

THE UNIVERSITY OF LONDON

**Patterns of Audience Appreciation Ratings
for Television Programmes**

Douglas George CARRIE

A Dissertation in Marketing

Thesis Submitted to the Faculty of Economics
in Candidacy for the Degree of Doctor of Philosophy

LONDON BUSINESS SCHOOL

1997

ABSTRACT**PATTERNS OF AUDIENCE APPRECIATION RATINGS
FOR TELEVISION PROGRAMMES****Douglas George Carrie**

(Under the supervision of Professor A.S.C. Ehrenberg)

This thesis is about patterns in the way viewers rate their appreciation (i.e. their *liking or enjoyment*) of television programmes. It is based on the analysis of an extensive set of previously confidential UK panel data, and involves replicating and building upon a well-established body of prior work. The major conclusions reached are:

- Individual viewers can be segmented (on both behavioural and attitudinal bases) using this type of audience appreciation data.
- Differences in the audience composition of programmes (i.e. certain programmes might attract audience segments that tend to give higher average appreciation scores) do not provide a simple explanation for systematic variations in the aggregate appreciation scores achieved by different programmes and by different programme types or genres.
- Limited evidence is found in support of a 'narrowcasting' proposition (i.e. where certain programmes might be expected to reach small, well-defined, *and especially involved or appreciative audiences*). This is not however the strong pattern that many practitioners might expect. There is no evidence, even in extreme cases, of any programmes that are watched by very targeted and *highly* appreciative audiences.

-
- There are systematic relationships between audience appreciation ratings and more traditional television ratings of audience size. When allowances are made for scheduling factors, programme type effects, and variations in audience composition, the size of a programme's audience is positively related to the appreciation viewers have for it.
 - There is a strong relationship between appreciation and frequency of viewing. Programmes which achieve higher average appreciation scores also have higher levels of repeat viewing.

Overall, these patterns suggest a positive but limited future role for audience appreciation ratings as a supplementary media planning measure. The data presentation methods and findings outlined in this thesis should provide a useful guide to help aid the interpretation of results from subsequent studies.

To my parents, Jim and Dorothy Carrie, and to the proud memory of my grandparents, Douglas Carrie, Jessie Carrie, George Armstrong, and Edna Armstrong.

ACKNOWLEDGEMENTS

Professor Andrew Ehrenberg's influence on the work described here is immeasurable. He has offered his advice and encouragement generously, but his real gift to me is in the general perspective and outlook he has given me on research and on the effective communication of numerical, verbal, and written information. This will forever guide me in the approach I will take when faced with new teaching and research challenges.

I owe an enormous personal debt to my parents, Jim and Dorothy Carrie, for their endless love, encouragement, and financial support. This must have seemed like a crazy endeavour at times but I thank them for always giving me the freedom to follow my dreams. I am also grateful to the memory of my grandparents, Douglas Carrie, Jessie Carrie, George Armstrong, and Edna Armstrong, for the formative and never-ending influence they have had on my life. Finally, thanks to my wonderful sister Susan, to my brother-in-law Greg, and to my new nephew Jeffrey, for all your love and support.

For helping to make my life as a PhD student at London Business School so rich and memorable, this acknowledgement would be incomplete without a mention of the entire A-Team: Mark Bleackley, Sabine Küster, Luc Renneboog, Tomas Valnek, Rose Trevelyan, James and Kathy Taylor, Zoltan and Judit Antal, Oliver Hansch and Maureen Gillespie, Liz and J.R. Tracy, David Cannon and Bette Laing, and Mark, Karen and William Shackleton.

For their moral support at the final write-up stage, I would also like to thank all my new colleagues at the University of Auckland (with special thanks to Andi Martin, for her help in proof reading, and to Bill Murphy and his wife Toni, for keeping me company down the hall over many long evenings). Finally, for helping me maintain my sanity to the very end, I should thank “the lads” in Auckland; Matt, Mario, Frank, Goran, Wayne, Klem, Osh, and Whoppa.

TABLE OF CONTENTS

1.0	Introduction	
1.1	Chapter One Overview	13
1.2	Framing this Research	13
1.3	Background to the Choice of Topic	16
1.4	Thesis Structure	21
1.5	Summary of Findings	23
	(a) Individual Viewers' Appreciation Scoring Patterns	24
	(b) Programme Appreciation Scores	26
	(c) Programme Appreciation and Audience Composition	27
	(d) Programme Appreciation and Audience Size	30
	(e) Programme Appreciation and Repeat Viewing	31
1.6	Discussion and Conclusion	32
2.0	An Introduction to the Literature	33
2.1	Chapter Two Overview	33
2.2	The Wide Ranging Nature of Audience Research	34
2.3	Patterns of Viewing Behaviour	36
	(a) Audience Size	39
	(b) Audience Composition	39
	(c) Repeat Viewing	40
	(d) Duplication of Viewing	41
	(e) Double Jeopardy	42
2.4	Patterns of Television Audience Appreciation	42
	(a) British Audience Appreciation Research	43
	(b) Other Audience Appreciation Research	44
2.5	Summary	45
3.0	Methodology	49
3.1	Chapter Three Overview	49
3.2	Methodological Principles	49
3.3	A Question of Significance	52
3.4	The BARB Television Opinion Panel	57
	(a) The Appreciation Rating Scale	58
	(b) How Broadcasting Organisations Have Used this Data	62
3.5	The Raw Data Sample	64
3.6	Additional Variables Developed for this Research	67
	(a) Behavioural Classifications of Individuals	67
	(b) Classification of Programme Types	67
	(c) Calculation of Programme Appreciation Scores	69
	(d) Calculation of Audience Size Ratings	71

3.7	Sampling Frames for this Research	71
3.8	Summary and Sequence of Analyses	74
4.0	Individual Viewers' Appreciation Scoring Patterns	78
4.1	Chapter Four Overview	78
4.2	Viewers Like What They Watch	78
4.3	Demographic Patterns	79
4.4	Behavioural Classifications of Viewers	81
	(a) Light, Average and Heavy Viewers of Television	82
	(b) Low, Moderate, and High Appreciation Scorers	85
	(c) Consistent, Regular, and Varied Appreciation Scorers	87
	(d) Composition of the Different Behavioural Categories	89
4.5	A CHAID Segmentation Modelling Analysis	96
	(a) What is CHAID?	97
	(b) Results of a CHAID Analysis	99
	(c) Comparing CHAID to the Data Reduction Approach	104
	(d) Comparing CHAID to other Segmentation Approaches	109
4.6	Patterns Across Programme Types and Channels	112
	(a) Individuals' Viewing of Programme Types and Channels	112
	(b) Individuals' Appreciation of Programme Types and Channels	117
4.7	Summary	126
5.0	Programme Appreciation Scores	128
5.1	Chapter Five Overview	128
5.2	Programmes are Mostly Quite Liked	128
5.3	Appreciation for Different Programme Types	129
5.4	Programme Appreciation by Channel	132
5.5	Summary	134
6.0	Programme Appreciation and Audience Composition	135
6.1	Chapter Six Overview	135
6.2	The Audience Composition of Programmes	136
6.3	The Audience Composition of Programme Types	139
6.4	The Audience Composition of Different Channels	141
6.5	Programme Appreciation and Audience Composition	144
6.6	Programme Appreciation and 'Narrowcasting'	154
6.7	Programmes Appreciation and Audience Bias	156
	(a) Framing the Research Question	157
	(b) Detailed Analysis for Programme Audience Bias by Gender	161
	(c) Results from Similar Analyses for other Segmentation Criteria	170
	(d) Summary	172

6.8	Programmes Most Liked by Different Segments	173
6.9	Summary	175
7.0	Programme Appreciation and Audience Size	177
7.1	Chapter Seven Overview	177
7.2	Main Findings of Past Research	177
7.3	Methodology	179
	(a) Definition of Variables	179
	(b) The Analysis Approach Taken	180
	(c) A Methodological Note on Best Fit	183
7.4	Analysis and Results	186
	(a) The Relationship Allowing for Scheduling & Channel Effects	187
	(b) The Relationship within Programme Types	190
	(c) The Relationship Allowing for Audience Differences	198
7.5	Summary	200
8.0	Audience Appreciation and Repeat Viewing	202
8.1	Chapter Eight Overview	202
8.2	Main Findings of Past Research	202
8.3	Appreciation and Frequency of Viewing	205
8.4	Appreciation and Repeat Viewing	209
8.5	Summary	210
9.0	Discussion and Implications	211
9.1	Chapter Nine Overview	211
9.2	Summary of Overall Findings	211
9.3	Limitations	215
9.4	Directions for Future Research	218
9.5	Conclusion and Contribution	221
	Appendix: "Programme Involvement and Context Effects"	223
A.1	Overview	223
A.2	The Literature on Programme Context Effects	223
	(a) Negative Effects	225
	(b) Positive Effects	228
	(c) Positive and Negative Effects - An Inverted-U Relationship?	232
	(d) Recent Advertising Industry Research	232
A.3	Audience Appreciation and Media Planning	235
	References	240

TABLES

3.1	Extent of the Thesis Database	66
3.2	Classification of Programme Types	68
3.3	Breakdown of Individuals in the Panel According to Number of Diary Booklets Returned During this Five Week Period	72
3.4	Demographic Breakdown of Final Thesis Database	73
4.1	The Range of Appreciation Scores	79
4.2	Demographic Variations in Appreciation	80
4.3	Light, Average and Heavy Viewers of Television	83
4.4	Low, Moderate and High Appreciation Scorers	86
4.5	Consistent, Regular and Varied Appreciation Scorers	89
4.6	Composition of Light / Average / Heavy Viewers	91
4.7	Composition of Low / Moderate / High Appreciation Scorers	94
4.8	Composition of Consistent / Regular / Varied Appreciation Scorers	95
4.9	CHAID Segment Profiles in Decreasing Order by Segment Members' Average Overall Appreciation Scores	102
4.10	Individuals' Viewing of Different Programme Types	113
4.11	Individuals' Viewing of Different Channels	116
4.12	Individuals' Average Appreciation of Different Programme Types	118
4.13	Observed and 'Predicted' Average Appreciation by Males of Different Programme Types	120
4.14	Difference between Observed and 'Predicted' Average Individual Appreciation Scores for Different Programme Types	122
4.15	Individuals' Average Appreciation of Different Channels	124
4.16	Difference between Observed and 'Predicted' Average Individual Appreciation Scores for Different Channels	125

