only weakly better than the simplest OLS framework.

Support for the effect of original brand quality was found in the comprehensive, pooled data set only via moderation, or interaction, with the extrinsic effects of complementarity and substitutability. Since neither extrinsic effect is practically limited by any ideal point behavior over its measurable range, then brand managers can usefully explore the leverage of existing high quality original brands via extensions certainly with high levels of complementarity, and possibly with high levels of substitutability, but with some qualification regarding the tradeoff between the market-specific financial impacts of substitutability and original brand cannibalisation.

The direct effect of the second posited extrinsic attribute difficulty is not supported. No moderating role of quality on either transferability or difficulty is supported. No direct effect of original brand quality on extension preference is supported.

The three main-effect brand extension ‘fit’ attributes, namely complementarity, substitutability and transferability and the indirect effect of original brand quality via quality’s moderating role on the intrinsic effect of complementarity and are supported.
Direct fit effects of the extrinsic effect transferability demonstrates ‘ideal point’ behaviour over the practical range of measures, limiting the direct evaluative influence to moderate levels when compared to the effectively linearly increasing intrinsic effects of complementarity and substitutability.

The GLMM framework in conjunction with intra-fit effects specification improves brand extension model inferences by study location in a more consistent way amongst all effects compared to residual centering or models without intra-fit effects with respect to the GLMM pooled result with intra-fit interactions. Only the GLMM framework, together with intra-fit interaction specifications and the important ideal point specifications strictly allows generalisability with respect to the populations that subjects and brand extensions were sampled from. Thus the GLMM inference framework is seen as an essential set of analysis tools when testing important hypotheses on data with high levels of heteroskedasticity.

Aaker and Keller’s seminal brand extension theory is further supported both theoretically and empirically and is strictly generalized to all study locations by including study location as a random effect in the General Linear Mixed Model.

Additionally, the limited usefulness of the ideal point or inverted “U” response to the extrinsic transferability effect, which is linked in the literature to broad brands, formed by sequential successful extensions, and typically supported by advertising directed at maximally enhancing positive affective attitudes to the brand, are in stark contrast with the opportunities afforded by the monotonically increasing response to the significant intrinsic effects of complementarity and substitutability.

We also draw several useful conclusions from this study by employing hindsight from the beneficial position of finally having produced unbiased inferences from the pooled data using an appropriate inference generation framework and model specification. From this position we can fully explain the lack of agreement amongst the original study and the close replications. Difficulties in previously published attempts at Empirical Generalisations of Aaker and Keller’s seminal Brand Extension preference theory are fully explained due to:
(1) the inappropriate use of the residual centering technique (Echambadi et al. 2004),
(2) the positive bias of Ordinary Least-Squares T-statistics for weak instruments (Zivot et al. 1998),
(3) a failure to account for within-fit interactions and the second-order ideal-point fit specifications in the empirical models, leading to severe specification error,
(4) poor experimental design efficiencies associated with a lack of balance and the presence of significant heteroskedasticity, coupled with
(5) lack of sufficient type I and type II error control due to insufficient sample size over all experimental units, both locations and subjects.
(6) the subsequent lack of application of Generalized Linear Mixed or Hierarchical Bayesian models and Maximum Likelihood information criteria to both single sets and pooled analysis of replicated study data to adapt to conditions (2) and (4), and finally
(7) some basic procedural errors in the original studies (Holden and Barwise 1996).

Note the citations above in points 1-7 refer to papers in which the problem is noted, not evidenced.

2. Implications

It is envisaged that these results will have wide ranging academic and commercial applications.

Implications for Brand Extension Research

Academics interested in the literature of brand extension will have the causes of the prior disagreements about the roles and inter-relationships amongst the three ‘fit’ attributes of the Aaker and Keller model more fully explained. These disagreements are fully described in chapter 2 section 5 and chapter 3 of this thesis. Academics in the brand extension area will be able to focus on incremental improvements to an accepted set of core components in the theory of brand extension preference formation, particularly the direct and indirect effects of fit and the indirect effect of original brand quality completely moderated by the fit attributes.
Additionally, the importance of integrating the two literature streams of brand extension theory and ideal point evaluation models, and therefore including the inverted “U” relationship in models of the extrinsic transferability brand extension fit attributes, introduced in chapter 2 section 5 and chapter 3, is highlighted in this thesis.

Given the importance to marketing practitioners, especially those with brand management responsibilities, this ideal point model form should be included in all future brand extension evaluation research, further explored and generalised, and the implications then included in future marketing texts.