5.1	Average Programme Appreciation Score for the Average Programme	129
5.2	Appreciation Scores by Programme Type	131
5.3	Appreciation for Programmes on Different Channels	132
5.4	Programme Types Broadcast on the Different Channels	134
6.1	The Audience Composition of Programmes	137
6.2	The Audience Composition of Programmes Within Two Starting Time Bands	138
6.3	Audience Profiles for the Average Programme of Different Programme Types	140
6.4	Audience Profiles for the Average Programme on Different Channels	142
6.5	Appreciation Patterns for the Average Programme of Different Programme Types	145
6.6	Difference between Observed and 'Predicted' Average Programme Appreciation Scores for Different Programme Types	148
6.7	Appreciation Patterns for the Average Programme on Different Channels	151
6.8	Difference between Observed and 'Predicted' Average Programme Appreciation Scores for Different Channels	152
6.9	Enjoyment Ratings: 16 to 34 year olds	157
6.10	Enjoyment Ratings: 16 to 34 year olds. With Table Rearranged by Audience Size	158
6.11	The 20 Programme Showings Most Biased Towards Men	162
6.12	Programme Appreciation and Those Programmes with Most Biased Audience Composition by Gender	165
6.13	Programme Appreciation and Programme Audience Bias by Gender	170
6.14	Programme Appreciation and Programmes "Most Liked" by Men	174
7.1	Appreciation and Audience Size: The Overall Relationship	187
7.2	Allowing for Time and Channel Effects	189

7.3	Earlier Research Allowing for Programme Type (1)	191
7.4	Earlier Research Allowing for Programme Type (2)	192
7.5	Allowing for General Programme Types: BBC1 & ITV	193
7.6	Allowing for Specific Programme Types: BBC1 & ITV	195
7.7	Residual Standard Deviations (rsd) for the Linear Regression Equations Outlined in Table 7.6	197
8.1	Appreciation and Individuals' Frequency of Viewing	206
8.2	Individuals' Frequency of Viewing	208
8.3	Programme Appreciation and Repeat Viewing	209

FIGURES

2.1	Patterns of Television Audience Viewing Behaviour	38
4.1	Tree Diagram for Viewer Segmentation by Appreciation Scoring Patterns	100
7.1	Audience Size and Programme Appreciation	186
7.2	Shape of the Viewing Curve Over the Day	188

1. INTRODUCTION

1.1 CHAPTER ONE OVERVIEW

This introductory chapter outlines the scope and structure of this thesis, beginning with some background on the topic itself and concluding with a summary outline of the key results and conclusions which will then be detailed in later chapters.

1.2 FRAMING THIS RESEARCH

Traditionally, television media scheduling and advertising decisions have been based almost entirely on quantitative estimates of the size and demographic composition of programme audiences. In essence, the ‘ratings’ assess the relative popularity or success of programmes according to how many people are watching. These quantitative ratings then determine the value of specific programmes to broadcasters and advertisers. Such ratings systems have evolved over time (e.g. moving from simple diary collection methods to sophisticated electronic ‘peplemeters’), but the essential ‘currency’ of audience size ratings remains one that was originally devised to serve a simpler media environment. Up until the recent past, most television markets were dominated by relatively few mass-audience channels. Advertisers had limited TV broadcasting options and often very little control over the exact placement of advertisements. There had to be a currency for buying and selling advertising time, but what mattered most was industry acceptance of a standard and clear-cut ratings system. Technical (e.g. statistical) questions would occasionally be raised and addressed concerning this system’s *accuracy* but its overall appropriateness or versatility was seldom questioned.

In recent years, television markets worldwide have rapidly grown much more complex, fragmented, and competitive. In line with this, advertisers now have many more television advertising alternatives to consider, as well as ever greater choice and control over the specific programmes and environments in which their advertisements appear. This raises many interesting issues and problems. In an increasingly multi-media environment of satellite television, cable television, video recorders, video games, and remote control ‘zapping’, the fact that a programme reaches the home television screen can no longer be seen as any guarantee that it will actually reach the viewer. Viewers may not be paying full attention; they may be engaged in other activities; they may be constantly changing channels; or they may not even always be physically present (e.g. they may be continually leaving and re-entering the room during the course of a programme). And even if the number of viewers who are ‘watching’ a programme can be accurately measured, does this mean that these viewers will remain attentive to advertisements shown during the programme?

Few of the above concerns are truly new. Over the past thirty to forty years, academic and television industry researchers have periodically returned to the fundamental issue of how to define ‘viewership’ in terms of some sort of audience size ratings. Various kinds of data have been collected, based on more restrictive definitions of viewing than the traditional standard of being ‘in the room with the television set turned on’. All such studies, however, naturally lead to smaller estimates of the number of ‘viewers’ for a programme and the industry has invariably then fallen back on its more broadly based viewing estimates. It seems that broadcasters and the advertising industry regard *big* numbers as the only *good* numbers. (This insecurity is presumably a legacy of the early days of television’s rivalry with print advertising.)

While the use of audience size ratings continues essentially unchanged, there has been a recent resurgence of interest by media planners in other more 'qualitative' television audience research measures. Such measures are designed to consider not only the *number* of viewers who may be 'in the room and able to watch', but also the relative *'quality'* or *'value'* of audiences for particular programmes or channels. While traditional audience size ratings implicitly assume that 'if viewers watch a programme then they like it', another way of addressing the relative popularity of programmes is to ask viewers directly and to try to formally measure such constructs as 'involvement', 'attention', or 'liking'. Again, many such possible audience research options have been proposed, developed, and analysed over the years, but the simpler television environment of the past mostly precluded their widespread use and acceptance in the buying and selling of TV advertising time. In today's much more complex viewing environment, might the measurement of audience attitudes and reactions now be seen as an important and useful complement to traditional 'head-count' ratings?

Within this relatively broad television audience research framework, this thesis concentrates on television programmes (rather than advertisements) and revisits one particular form of audience reaction measurement: appreciation ratings for television programmes (i.e. ratings of how much viewers like or enjoy the programmes they are watching). Utilising an extensive sample of previously confidential UK panel data, this research follows principles of data reduction to look for patterns in such audience appreciation ratings.

The thesis is structured as a programme of research into aspects of Audience Appreciation, rather than as a single self-contained study with one sharp predetermined focus. There are a series of largely independent (but

complementary) analyses carried out in separate chapters. Among the issues investigated, for example, is how far aggregated appreciation ratings for programmes might be related to traditional ratings of audience size. Also, to what extent can individual viewers perhaps be segmented both demographically and behaviourally along audience appreciation lines? A better knowledge of the existence and extent of such patterns will help both practitioners and academics in considering the practical application of this type of data for media planning and evaluation. The findings of this research will lead to a better understanding of the existing television medium and will be of continued interest in line with future television and multi-media developments (e.g. the issue of 'narrowcasting' to be explored in Chapter 6).

The scope of this thesis does not extend to directly investigating programme context effects (e.g. are advertisements contained within highly appreciated programmes more likely to be noticed?). However, an appendix will review the marketing literature in this related area - of great fascination to advertising practitioners and to theory building academics - in light of the findings of this new research.

1.3 BACKGROUND TO THE CHOICE OF TOPIC

To place this research in a personal context, I arrived in London (from a prior educational background in Canada and the US and from a number of years of work experience in Taipei, Taiwan) with a strong interest in exploring various issues related to television audience viewer behaviour and the measurement of advertising effectiveness. This general field of study would take advantage of the strengths and expertise of my academic supervisor, Professor Andrew Ehrenberg, and of other faculty at London Business School. It would also

capitalise on London's position as an important base for international television companies and for major advertising agencies. Indeed, throughout the formative stage of developing and narrowing down a specific thesis topic, I benefited greatly from ready access to experts and professionals in the field (e.g. at Channel 4, the BBC, the Broadcasters Audience Research Board, etc.), and from the availability and use of London's specialist media-related libraries.

As is often typical of new doctoral students, one sets off somewhat naively on the process of 'getting a PhD' with grand hopes and expectations for the scale and scope of a proposed thesis. In this regard, my original intention was to carry out an internationally oriented study that would encapsulate cross-cultural measures of audience viewing behaviour and contribute to important debates on the measurement of advertising effectiveness in dynamic and rapidly evolving international broadcasting environments. My first attempts at developing specific research questions and designing possible experiments soon floundered over issues of sheer practicality (i.e. expenses that would be involved in cross-border studies, related time constraints, notoriously difficult constructs to research, etc.). In continuing to move through the early stages of developing a formal thesis proposal, I therefore started to consider previous research that had been carried out by my supervisor and his colleagues and to explore possible options for building or extending on this work. Accordingly, I soon found myself reading intensively about and developing an interest in audience appreciation ratings, one of their relatively minor prior research topics but one which led to substantive early findings.

There is a long history of audience appreciation measurement in the United Kingdom, dating back to the first survey of radio audience appreciation by the British Broadcasting Corporation (the BBC) in 1941. With a current television audience appreciation measurement system that can be traced back to

development work in the late 1960s and 1970s, the UK remains one of only a few countries worldwide to have regularly collected this type of appreciation data (Gunter and Wober 1992). The key measure traditionally calculated for a television programme is its Audience Appreciation Index or “AI” score. This will be outlined in detail later but is essentially an aggregated score on a scale from 0 to 100 which represents how much viewers who watched a programme actually enjoyed or appreciated the programme.

Far from being advertising-related, the collection of such audience appreciation data over the years largely reflects the UK’s public service broadcasting ethos and a related desire by broadcasters to demonstrate that they are serving and satisfying their viewing audiences. Audience appreciation data is currently collected under the auspices of the UK’s Broadcasters’ Audience Research Board (BARB) in the form of the TV Audience Reaction Service. BARB is also responsible for standard UK audience size ratings in the form of the TV Audience Measurement Service, which provides quantitative audience data on a regular basis. Although regular audience ratings data is used both by UK broadcasting organisations and by commercial organisations (especially those in the advertising industry), audience appreciation data has always been reserved for the sole internal use of the main UK broadcasting organisations who jointly commission it (the BBC, the Independent Television Association, Channel 4, S4C and the Independent Television Commission). This data has remained highly confidential and has never been released or sold for use by advertisers and media planners.

It is an interesting reality that if something is seen as ‘forbidden’ or ‘confidential’, it somehow automatically seems to assume an air of importance or relevance. As will be discussed later in this thesis, there have been few

published studies on BARB's audience appreciation ratings over the years. A great deal of column space on this topic, however, has been filled with conjectures and editorials in the popular media and advertising industry press. The following are just two recent perspectives:

“Why are broadcasters willing to let us know how many million people watch their programmes but not willing to let us know how much they enjoy them? There is a strong case to be made..... that viewers' appreciation of programmes is as important an index as audience size. BARB argues that an extra set of figures would confuse matters. What are they worried about?” (Stewart 1995)

“Audience appreciation has been a hot potato for years. While an industry measurement of audience appreciation has long been available to the broadcasters selling the TV airtime, the buyers and advertisers have been kept in the dark. At the 1993 TV conference in Monte Carlo, a presentation from Zenith Media suggested that audience appreciation was a red herring. Other agencies, including TMP partners, believe that an appreciation measurement can give them and their clients an edge in media negotiations.” (Beale 1993)

In September 1992, as I was in the process of narrowing and developing my thesis topic and as I had coincidentally begun reading a great deal about audience appreciation research, BARB began advertising for approaches from companies interested in running a new audience reaction measurement service to replace the panel diary system that was currently in existence. The new BARB contract generated a great deal of discussion and interest amongst broadcasters and advertising agencies, both in the UK and abroad. As an expensive electronic collection system was one option under consideration (Dawkins and Samuels 1994), there was now talk of the possibility that audience appreciation data could for the first time be released and sold as a supplementary aid to television time buying.

Even without the release of such syndicated audience appreciation data, advertising agencies and media planners at the time were showing increasing interest in the self-collection of similar types of proprietary data. Consequently, a number of closely related audience appreciation studies were being undertaken within the industry. This seemed to suggest to me a number of interesting implications for further research. Whether or not BARB did ultimately decide to release and sell their own data, anyone could conceivably set up a similar audience appreciation data collection system. How might media planners use and interpret such data, and would it really be of any additional value over the ongoing use of traditional audience size ratings? Even if such data was of value, collecting it on a continuous week by week basis (as done by BARB) is also a very expensive undertaking. From a media planning perspective, could such continuous and costly data be justified or would more occasional ad hoc audience appreciation surveys provide sufficient and appropriate data?

Overall, in line with this high level of industry interest and with a rapidly transforming television media environment, I saw the opportunity for new academic research in this area to be relevant, timely, practical, and of use to both practitioners and to marketing academics. With BARB's cooperation, I was granted access for academic purposes to a substantial sample of previously confidential audience appreciation data. The extent of this data would allow me to revisit past research to establish findings at aggregate and individual levels which until now had been based mostly on small samples and on incomplete diary methods. Furthermore, this data would provide ample scope for new analyses and for new interpretations in light of a changing TV advertising marketplace.

1.4 THESIS STRUCTURE

This dissertation replicates, builds, and extends directly on a well established body of prior research. The structure of the thesis as a sequence of largely self-contained chapters reflects the fact that this write-up summarises the results of a varied series of approaches to audience appreciation instead of a single main study. For this reason, rather than having a comprehensive but multi-faceted literature review early in the thesis, a more limited initial review provides a brief introduction to key prior research and to a few fundamental concepts. Similarly, the main methodology chapter concentrates primarily on defining an overall research philosophy and on describing the data. Finer methodological issues, as well as specific comments on the findings of past research, are then mostly incorporated and discussed where they are *directly applicable* within subsequent chapters. This integrated approach ensures that replications and extensions are discussed in terms of earlier results, and that generalisable patterns and results are identified as such.

Following the introductory chapter you are now reading, this thesis moves through an overview of relevant literature (Chapter 2) and then directly into the main chapter on methodology (Chapter 3), which details the overall research framework and describes the data to be analysed. This is followed by five self-contained chapters (Chapters 4 through 8), each investigating a specific area of research and incorporating essential prior literature, methodological issues, analysis, and the presentation of results.

Chapter 9 concludes the thesis with an overall discussion, consolidating the findings of the previous six chapters and including discussions on limitations of this research and on possible future research directions. For interested readers, a broader review of the literature on programme involvement and its

relationship to advertising effectiveness is then included as an appendix. Although this thesis research does not directly investigate such advertising issues, this appendix offers a relevant opportunity to summarise and (where pertinent) discuss such past research in light of the patterns now established through my studies. It is in this appendix, therefore, that some of the direct managerial (i.e. advertising and media planning industry) implications of this thesis are discussed.

As will be detailed in Chapter 3 (“Methodology”), much of the value of this thesis lies in data reduction. This involves reducing large volumes of data into clear and simple tables where patterns, if they exist, should become readily apparent. The guiding principle of this research is therefore not merely to look for statistical significance and ‘best fit’ within a single set of data. (Given the large sample sizes available for analysis, attaining statistical significance in many cases here is not really even an issue of concern. Even differences too small to be of practical or theoretical interest are ‘significant’ - i.e. they actually happened in the population.)

Instead, through extending on past research, the aim here is to determine whether past relationships generalise to new and varied replications (or not), and whether results from new analyses generalise across multiple sets of data (e.g. across different programmes, across different programme types, across different demographic groups, etc.). For these reasons, this thesis is kept free and uncluttered wherever possible of complicating statistical jargon and measures of significance. The reader is advised that specific notes on such matters will be made wherever issues of statistical significance or methodology do become of concern, but that findings will otherwise be presented throughout this thesis in simple and unembellished summary tables.

Finally, in line with the style and structure of this thesis, note that space is devoted at appropriate points throughout to outline and provide background to the *process* that was involved in completing this dissertation and ‘getting a PhD’ (i.e. what was done and why it was done at every stage of the project - as, for example, with the earlier discussion in this chapter providing background to my choice of this thesis topic area). This discussion has been explicitly included at the request of my examiners, both to demonstrate and detail the development of my own research skills, and to aid such development in other students who may refer to my thesis.

1.5 SUMMARY OF FINDINGS

This overview provides a more detailed summary of the analyses undertaken and the conclusions reached in Chapters 4 through 8 of this thesis. Moving in order on a chapter by chapter basis, the key questions addressed are briefly summarised and the main findings are outlined.

An important distinction in terminology to note here, and one that is carried throughout this thesis, is that between *individual* appreciation patterns and *programme* appreciation patterns. The notion of *individual appreciation scoring patterns* takes the individual viewer as the unit of analysis, considering the ways individuals (or various groupings of individuals) differ in their appreciation scoring patterns. The average appreciation score of Individual A, for example, is the average score given by that individual to all the programmes he or she has viewed. In contrast, the notion of *programme appreciation patterns* takes the individual television programme as the unit of analysis, considering the ways programmes (or various groupings of programmes) differ in the appreciation scores they receive. The appreciation

score (or “AI” score as mentioned earlier) achieved by Programme X represents the average appreciation score that particular programme has received across all of its aggregated viewers.

(a) Individual Viewers’ Appreciation Scoring Patterns:

Following the general introduction, literature, and methodology chapters, Chapter 4 summarises patterns in the ways *individual viewers* differ in their appreciation response styles. Much of this chapter involves replicating and generalising from past research, with an aim of establishing and confirming the main groundwork for subsequent chapters (i.e. presenting already well-established individual level patterns in this data and then outlining the main dimensions on which viewers will be segmented and classified throughout later chapters). Patterns outlined include:

- Viewers mostly like what they watch. Individuals tend primarily to use the positive end of the six point audience appreciation scoring scale. A variety of demographic variations are also reconfirmed (e.g. women tend to give slightly higher appreciation scores on average than do men; older viewers tend to be more generous in their scoring patterns than are younger viewers; and lower income social groups tend to give slightly higher appreciation scores than higher income groups).
- Additional behavioural classification categories are then developed in order to establish further variations in individual scoring patterns (e.g. classifying individual panel members in terms of their overall weight of television viewing; in terms of their average overall level of appreciation

scores across all programmes viewed; and in terms of how consistent or varied they are in the way they make use of the appreciation scoring scale). Through classifying individuals as light, average, or heavy viewers of television, for example, previous findings are again confirmed whereby heavy viewers are demonstrated to give higher appreciation scores on average than do lighter viewers.

- In one exception to the straightforward analysis and data presentation techniques used elsewhere in this thesis, Chapter 4 also contains a CHAID (Chi-Squared Automatic Interaction Detector) segmentation modelling analysis. This exercise is based on the use of advanced statistical software to explore the issue of which of the demographic and behavioural variables discussed above contribute most to the prediction of whether individuals will be low, moderate, or high appreciation scorers on average across all the programmes that they view. The most important predictor of individuals' appreciation scoring patterns is their age, with older viewers tending to be more generous on average than younger viewers. Gender is also an important variable, with women on average being more generous than men, while social class and weight of viewing are less effective predictors of whether individuals will be classified as high or low appreciation scorers.
- Chapter 4 concludes by examining individuals' appreciation and viewing patterns for different programme types and for different channels. The findings again confirm, illustrate, and build on past research. Although specific individuals can vary greatly in their viewing and appreciation scoring patterns, there is a (perhaps surprising to many) lack of dramatic segmentation differences when individuals are grouped and compared across various demographic and behavioural subgroups.

(b) Programme Appreciation Scores:

Chapter 5 summarises patterns in the appreciation scores (or AI scores) achieved by different television *programmes*. Again, much of this chapter involves foundation work in replicating and reconfirming past research findings. Patterns outlined include:

- In summing and averaging responses to calculate programme appreciation scores (i.e. the average score given by all viewers who watched a particular programme), it is unsurprising that most programmes tend to be quite liked on average by their viewers. In practice, most programme appreciation scores fall within a fairly narrow range of from about 60 to 80 on a possible scale of 0 to 100.
- Some variation in programme appreciation scores can be found across different programme types or genres but these are smaller than is often supposed. For example, sports, information, light drama and some light entertainment programmes (e.g. situation comedy and quiz shows) generally achieve higher than average programme appreciation scores. Somewhat lower than average scores are then received by films, heavy drama, news, and much other light entertainment.
- The *range* of programme appreciation scores achieved by different programme types does vary more widely. For example, one news programme is typically not all that much different from another news programme, so the range of scores achieved by different news programmes is relatively narrow. In contrast, films are a more heterogeneous programme type and achieve a wider range of appreciation scores.

- Systematic audience appreciation patterns are similarly established across programmes shown on different channels, with the two smaller channels (BBC2 and Channel 4) achieving on average higher programme appreciation scores than programmes shown on the two larger channels (BBC1 and ITV). Such patterns seem to reflect differences in the range and types of programmes broadcast on the larger channels as compared to the smaller channels.

(c) Programme Appreciation and Audience Composition

Chapter 6 examines how the audience make up of programmes (i.e. ‘audience composition’) varies in general and across different programme types and different channels. Having established patterns of individual response styles (in Chapter 4) and having established overall appreciation patterns for different programme types and different channels (in Chapter 5), the question arises as to whether programme appreciation scores are simply a reflection of the programme’s audience composition. For example, does Programme A achieve a higher AI score than Programme B simply because Programme A attracts people (or viewer segments) who tend on average to give higher scores to everything they watch?

- Given scheduling factors and the availability of different audience segments to view, it is unsurprising that audience profiles (e.g. audience composition by demographic and behavioural segments) of different programmes can vary considerably. Nonetheless, these differences are smaller than is often supposed, especially when one considers programmes shown at roughly the same time on different channels. Such programme

audiences will then be largely unsegmented, reflecting the population composition of all those available to view at that time.