Authors associated with previous studies that have employed the dimensions of complementarity, substitutability and transferability in brand extension research will benefit directly from studying these findings. This should stimulate future research that seeks to modify previous experimental designs and promote critical evaluation and further understanding of brand extension preference formation by eliminating potential specification bias embedded in many previous studies (Aaker 1990; Albion 1985; Barrat et al. 1999; Bottomley and Doyle 1996; Bottomley and Holden 2001; Duncan and Nelson 1986; Elrod, Russell, Shocker, Andrews, Bacon, Bayus, Carroll, Johnson, Kamakura, Lenk, Mazanec, Rao, and Shankar 2002; Flaherty and Pappas 2000; Kapferer 1992 p 96; Keller 1998 pp 608 - 12; Kerby 1967; Martin, Stewart, and Matta 2005; Mather 2005; Mather and Sunde 1998; Milewicz and Herbig 1994; Monga and John 2004; Rangaswamy et al. 1993b; Reast 2005; Romaniuk and Sharp 2004; Samu and Ducey 2002; Sharp 1991; Smith and Andrews 1995; Sunde and Brodie 1993; Sunde 1991, 1998; van Riel, Lemmink, and Ouwersloot 2001; Völckner 2003; Wänke, Bless, and Schwarz 1998; Zatloukal 2002).

Furthermore, any researchers in (a) marketing of brand extensions, (b) brand equity or (c) cognitive psychology that have employed either the dimension of congruence (Bacon 1979; Maoz and Tybout 2002; Meyers-Levy and Tybout 1989) or typicality (Boush and Loken 1991; Farquhar, Herr, and Fazio 1990; Loken and Ward 1987; Park, Lawson, and Milberg 1989; Rosch et al. 1976) would similarly benefit from these thesis findings.

In addition to considering potential specification bias that may be embedded in previous studies, the particular relevance of the ideal point behaviour of the extrinsic component of overall congruence, as evidenced by the finding for the role of transferability in this study,
and its implication for research design will be of interest to the researchers cited above when planning future research along these themes.

The particular relevance of the ideal point behaviour of the extrinsic component of overall congruence and its implication for research design will be of interest to the researchers cited above when planning future research along these themes. This relevance is evidenced by the finding for the ideal point behaviour of the extrinsic effect of transferability in this thesis, suggesting fruitful roles for inducing cognitive elaboration as a guiding mechanism for successful brand extension strategies. Indeed several recent publications seem to be heading in this direction already (Ahluwalia 2008; Batra et al. 2010; Monga and John 2010).

A number of guidelines have now been given for practitioners when developing brand extensions, including several optimal tactics depending on the contextual branding strategy, as well as the importance of designing extensions that enhance cognitive evaluative processing so as to take advantage of the significant ‘ideal point’ extrinsic evaluative consumer response.

**Implications for Other Brand Research**

**Implications for Empirical Generalisations in Marketing Research**

The skill set of researchers publishing replication studies and substantiating general theories in marketing will be generally enhanced by the inclusion of general linear mixed model frameworks with appropriate interaction effects and variance-covariance structure specification. The applicability of parsimonious General Linear Mixed Model estimation frameworks to empirical generalisations of marketing theory on multiple sets of similar, close replication, data has been demonstrated.

The warning that residual centering should only be used with care and in situations where main effects only are an appropriate model effects specification should be heeded by all.
Experimental unit sample size considerations should be more carefully reviewed in the light of possible hierarchical effects and significant variance-covariance structure. Simple OLS-based sample size calculations may not apply, and the implications for increased sample size can be considerable. Sample size should be guided by recommendations for the nearest equivalent SEM specification, which already accommodates the sample size implications for estimates of additional heteroskedastic effects.

The significance of the ideal point or inverted “U” behavior of the extrinsic transferability fit parameter estimates in this comprehensive study also provides valuable empirically generalised support for the findings of the original authors’ research on this topic in a branding context (Ozanne et al. 1992). Furthermore, this finding may now offer a credible process-based hypothesis to explain another disturbing empirical generalisation of lower brand profits from higher communication investment (Albion and Farris 1987), and so should stimulate more research in this direction.

**Implications for Other Brand Management research**

In addition, researchers of brand equity and its management (Ambler and Styles 1996; Andrews 1996; Balachander and Ghose 2003; Barwise 1993; Broniarczyk and Alba 1994; Chen and Chen 2000; Dawar and Anderson 1994; DelVecchio and Smith 2005; Keller 1993, 2002 pp 27-41; Lane and Jacobson 1995; Lane 1998; Lebar, Buehler, Keller, Sawicka, Aksehirli, and Richey 2005; Martin et al. 2005; McMath 1997; Mills 1995; Randall, Ulrich, and Reibstein 1998; Schuiling and Moss 2004; Smith 1990; Sullivan 1992; Swait, Erdem, Louviere, and Dubelaar 1993; Urde 1994; van Riel et al. 2001; Yoo, Donthu, and Lee 2000) will benefit from the additional insights that the findings of this study provides. The new realization of the limited role that the extrinsic attribute of transferability can play in leveraging brand equity, compared to the more flexible and extensive brand equity leverage value that the intrinsic attributes of complementarity and substitutability may stimulate a revised set of brand equity management guidelines from some of these researchers cited above who are actively publishing in this theme area.
Implications for Marketing Practitioners

Marketing practitioners have a more generally accepted and more easily understood core theory of brand extension to reliably guide their successful brand portfolio development. Practitioners cannot assume without further market specific research that existing brand value will be transferred to any specified brand extension. Concept fit between the original brand and the extension must be examined, understood and maximised, so that the likelihood of brand extension success can in turn be maximized following the recommendations below.