- In line with the above, and in spite of the individual level scoring pattern differences outlined in Chapter 4 (i.e. by different viewing segments), it is then established that little of the overall variation in average programme appreciation scores for programmes classified by type or by channel is due to audience composition effects. Light drama programmes on average achieve higher programme appreciation scores than films because such light drama programmes are ‘liked’ relatively more on average by *all* viewing segments, rather than because such light drama programmes tend to attract audiences on average that are biased towards high appreciation scoring segments (e.g. with higher proportions of older, female viewers). This is an important point as it suggests that an audience appreciation score (AI) for a programme reflects something qualitative about the programme itself, rather than just something about differences in audience composition and the appreciation scoring styles of the particular individuals who watch it.
- As the television medium continues to fragment, there is often talk of the rise of ‘narrowcasting’, with channels and programmes that are expected to reach small but well-defined and highly involved audiences. Cable and satellite channels, for example, would love to be able to demonstrate to advertisers that their programming delivers highly targeted and especially attentive audiences. The latter half of Chapter 6 represents an entirely new contribution of this thesis. It addresses the above issue through analyses to establish the extent to which the audience bias of a programme towards specific viewer segments might relate to enjoyment of the programme by those ‘targeted’ viewers in comparison to all other viewers who may be

watching. If a programme is developed for and targeted at younger viewers, for example, do these younger viewers actually tend to appreciate the programme more than the other non-targeted viewers who may be watching?

The results here show tantalising but very limited evidence in support of a narrowcasting thesis. Within this data, some examples can certainly be found where programmes do reach small and ‘targeted’ audiences (based on a variety of demographic and behavioural segmentation criteria). For example, there are programmes that attract biased audiences (e.g. heavily skewed towards male viewers). Furthermore, such programmes tend also to be enjoyed *slightly* more (in comparison to average appreciation scoring patterns) by these ‘targeted’ viewers (e.g. men) than by other viewers who are also watching (e.g. women).¹

Such examples in support of ‘narrowcasting’ are rare, however, and tend to be dramatic only at the extremes (e.g. the relatively few programmes that are very heavily biased towards male or female viewers). In segmenting viewers along other dimensions (e.g. age and social class), there is very little evidence of programmes with audiences that are not only heavily skewed towards a particular segment of viewers, but which are also liked relatively more by these ‘targeted’ audience segments than would be expected given already established patterns of appreciation.

¹ Such comparisons of liking have to take into account the individual variations in appreciation scores summarised in Chapter 4. For example, women on average tend to give higher appreciation scores than men. It is significant then to find programmes where men give higher scores on average than women.

(d) Programme Appreciation and Audience Size:

Chapter 7 replicates and then extends considerably on past research which explored the question of how much ratings of audience appreciation for a programme are related to ratings of audience size. This again is an important issue, especially when considering the possibility of utilising audience appreciation data as a supplementary media planning currency to traditional ratings. Does audience appreciation measure something different from audience size, and just how far are these two types of measures correlated?

As expected, and as demonstrated in past research, overall (for all programmes) there is little if any relationship between audience size and audience appreciation. This however ignores the many factors that can influence either audience size or audience appreciation ratings for a programme (e.g. scheduling effects, channel effects, programme type effects, and audience differences in the way viewers score programmes). The analyses carried out in this chapter therefore aim to explore the relationship between audience size and audience appreciation first within progressively narrow sub-samples of similarly grouped programmes, and then within even more narrowly defined specific audience segments for these programme sub-samples.

The results are mostly consistent with past research, although there are interesting instances where different findings emerge (to be described in detail within the chapter):

- Overall, when allowances are made for programme and audience differences, a definite ‘Double Jeopardy’ pattern does emerge (e.g. ‘bigger’ programmes are liked more on average by those who view them).

- In the most closely defined attempts to compare ‘like with like’ (e.g. looking at a particular time, on a particular channel, within a particular programme type, and amongst similar categories of viewers), a correlation of about 0.6 can be found between audience size and audience appreciation. The highest correlations and the most steeply sloped regression lines are mostly found for entertainment programmes. The lowest correlations and the least steeply sloped regression lines are found for more demanding and homogeneous programme types like “News”.

(e) Programme Appreciation and Repeat Viewing:

Chapter 8 considers how far audience appreciation for a programme episode might relate to the likelihood of viewers repeat viewing in subsequent weeks. Is audience appreciation therefore a useful construct to advertisers and programmers due to its association with peoples’ ‘loyalty’ to the programmes they watch?

Replication again plays a central role in this chapter, although this is replication with an important data-based difference. A variety of prior studies have suggested that the higher the appreciation score given by an individual, the more likely the person was to be a repeat viewer. Such findings, however, have invariably been based on quite limited data sets which typically only measured individuals’ claimed (and not actual) frequency of viewing. In this research, the extensive individual-level panel data available means that a viewer’s audience appreciation and repeat viewing patterns can be tracked longitudinally in detail from day to day and from week to week over this five week period. Furthermore, this can be done for many hundreds of programmes in contrast to the much more limited samples available in much prior research.

Results in this chapter provide powerful confirmation of patterns suggested in earlier research:

- There is a strong individual-level relationship between appreciation and frequency of viewing.
- Programmes which achieve higher average appreciation scores (AI's) also have higher levels of repeat viewing.

1.6 DISCUSSION AND CONCLUSION

Each of the above analysis chapters concludes with a brief summary. A broader overview of the key findings and patterns established is then given in Chapter 9. Overall, this thesis makes a contribution to knowledge in that it replicates and extends on a previously well-established but limited body of prior research. The key contribution of this research lies in providing further insights into how to interpret and use formal measures of television audience appreciation, including facing up to and bringing out the limitations of the operational data studied here. Many of the findings and patterns outlined will be of interest and of potential use in line with concerns that media planners are almost continually discussing or worrying about. If traditional television ratings measure viewing as being in the room with the set turned on, then measures of audience appreciation can at least to some extent move beyond such simple 'presence' to some indication of the 'quality of exposure' to programming (and potentially also to advertising).

2. AN INTRODUCTION TO THE LITERATURE

2.1 CHAPTER TWO OVERVIEW

As mentioned in the introduction to this thesis, this chapter does not aim to provide a formal and comprehensive review of all the literature which might relate to this research programme. Instead, this initial review is intended to provide an introductory foundation and to frame the boundaries of what will be covered in the remainder of the thesis. Specific details on the methods and findings of past research are then mostly incorporated within subsequent ‘analysis’ chapters where they can help refine or develop expectations for this new research, and also aid in the interpretation of any new findings. As mentioned in the introductory chapter, a discussion of some more broadly related bodies of literature (e.g. the literature on programme context, effects) has also been included in an appendix.

From a brief general look at various academic and commercial perspectives on television audience research, this chapter begins to narrow its focus to a more detailed overview of the particular body of research on which this thesis builds. This particular prior research into ‘patterns of television audience viewing behaviour’ is still very wide-ranging, encompassing many individual studies over an extended time period. A number of key established patterns of television viewing behaviour are highlighted here, since these will arise and be referred to repeatedly in later chapters. The focus of the chapter then narrows even further to introduce and outline prior research which has looked specifically at issues related to television audience appreciation. It is this specific research stream which forms the basis for the work that is undertaken in this thesis.

2.2 THE WIDE RANGING NATURE OF AUDIENCE RESEARCH

In less than fifty years, television has grown to become a dominant cultural institution in developed countries and a force of still growing influence throughout the world. As such a central feature of life, television (and the mass media in general) has consistently attracted a wide range of discussion and debate. Researchers in fields as diverse as economics, sociology, psychology, political science, communications, marketing, and advertising have all focused much attention on television and its audience, but often from quite different angles of interest.

A great deal of academic research in the social sciences, for example, has been conducted to explore the effects of television on its audience. Concerns raised have included violence and aggression, sexism and promiscuity, bad language, racial prejudice, reporting bias, political agenda setting, television's impact on children, and television's influence on societal values. While there are a wide variety of issues, several recurring themes have been identified in the literature (Barwise and Ehrenberg 1988). In particular, it is noted that research typically concentrates on concerns over television's *negative* rather than positive effects. Such criticisms often involve an implicit assumption that specific types of programming (and advertising) content will affect the audience's attitudes and later behaviour. Interestingly, people invariably seem to be most concerned about the effects of television on *other* people, rather than about any perceived negative effects that television may have had on themselves.

Rather than concentrating on the influence that television exerts on its audience, another important research stream in the academic literature has focused instead on how the audience *uses* television. 'Uses and

gratifications' research borrows from psychology and tries to explain how and why people watch TV (viewer behaviour) in terms of individuals' motives and needs. Television is seen as giving gratification to the user by satisfying these personal needs (McQuail 1983).

Media and advertising industry perspectives on television audience research tend by nature to be far more parsimonious and commercially directed than some of the more socially oriented academic perspectives. Despite numerous theoretical perspectives on the effects of television and on audience uses and gratifications, Judy Thomas of the DMB&B Worldwide Media Group recently summarised current advertising industry knowledge and research methods by saying "no-one knows what viewers want because they are hardly ever asked" (Thomas 1992). Instead, advertising and media planners draw inferences about what viewers want from such sources as television ratings data (providing audience sizes and demographic profiles for different programmes), industry data on the sales and ownership of home entertainment hardware, and sporadic surveys into specific areas of interest.

Ratings data provides the yardstick of 'success' within the television industry. For commercial broadcasters, the ratings indicate a programme's success at delivering audiences to advertisers and therefore largely determine the advertising rates that can be charged. For producers, ratings influence programme pricing, cancellation and renewal decisions. Even for public broadcasters, not dependent on advertising income, ratings are seen as a vital indication of how well the channel is serving the public's needs.

Reflecting the interests of broadcasters and advertisers, as well as the rise of such factors as cable and satellite television, home video, and the remote control 'channel zapper', audience measurement has become an increasingly

complicated affair in recent years. Measurement technology has responded to such challenges by moving from traditional panel diary research, to the widespread use of push-button electronic ‘people meters’. Although even the most current measurement techniques still suffer from some inadequacies and can often be criticised, they do offer extremely detailed and precise audience measures, particularly in terms of providing demographic information (Beville 1985; Ehrenberg and Wakshlag 1987, Ang 1991; Buzzard 1992). Although no ‘perfect’ measurement system is ever likely to arise, it can be expected that future advances will lead to ever more efficient and detailed analyses of the viewing audience, and to ever increasing volumes of generated data.

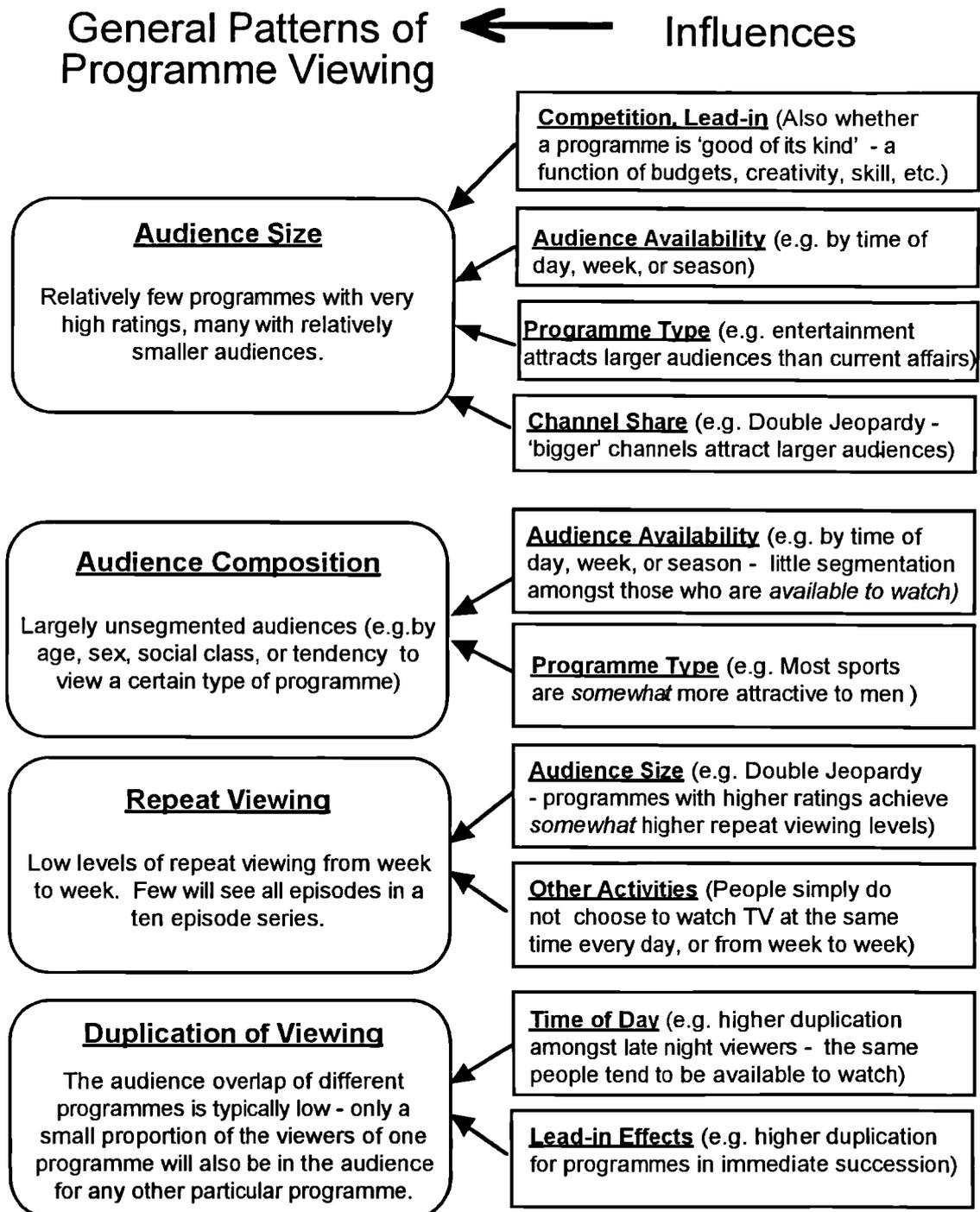
2.3 PATTERNS OF VIEWING BEHAVIOUR

While there are many theoretical perspectives on the effects of television and on the gratifications it offers to its viewers, do such theories reflect or account for actual patterns of viewer behaviour? Similarly, while there is an extremely large amount of commercial research into television ratings, what do these marketing information systems actually tell us about viewer behaviour?

One problem with having such a great continuing volume of detailed ratings data, is that there is often too little empirical research and reflection into the ‘big picture’. A body of research into ‘patterns of television viewing behaviour’ has addressed this issue over the years through conducting secondary analyses of commercial television industry ratings data. Although this body of research reflects many separate studies, spanning a period of more than 30 years, most of the main findings are summarised in two books (Goodhardt, Ehrenberg, and Collins 1975; Barwise and Ehrenberg 1988). Essentially, this research has aimed to establish generalisable patterns in the

way viewers actually behave and 'use' television. Overall, this body of work forms a key foundation for both the methodological outlook of this thesis, and for many of the replication oriented analyses that are undertaken. (For simplicity, this body of research will be referred to from now on as the "Ehrenberg Group's Research". This acknowledges the continuing influence of Professor Andrew Ehrenberg throughout the many studies being referred to here as an aggregate "body of research". References to specific studies will of course recognise the important roles of the other individual authors involved.)

Figure 2.1 provides a summary of the patterns of viewing behaviour that have been well established through the Ehrenberg Group's research (Goodhardt, Ehrenberg and Collins 1975; Barwise and Ehrenberg 1988). These studies have concentrated on the UK and on the US, although similar methods have been applied to data in Finland (Kasari 1985), and advanced media models have continued to be developed based on these now established patterns of viewing behaviour (Rust 1986). The essential conclusion is that television is very much an unsegmented mass medium (especially in comparison to other media such as radio or magazines), in which there are regular and systematic patterns of audience viewing behaviour. The following paragraphs provide further elaboration on these fundamental patterns of viewing behaviour.

Figure 2.1 Patterns of Television Audience Viewing Behaviour

(a) Audience Size:

A key determinant of audience size ratings is availability to view. For example, audience sizes are largest for prime-time evening programmes simply because more people are available to view at this time. Of the more than 1,300 or so programmes that will be broadcast during the course of a week on the four main UK channels, most will not be broadcast during the prime evening hours. The vast majority of programmes will therefore achieve relatively low ratings (although even say a 0.5% share of the UK audience would imply a considerable audience of almost 300,000 viewers). Individual viewers may differ greatly in their tastes and preferences, but in viewing 25 hours or so of television each week, they will choose a varied range of programmes including a few of the most popular high rating programmes and many more low rating programmes.

(b) Audience Composition:

Different programmes shown at the same time on different channels may have different sized audiences, but these audiences will be largely unsegmented, simply reflecting the population composition of all those available to view at that time (e.g. housewives, children, and the unemployed make up a higher proportion of the daytime audience). There are systematic exceptions, such as with men spending slightly more of their viewing on sports, and women on soap operas, but even these differences are not all that dramatic in relation to the total time people spend viewing. People do allocate their viewing time differently across different programme types (e.g. they spend more time watching entertainment programmes than informational programmes), but on

average across possible audience segments (e.g by sex, age, and weight of viewing) the proportions of time allocated to the different programme types on offer is much the same.

(c) Repeat Viewing:

One feature of television programming is that it is very repetitive, with different episodes of the same programme typically being shown at the same time each week (at least between major changes in programme schedules which may occur several times each year). Repeat viewing refers to the regularity of viewing of on-going television series. In this respect, viewers are shown not to be highly loyal to specific programmes. Less than half of the viewers watching a specific prime time programme in one week, will have seen the previous week's episode. Furthermore, virtually nobody watches every episode of a regular series of say 10 to 20 episodes over a season. In fact, typically far more than half the people who viewed any episodes at all out of a series of 10 will have seen three or less.

Low repeat viewing does not mean low reach, however, as the number of people who see at least one episode of a series over a period of time is much higher than the audience size for any given episode. If a series is seen in a given week by 15 percent of the population, for example, typically as much as half of the population will see it *at least once* over a 10-week period. Low repeat viewing levels also reflect the availability to view factor. Of the 25 hours or so of television an individual may watch over the course of an average week, there may be several programmes of high interest, concern, or involvement. Given the much larger number of programmes the individual will actually view, however, most programmes will be of no great concern to the

individual and will be watched rather impassively for relaxation or entertainment. In such a low involvement situation, people may not make themselves available, or may simply not choose to watch television at the same time every day, or at the same time from week to week.

(d) Duplication of Viewing:

This concept refers to the common audience of two *different* programmes. To what extent are viewers of one programme also viewers of another programme (e.g. one that may be shown on a different channel, on a different day, or at a different time)? In general such audience duplication is low and audiences for any two programmes are largely made up of different people, although there are exceptions (e.g. for two programmes shown in succession on the same channel, where there is a significant 'lead-in' effect; and for late night programmes, where audience overlap - though still low - is somewhat higher than normal because the same people tend to stay up late from night to night). The basic pattern that has been established here is known as the Duplication of Viewing Law. This demonstrates that the major influence on the level of audience duplication between two programmes is the *size of their respective audiences*: the proportion of the audience of programme B that also watches programme A simply varies with the rating of programme A, with only small deviations. (This basically says that if ten percent of the whole population watches programme A, then roughly ten percent of the viewers of B will also watch A, as will roughly ten percent of the viewers of programmes C, D, and E.) Given the predictability of audience duplication across different programmes, *independent of programme type or content*, an important implication is that there is no special tendency for viewers of one type of programme (say prime-time drama) to watch others of the same type.

(e) Double Jeopardy:

One final pattern mentioned in Figure 2.1, and one that will be returned to again, is that of Double Jeopardy (Ehrenberg, Goodhardt, and Barwise 1990; Ehrenberg 1993). Not only do smaller audience programmes (or smaller television channels) have fewer viewers, but they also tend to attract lower repeat viewing from week to week and to be liked less by those viewers who do choose to watch. This widespread phenomenon, first noted by the Columbia University sociologist William McPhee in the 1960's, occurs when people have to choose between broadly similar items that differ in popularity. McPhee named the effect Double Jeopardy because he saw it as unfair that an item chosen by fewer people should also be liked less by them (e.g. a small brand will be bought by fewer buyers than a big brand, it will be bought somewhat less frequently and it will be liked somewhat less on average by those individuals who do choose to buy it).

2.4 PATTERNS OF TELEVISION AUDIENCE APPRECIATION

There is a long history of measuring audience responses and reactions to television programmes in attempts to move beyond the measurement of audience numbers to some indication of how *involved* the audience feels with the programme. It is precisely here within the marketing and media literature where this thesis on “patterns of audience appreciation ratings for television programmes” provides its new contribution to knowledge.