To this author’s best knowledge, these findings also offer a unique and robust explanatory model that questions the widely employed broad brand management strategy surrounding the use of advertising targeted to maximizing affective response to the brand. These results demonstrate quantitatively that brand equity is most simply leveraged to an extension via a strategy of brand extension design and communication that maximises positive effective evaluation of the extension brand’s complementarity. Potential cannibalisation of the market share of a parent brand is likely to be also minimized with this strategy.

However, no one clear globally optimal strategy emerges. Rather it becomes apparent that (a) intrinsic brand support should be limited to that which optimizes perceived original brand transferability to future extensions at moderate levels, whilst product design, distribution, customer service and communication efforts all support maximum original brand quality, and (b) future brand extension research should integrate choice models so as to enable an optimized trade off amongst potential extensions exhibiting positive evaluations to optimum levels of high complementarity and moderate transferability with relatively little parent brand cannibalisation, and other potential extensions with additional high levels of substitutability and associated higher levels of parent brand cannibalization.

Even in situations where brand managers consider employing original brand cannibalization, additional high extension complementarity (where realizable, possibly a difficult creative ideation task in itself) to high substitutability will further enhance the desired extension preference. Choice models (Ben-Akiva, Bradley, Morikawa, Benjamin, Novak, Oppewal, and Rao 1994; Louviere, Hensher, and Swait 2000; McFadden 1972 pp 3-11; McFadden and Train 2000) will likely still be included as part of the applied research mix as predictions of parent brand cannibalization will still be required in this strategic context.
3. Limitations

One limitation of this research is that the response variable in all the contributing studies’ data is an interval measure of brand extension quality and therefore is a measure of an attitude or judgment. Caution must be exercised when extending this newly generalized model of brand extension evaluation through to consumer behavior. There is a wealth of literature on the discordance between preference and behavior (Agarwal and Teas 2001; Belk 1985; Kahneman et al. 1982 pp 101-53; 211-68).

Another limitation is that the collection and integrity of the data used for this analysis was not under the direct control of this author. Additional analysis does not indicate any data corruption but there is always a possibility that this has occurred.

Although no literature was found to support 3-rd order and higher effect specifications, further specification bias cannot be ruled out for higher order effects of mediation and ideal point relationships already tested in this second order effect specification model used in this thesis. The post-hoc results of the relationship between sample size and power in section 8 of chapter 7 indicate that even with the relatively large number of observations arising from the pooling of 13 data sets, there is insufficient data to support the reliable analysis of additional effects implied by 3rd and higher order general linear mixed models necessary to investigate the existence of 3-way mediation and 3rd order polynomial curvilinear effect responses.

Finally, all data supporting inferences in this study were collected from consumers who were not necessarily buyers or purchasers of the original brand. However in every case the original brand was well-known to the market the sample was drawn from, and these consumers, even if not the final decision maker, likely had knowledge, involvement and influence on relevant decision makers and vice-versa, making this a minor limitation overall.

Although the XLRI data showed a low level of design balance, and was missing some original brand design points, the inclusion of it in the pool is not a limitation as the inference framework used was demonstrated to be more than adequate to the job of accommodating the
challenges and generating robust inferences despite the data limitations via the results of the bootstrap and associated power and sample size analysis.
References


Brucks, Merrie D. and Samar Das (1999), "A Structural Model of Brand Equity," in *American Marketing Association Advanced Research Techniques Forum*, Santa Fe, New Mexico, USA.


Glenn Gray


--- (2006), "The 100 Top Brands. (Cover Story)," *BusinessWeek* (3996), 60.


--- (2009), "100 Best Global Brands. (Cover Story)," *BusinessWeek* (4148), 50-56.


New York: Cambridge University Press.


Labat, René ed. (c.1600 BC), *Treatise of Medical Diagnosis and Prognoses (English Translation by Labat 1951)*, Nineveh, Mesopotamia.


Loken, Barbara and Deborah Roedder John (1993), "Diluting Brand Beliefs: When Do Brand Extensions Have a Negative Impact?,” *Journal of Marketing*, 57 (3), 71.


Shao, Jun and Dongsheng Tu (1995), The Jackknife and Bootstrap, New York, NY, USA Springer Verlag.