An excellent and very comprehensive review of prior research in this field, examining research on audience reaction measurement both in the UK and

elsewhere, can be found in a recent monograph of the Independent Television Commission (Gunter and Wober 1992). For broad general background information, readers are therefore referred to this monograph. The literature review in this thesis can then be kept very directed and focused. While all prior studies of direct relevance will of course be discussed here, the ITV monograph does provide a variety of interesting additional material. In particular, more detail is provided on the historical and ‘political’ development of audience appreciation ratings systems in the UK, and especially on the development of the “interesting and/or enjoyable” measurement scale which will be introduced and discussed in Chapter 3.

(a) British Audience Appreciation Work

The Ehrenberg Group’s research into patterns of viewing behaviour, discussed earlier, has not all been based on the analysis of traditional ‘audience size’ ratings data. Indeed, a substantial substream to this body of research has centred on audience appreciation ratings. While other research into general patterns of viewing behaviour provides broad conceptual and methodological frameworks, it is these specific audience appreciation centred studies that provide the direct foundation to be built upon by this thesis.

The model for the UK’s audience appreciation measurement system can in fact be traced to development and analysis work by the then Independent Broadcasting Authority and its consultants, Professor Andrew Ehrenberg’s consulting company Aske Research Limited (Aske Research Ltd. 1973a; Aske Research Ltd. 1973b; Aske Research Ltd. 1975; Aske Research Ltd. 1978; Aske Research Ltd. 1980; Aske Research Ltd. 1981a; Aske Research Ltd. 1981b). Subsequent academic research by the Ehrenberg Group (particularly

including work by Patrick Barwise) has further built and extended on this initial development work (Barwise, Ehrenberg and Goodhardt 1979; Barwise and Ehrenberg 1982; Barwise 1985; Barwise and Ehrenberg 1987). These studies have used data from various earlier incarnations of the UK audience appreciation collection system (but not from the recent Broadcasters Audience Research Board system which provides the data for this thesis) as well as additional self-collected data from the U.S.A. In line with the replication work and new analyses to come in this thesis, details on the methods and findings of this prior research will be outlined in full detail in later chapters.

With the exception of the Ehrenberg Group's research on audience appreciation ratings, virtually all other studies that have centred on UK audience appreciation ratings have originated from within the broadcasting organisations involved in collecting this data. Two key authors of various studies which will be referred to, replicated, and built upon in this thesis are Peter Menneer, then head of the British Broadcasting Corporation's Broadcasting Research Department (Menneer 1987a; Menneer 1987b), and Dr. Mallory Wober, then Deputy Head of Research at the Independent Television Commission (Gunter and Wober 1992).

(b) Other Audience Appreciation Research

The measurement of audience responses and reactions to programmes is by no means limited to UK work based on BARB data (or on earlier incarnations of the current audience appreciation rating collection mechanism). A number of public broadcasters in other countries developed similar audience reaction measurement systems during the 1960's and 1970's and continue to collect

this type of data (e.g. Canada, Australia, New Zealand, and the Netherlands). In the US, during the same period, related systems of qualitative ratings were developed by various private companies. None of these companies found great success in trying to market measures of audience reaction to advertisers and media planners as a supplementary currency to audience size measures. Although several such research systems survive in the US, these are conducted privately by television and cable companies for internal programme development purposes (Gunter and Wober 1992).

In the 1980's, a company called Television Audience Assessment (TAA) was set up in the US to develop and market measures of audience reaction to the major television networks and to advertisers. TAA's work was in line with, and built on the Ehrenberg Group's earlier research in the UK. Although for a variety of reasons this effort also failed commercially, it did leave behind a seminal body of carefully conducted research (TAA 1983a; TAA 1983b; TAA 1984a; TAA 1984b). As with the key UK research outlined above, the specific findings of this research will be detailed and referred to where appropriate in later chapters.

2.5 SUMMARY

This chapter has provided a brief overview and introduction to the literature which frames and provides a foundation for the analyses to come in this thesis. This review has narrowed from an initial broad perspective to identify some key research streams. These include:

- The Ehrenberg Group's research on general patterns of how and what television audiences view (e.g. this thesis will reflect and allow for known

patterns of viewing, such as established repeat viewing, and audience composition/segmentation patterns, which will now form a basis from which to explore patterns in how much audiences actually like or appreciate what they view).

- The Ehrenberg Group’s specific prior work using audience appreciation data (e.g. this thesis will build and extend specifically on foundation work looking at the relationships between audience size and audience appreciation, and between repeat viewing and audience appreciation).
- Other earlier UK audience appreciation work that has been based on and has established general patterns in, BARB data (e.g. studies mentioned earlier by Menneer and Wober), as well as US research findings based on similar forms of data (e.g. the extensive work by Television Audience Assessment in the 1980s which in turn largely built on the earlier work of the Ehrenberg group here).

These “key research streams” do largely represent a series of prior empirical studies that have established (or in certain cases have just begun to establish) patterns and findings related to audience appreciation ratings. All of these earlier studies, however, have had certain limitations that this thesis, based on a new and extensive set of data, can begin to overcome. (For example, the large data sample available here will help minimise sample size concerns that arose in prior research looking at audience appreciation and its relation to both audience size and repeat viewing patterns. Similarly, the detailed longitudinal nature of this data will overcome limitations in earlier repeat viewing studies where conclusions were based on non-continuous data and on measures of viewers’ claimed as opposed to actual frequency of viewing.)

The thesis is therefore more than a data mining exercise. Everything that has come before provides ‘hypotheses’ or ‘expectations’ for this new research (i.e. empirically based expectations). One aim of this research is to check on and confirm prior findings (i.e. the search for *generalisable* patterns). This does not just mean a series of straight replications, but differentiated ones (under different conditions) where the aim is to understand and/or quantify these patterns better. A second aim of this research programme is then to expand knowledge through completely new analyses which nonetheless build and extend directly from this established base of prior findings. Overall, this research is about first exploring and then establishing patterns; patterns in the ways that *individuals* rate their appreciation of television programmes, and patterns in the scores that *programmes* achieve when appreciation scores are aggregated across all individuals who are watching. What are these types of audience appreciation ratings in effect actually measuring, and how do they relate to more traditional ratings of audience size?

As a final comment on the literature in this field, it should be noted that many of the empirically based viewing patterns outlined in this literature review (e.g. repeat viewing as summarised in Figure 2.1) would seem to suggest an overall low-involvement framework for television viewing behaviour. This is in line with Krugman’s often cited and largely speculative earlier papers on the impact of television advertising (Krugman 1965; Krugman 1967). The simple fact is that most television viewing takes place for relaxation or entertainment purposes, often without detailed prior planning or out of habit, and quite often in a context where the viewer is concurrently involved in other activities (i.e. reading, talking, eating, dozing, etc.).

Few of the many programmes a viewer may see during the course of a week will demand his or her complete devoted attention and enthusiasm, and even the most involving and demanding television programmes will very seldom be all that involving. Television is a mass medium, and even special interest or other targeted programmes must be developed to appeal to relatively broad audiences. This is especially true in comparison with the print media, for example, where much more detail and depth is possible on a subject and where low production costs mean that extremely narrowly-targeted magazines and journals can still make economic sense. This overall low-involvement reality must be recognised before one even begins considering or developing specific measures of television audience involvement levels (i.e. audience appreciation or other involvement measures). The patterns of television audience appreciation that will be explored in this thesis should therefore be seen in the context of variations in interest and enjoyment levels across what is essentially a low-involvement media.

3. METHODOLOGY

3.1 CHAPTER THREE OVERVIEW

In this chapter, the methodological principles that underlie this research programme are described and explained. This is followed by a detailed overview of the actual data sample analysed and a brief outline of the general sequence of analyses undertaken.

3.2 METHODOLOGICAL PRINCIPLES

This thesis represents a programme of empirical research that is based on the analysis of a substantial set of UK audience appreciation panel data. Although a primarily inductive and data driven approach is followed, this embodies much more than the relatively blind application of a sophisticated analysis technique to a large and messy set of data. Instead a more methodical ‘step-by-step’ approach is undertaken, where the different analyses to follow incorporate many expectations and implicit hypotheses that are either derived from past research or are newly framed and developed through the course of this research programme.

As discussed in the introduction to this thesis, the guiding methodological principle here is not merely to look for ‘significance’ and ‘best-fit’ but to go well beyond this. The word “patterns” is particularly central to the outlook of this research. The value of this thesis lies in data reduction, using relatively straightforward analysis and data presentation techniques to reduce large volumes of data and to identify systematic patterns (Ehrenberg 1991). The concern is with “*modelling* data, not in the sense of representing data in

some more or less abstruse mathematical or graphical form but in the sense of discovering, displaying, and seeking to generalise patterns in the data” (Collins 1992). The implications of these patterns can then be considered in line with the possibility of this type of audience appreciation data ever being used as a supplementary media planning currency.

Much of this research does first centre on revisiting, replicating and extending directly upon past findings. But a variety of new and completely original analyses are also undertaken. Even where replication work is involved, this does not imply an exact duplication of past studies (Lindsay and Ehrenberg 1993; Hubbard and Armstrong 1994; Ehrenberg 1995). Rather, by building upon a well-established body of prior work, the distinct contribution to knowledge offered by this thesis can be seen as both incremental and measurable. Where the results of new replications are consistent with the findings of the original study, such replicated patterns are a powerful confirmation that imply statistical significance and good fit. Where new results might differ from past findings, this opens opportunities for discussion, debate, and statistical testing.

This overall emphasis on replication and extension is justified for two key reasons. Firstly, a primary aim of this research is to look for and establish systematic and generalisable patterns in this type of audience appreciation data. The large and comprehensive set of recent data available here means that previous research hypotheses and findings, which until now have often been based on small samples or on incomplete diary methods, can be revisited and either re-established or refuted. Secondly, although a substantial body of prior research in this area exists, the television media environment has changed dramatically in recent years. This can only lead to new stances for interpretation, even when one is replicating and extending on past findings.

Further support for the approach taken in this thesis can be found in recent articles by a number of leading academics. Indeed, an entire recent issue of *Marketing Science* (1995, vol. 14, No. 3) was devoted to the topic of “empirical generalisations in marketing”. A key point is that replications, based on the examination of multiple sets of data, are needed in order to build knowledge and to develop *empirically grounded theory*. Academic research in marketing is currently quite at odds with most research in the natural sciences in that it is almost completely dominated by the hypothetico-deductive approach (Barwise 1995). Such an approach sees the development of theory *followed* by empirical observation even though there is no reason why theory must or should always precede empirical observation (Bass 1993). It is quite fair to raise the argument that this “mainstream-academic” approach has produced few generalisable or lasting results over the years (Ehrenberg 1994). A consequence of the striving for originality in journal articles is the development of explanatory theories and models which are then typically verified with a limited empirical test against a single set of data.

The results of two recent studies illustrate why the above issues should be of concern to marketing academics. One study shows that few replication papers are published in the major marketing journals, yet published replications typically produce results that conflict with the original study (Hubbard and Armstrong 1994). A second study questioned the predictive value of ‘scientific’ knowledge of consumer behaviour (Armstrong 1991). It was expected that people familiar with many years of research on consumer behaviour would be able to make better predictions (e.g. to successfully predict an already established result about phenomena in this field). Academics, high school students, and practitioners were asked to make predictions for the outcomes of research into 105 hypotheses from 20

empirical studies selected from the Journal of Consumer Research. Contrary to expectations, the experts (academics and practitioners) were no more accurate than the novices (high school students), and academics were no more accurate than practitioners. In fact, none of the subject groups made predictions that were more accurate than chance. This study raised one important issue. In order to be of value, a body of research in consumer behaviour (or in any other field), should be cumulative and should build grounded empirical knowledge. Has extensive past research into 'consumer behaviour' provided a coherent foundation of use in predicting and modelling how consumers will actually behave?

This thesis reflects and incorporates many of the concerns outlined above. This is not to imply that research based on the hypothetico-deductive approach is 'bad', only that the marketing field (and its various publication outlets) should encourage a more normal and balanced range of approaches. Perhaps the day will yet come when doing a "replication of someone else's work and/or reanalysing their data might become a routine part of doctoral training" (Barwise 1995).

3.3 A QUESTION OF SIGNIFICANCE

Before outlining the data sample in detail, and in preparation for the subsequent ways that this data will be presented, analysed, and discussed throughout this thesis, some further methodological discussion needs to take place here on the question of 'statistical significance'. The philosophy underlying this research is that significance levels do not always need to be shown or discussed (unless there are realistic concerns that patterns in the large samples presented may *not* be significant).

With the often very large sample sizes available here, merely attaining statistical significance in most (but not all) cases is not really even an issue of concern. Even differences too small to be of either practical or conceptual interest are significant! In any event, this thesis typically incorporates even more explicit tests of the robustness or 'significance' of its findings. The appreciation scoring patterns to be outlined in Table 4.2, for example, *generalise* from prior findings (i.e. one can 'expect' these patterns). This process of generalisability is very much a form of sensitivity analysis - even where conclusions in the past may have been limited or based on inadequate data samples, they do lead to 'expectations' for what one might find. If such patterns and relationships do then replicate, especially in line with new and more extensive data samples or with new methods of analysis, then this is indeed a confirmation of the 'significance' or 'existence' of such patterns.

Statements that a result is 'statistically significant' are often over-used and over-interpreted in academic research. There are several things that this phrase does *not* mean: (i) that this observed difference is large (only that it exists and is probably real) and (ii) that the result is actually important (Ehrenberg 1991). With this in mind, I would like to make the following six bullet points to support my reasoning for not widely and repeatedly reporting significance levels throughout this thesis:

- **The Theoretical Basis:** Traditional significance levels (probabilities) are calculated on the null hypotheses that there is no (significant) difference. Once a significance level is established, the probability levels become 'incorrect' for that data. So, one should ask, why report them? There is no real reason except possibly as an, I think, inappropriate and certainly indirect way of saying 'large'.

- **Other Meanings:** Significance levels are indeed widely reported as largely unthinking substitutes for “Large” or “Important” or “Interesting”. In this thesis, I basically reject that use for the following reasons:

“Large”

This has to be established from the difference in mean values (or the %), by comparing that difference (e.g. 70-74 = 4 in average Audience Appreciation Scores given by males and females - as will be presented in Table 4.2) with:

- a) The relevant scale (e.g. the 0 to 100 audience appreciation scoring scale used in this thesis), or part of the scale (60-100).
- b) Other such differences (e.g. Table 4.5: will show a 74-69 = 5 difference in mean audience appreciation scores for heavy versus light viewers of television).
- c) Apparent practical or conceptual implications or applications (if any) of the difference.
- d) The actual number of women scoring higher than men.

In all these respects “70-74” is small. Whether the difference is ‘significant’ will ‘of course’ depend on the sample size and also on the internal variability of the data which again ‘of course’ has nothing directly to do with how big the male-female difference in the population actually is!

I also note that what makes the use of ‘significance’ to mean ‘large’ worse is that most reported tests of significance do not even report the difference in mean values (or % rates) in question. This is so not only in ‘bad’ papers, but also in much of say the Journal of Marketing Research and in virtually all statistical texts that I have seen. I have even been told by my supervisor Andrew Ehrenberg, that in many years of Rothamstead Annual Reports on ‘significant’ differences in agricultural trials, the crop yields for these ‘significant’ differences were never reported (F ratios and probability levels Yes, actual yields No.) [Rothamstead is where significance testing was virtually invented by R.A. Fisher.]

“Important” A 1 or 2 degree difference in body temperature can kill at around about 105 degrees Fahrenheit; not elsewhere. How does ‘significance’ come into that?

“Interesting” This depends on the subject matter, and not on the sheer numbers.

- **Opinions of Professional Statisticians:** Through my supervisor, I am also aware of a large and growing body of written opinions amongst professional statisticians who are against tests of significance (mostly, I think, for the kinds of reasons I outline here).
- **Similar or Differentiated Samples:** Statistical significance establishes (from the random variations of individual readings in the given sample)

how far other random but hypothetical samples from exactly the same population would ‘behave similarly’ (e.g. show a male/female difference in audience appreciation scores of about $70-74 = 4$).

In contrast, empirical generalisation establishes how far systematically different samples also do or do not show that (same or equivalent) male/female difference. For example, in this thesis, I will do replication analyses to see whether male/female average appreciation scoring differences hold from one week of viewing to the next. I will also make similar comparisons of male/female audience appreciation scoring patterns across and between different age groups and social classes. The male/female difference will then be explored still further across programmes of different types or genres.

Showing whether a difference actually holds again under different conditions (as I have done in this research) is much more powerful and informative than showing that it would hold again under (exactly) the same conditions.

- **Randomness:** Significance is predicated on the data being random samples from a defined population. But most data are not randomly sampled. As will be outlined later in this chapter (in Section 3.7), I have used two main sampling frames (a ‘large’ and a ‘small’ sample) for many of the analyses carried out in this thesis. There are reasons, and the choice is certainly not random, for why one or the other of these two sampling frames is used.
- **Population Data:** I note that statistical texts do not, in general, explicitly show or discuss how to ‘analyse’ and interpret differences between

populations (e.g. **all** males in the UK and **all** females). Yet that is what one has to do next once a sample result is ‘significant’ (i.e. approximately true for the whole population). One then has to start *interpreting* it, which is just what I will continually try to do through the analyses in this thesis.

3.4 THE BARB TELEVISION OPINION PANEL

The Broadcasting Audience Research Board (BARB) and its audience appreciation data collection process were briefly outlined in Chapter 1. As mentioned, this data is collected by BARB on behalf of the main UK broadcasters who jointly commission it. Following numerous meetings in 1992 with Bob Hulks, then BARB’s Chief Executive, I was granted access to this previously confidential data for academic purposes. The Audience Reaction Service, known as the Television Opinion Panel, was at the time provided by the British Broadcasting Corporation’s Broadcasting Research Department to BARB.¹ My contacts for securing this data on tapes were therefore through the BBC.

The BARB Television Opinion Panel, as managed from 1983 until 1993 by the BBC Broadcasting Research Department, was administered weekly by post using a paper diary system. Each week, panel members received a booklet covering 7 days programming, running from Monday through to Sunday. The

¹ In late 1993, the new contract for providing the Television Opinion Panel service went to RSL-Research Services Limited, a media and communications company which continues today to collect audience appreciation data on BARB’s behalf. Although an electronic collection service had been under consideration, and a lengthy pilot study was conducted, BARB’s final decision was to return to the traditional collection method of having viewers fill in paper diaries. This decision reflected timetable pressures more than any lack of confidence in electronic technology (Dawkins and Samuels, 1994; Kleinman, 1994).

Programme Diary enabled the panel member to simply and quickly give an appreciation evaluation for each programme seen during the week, using the scale to be outlined below. Respondents were instructed to: “Watch whatever you would normally watch. Please do not alter your viewing habits just because you are a panel member. Please do not give your opinion about a programme in the Programme Diary unless you have actually watched some of that programme for that week.”

(a) The Appreciation Rating Scale:

Programme evaluations were based on a six-point ‘portmanteau’ (or two-in-one) rating scale in which individuals said “how interesting and/or enjoyable” they felt each programme they viewed to be. In order to offset a tendency for responses to cluster at the positive end of the scale (people typically do not watch programmes they do not like), the six point scale was asymmetric, with three positive options, two negatives, and a neutral point. To calculate an easy to understand (i.e. a score out of 100) Appreciation Index (AI) for a programme, scores were attached to each point on the scale as follows:

Point on scale	Evaluation of Programme	Score
6	Extremely interesting and/or enjoyable	100
5	Very interesting and/or enjoyable	80
4	Fairly interesting and/or enjoyable	60
3	Neither one thing nor the other	40
2	Not very interesting and/or enjoyable	20
1	Not at all interesting and/or enjoyable	0

The aggregate AI score for a programme, calculated by summing the scores for all individual responses to that programme and then dividing by the number of respondents, could in theory therefore range from 0 to 100. Although AI scores for different programmes do vary, in practice they mostly fall within a fairly narrow range of from about 60 to 80. The average for all programmes, for example, is 73 or 74.

At the early stages of developing a thesis topic, there is a need to narrow down specifically ‘what to research’ within a broad general area. For this research, I made the decision early on that the BARB audience appreciation rating scale would be taken as it is (i.e. this thesis is not concerned with developing new or improved scales for audience reaction measurement). Working with this previously collected data, it was not necessary to experiment with new or enhanced scales for the measurement of audience reactions.

An alternative programme of thesis research might have concentrated solely on the development and testing of new scales for collecting this and other types of audience reaction data. This could, for example, have involved the collection of new primary data using this and various proposed alternative scales or methods. In terms of this thesis, however, such an alternative was not pursued due both to the high expenses that would be involved in self-collecting audience appreciation data, and to the fact that scale-development is not an area in which I felt I could best make a contribution to knowledge.

From the outset, I recognised that others have put much effort into the development of this scale and that it is the result of a great deal of research and testing (as I will outline briefly below and as is reviewed in greater detail

by Gunter and Wober 1992). The perspective of this thesis was therefore to recognise that this sort of data, using this “interesting and/or enjoyable” scale, was already being collected on a large scale for broadcasting purposes and that advertising agencies and media planners were also experimenting with data collected using the same scale (or quite similar variations on it). In this respect the value of this thesis lies not in identifying new types of data that *could* be collected, but in helping to understand and interpret findings from data that is already being collected and used at great time and expense to the broadcasting and advertising industries.

As will be raised repeatedly in later chapters, however, the relative ‘sameness’ of the data collected using this scale has proven to be a challenge and imposes a limitation on this research. In various presentations I have made on elements of this thesis, for example, a great deal of discussion is invariably generated by the use of a ‘double-barrelled’ scale that seemingly confounds all marketing research principles of ‘good scale design’. Furthermore, it is often argued that the ‘flatness’ of the data collected using this scale (e.g. most people quite like what they watch) may be due to a sheer failure of this scale to discriminate. Given these concerns, it is worth taking time in the following paragraphs to discuss this scale in further detail. How then did the “interesting and/or enjoyable” scale first evolve and come into accepted use, and what does this say about the type of data that is being collected using the scale?

Let us first consider the nature of the panel members’ ongoing task; that is to provide an evaluation of each and every programme viewed. Given this task, it is necessary to have a fairly simple scale which viewers can use across a wide range of programmes. It could be argued that viewers are capable of quite complex reactions to programmes and that more comprehensive multi-item scales might add further depth and insight. This would, however, turn the

weekly task of filling out a viewing diary into an unworkable and impractical exercise. (As will be summarised in the next chapter, the average individual may watch 30 or more programmes in a given week, and he or she is being asked to provide an appreciation score rating for each and every one of these programmes viewed.)

Given the necessity of having a single easy-to-use question, a great deal of research went into the original development of the “interesting and/or enjoyable” scale (Aske Research Ltd. 1981). This research showed that respondents *are* able to use it effectively to discriminate between programmes. The scale works in two ways. For programmes where the enjoyable part of the scale is more appropriate (e.g. comedy), this determines the answer given. Conversely, for programmes where the interesting part of the scale is more appropriate (e.g. news), this remains uppermost in respondents’ minds. The unusual asymmetric nature of the six point scale helps to bring out additional variance in that the positive options are the main ones that respondents actually use.

Subsequent research on alternative scales and wordings has shown that appreciation score results in any event do not seem to depend on the particular wording of the liking construct analysed (e.g. the same relatively ‘flat’ audience appreciation scoring patterns tend to be found whether one is using an “interesting and/or enjoyable” scale, separated “enjoyable” and “interesting” scales, a “liking” scale, or even a “good of its kind” scale Barwise and Ehrenberg 1982). All of this has actually been seen as encouraging for the future of appreciation ratings, as the use of different forms of wording will still lead to much the same conclusions about how much people like a given programme (Barwise and Ehrenberg 1987). Others have reached similar conclusions, after trying without any real success to improve

the discrimination of audience appreciation scales by experimenting with wider 10 point scales, or by allowing respondents to simply respond “on their own personal rating scale” rather than by using scales with specific key words such as “enjoyable” or “interesting”. (TAA 1983a; Dawkins and Samuels 1994).

These scaling issues will be returned to briefly in Chapter 9 (“Discussion and Conclusion”), once the reader has greater awareness of the detailed results of this programme of research. The question to be addressed then is whether any of these results are due to some failure of the “interesting and/or enjoyable” scale to discriminate or even to provide an appropriate measure of the ‘appreciation’ construct. In this sense, the analyses carried out in this thesis will contribute to increased cognizance of both the strengths and limitations of the audience appreciation scale and of the type of data it collects. For now, it is enough to emphasise that given this simple single-question measure of appreciation, this research is concerned with looking for patterns within data that represents individuals’ overall impressions or attitudes towards programmes (e.g. quite broad and general evaluations of programme likeability or appreciation).

(b) How Broadcasting Organisations Have Used this Data:

Despite the annual £2.5 million cost of collecting this data, it has been handled in a far from user friendly manner. As Hugh Johnson of Channel Four writes:

“Each week we get thick books of paper with thousands of numbers in them. This actually leads to the use of the data being largely restricted to the ‘comfort factor’ of examining individual AI figures for certain programmes. To be fair the BBC does sell computer tapes but Channel Four has not written the software to access the data this way.....The BBC

has not developed, to my knowledge, any special analysis programmes to help us examine the data. Tracking the data over time is probably the most actionable aspect - to quote an independent consultant who looked into the use of AIs for me 'the way the data are presented in the weekly books is useless.'" (Johnson 1992)

The "thick books with thousands of numbers" referred to here are BARB's "Audience Appreciation Summary" and "Audience Appreciation Report". These confidential weekly printed reports were delivered each week to the broadcasting organisations that jointly commissioned this research. The reports provided overall aggregated AI results for all programmes which satisfied the minimum sample size criterion (i.e. an AI was only reported if 25 or more responses were available for a particular programme). Further detail was then provided on the distribution of responses for each programme across the 6 point ratings scale, and AI figures were also calculated within different demographic sub-groupings (e.g. the AI score for each programme among viewers grouped by sex, age, social class, or BBC and ITV region).

As indicated in the quotation above, the broadcasting organisations involved in commissioning this appreciation data have developed few special analyses to explore the individual raw-level data in greater detail. Instead, programme makers and producers have simply relied on BARB's weekly printed summary reports. The "comfort factor" provided by these reports has nonetheless proven valuable. Broadcasters routinely use appreciation data to support programming decisions and to try and explain (and possibly take corrective action for) a programme's audience size. For example, although a new series may show disappointing or static audience size ratings over its first few months, it may also be showing steadily improving AI's over this same period. This could enable programmers to justify their storylines and predict that over time the series' following should likely build through improved audience size ratings (Menneer 1987a; Menneer 1987b).

In summary, the nature and confidentiality of audience appreciation data over the years has led to its being restricted to internal and largely non-sophisticated usage by the commissioning broadcasters. Few individuals have conducted extensive analyses of this data, especially from non-programming perspectives (i.e. media planning). Given that this is panel data, where the same individuals are responding week after week, it is especially surprising that the opportunity of tracking individual level raw data over time has not been exploited to any degree.

3.5 THE RAW DATA SAMPLE

For this research, a database of 5 consecutive weeks (the weeks of 19 April 1993 to 17 May 1993) of BARB Television Opinion Panel audience appreciation data was compiled. This particular 5 week period was chosen, after much discussion with BARB, because it is 'not peculiar'. (This period does not run over Christmas, or extended summer vacation periods where unusual programming may occur, and it begins the week following the Easter Holiday. Furthermore, there were few disruptions within this period to regularly scheduled programmes.) General patterns of viewing behaviour established over the years by the Ehrenberg Group (such as those outlined earlier in Chapter 2) justify why a five week period was felt to be more than sufficient. Apart from obvious exceptions (e.g. a 'big match', a royal wedding, the Olympics, etc.) one week of television programming is much like another. Furthermore, in following individuals' viewing behaviour over extended periods, repeat viewing patterns have established that few will see all episodes in even a five episode series.

This sample is extensive, comprising complete records of the individual level raw data from which BARB's weekly aggregated AI scores were computed. As will be outlined, a few previous academic studies have looked at BARB audience appreciation data. Such studies have mostly been based on summary programme-level aggregated data taken from BARB's compiled audience appreciation reports, rather than on the comprehensive individual level raw data which will be utilised in this research.

There are clear reasons why the commissioning broadcasters have been reluctant to spend much time developing software to explore and analyse this raw data. Although these broadcasting organisations could always have obtained data tapes from the BBC Broadcasting Research Department, these tapes used a very dated tape archiving process and the data contained was a simple and fairly unstructured 'dump' of raw data compiled from the paper diaries filled in by panel respondents. For this thesis, many months were taken up in the non-trivial process of writing programmes to read, clean, merge, and organise the data sample.² For example, as mentioned earlier, BARB has never tracked the same individuals over time. A great deal of data compiling and reorganisation was therefore necessary for this research in order to begin following each individual's audience appreciation and repeat viewing patterns longitudinally from week to week over this five week period.

² Other PhD students reading this thesis may be interested in many of the heated debates that occurred amongst those in my same year on the London Business School PhD Programme (in both Marketing and other disciplines). The thesis research of a number of these students involved developing questionnaires to collect new primary data for analysis. The remaining students were working with existing data (e.g. typically very large samples of stockmarket data, consumer panel data, or - in my case - data such as the audience appreciation data utilised in this thesis). Debates ensued about 'what was the easiest way to get a PhD?' Those collecting their own data firmly believed that it was far easier to be 'given' a set of existing data and to begin analysing this data than it was to go through all the time and effort involved in developing a questionnaire, in piloting it, and in collecting and inputting the data. Those working with large existing databases, however, often looked with envy at others who could 'control' what was being collected (i.e. avoiding much of the pain involved in compiling, cleaning, and managing very large secondary databases) and who would also typically end up with much smaller self-collected data samples that made the analysis process (once the data was there) much faster and simpler to complete. There was never any accepted resolution to this debate. From my own perspective, I can remember the 'joy' of the day I first received my data on a series of computer tapes. I felt my problems were over and that I could simply begin analysing this data right away in order to rapidly complete my thesis. If only life could be that straightforward!

Table 3.1 outlines the extent of the final database on which this thesis is based:

Table 3.1: Extent of the Thesis Database

Covers 5 weeks (BARB weeks 16-20 1993) on the main UK channels (BBC1, BBC2, ITV, Channel 4, and S4C)

Panel Data broken down by	Total Number
Individual respondents	3,918
Discrete programme showings	8,176
Unique programme titles *	1,469
Audience appreciation responses	548,834

* NOTE: Many titles will have had multiple showings or episodes over this five week period

In addition to providing an audience appreciation score (on the six point rating scale described earlier), each of the almost 550,000 individual responses in this database provides many fields of detailed information on the specific viewer and on the particular programme being evaluated. BARB's panel member information, for example, includes fields outlining such details as the respondent's sex (2 categories), age (6 categories), and social class (4 categories; AB, C1, C2, and DE)³. BARB's programme information then includes further fields listing the programme title and details on its broadcast date, day of the week, time, channel, and sequence (if the same title occurs

³ Demographic groupings listed for sex and age in the tables of this thesis are self explanatory. For social class, however, it is necessary to be familiar with the UK's standard social grade classifications. In arriving at its four social class categorisation, BARB has used the following six social grades but has combined grades A and B into a single AB social class and grades D and E into a single DE social class:

Social Grade	Social Status	Occupation
A	Upper middle class	Higher managerial, administrative, professional
B	Middle class	Intermediate managerial, administrative or professional
C1	Lower middle class	Supervisory or clerical, and junior managerial
C2	Skilled working class	Skilled manual workers
D	Working class	Semi and unskilled manual workers
E	Those at lowest level of subsistence	State pensioners or widows (no other earner), casual or lowest-grade workers.

more than once in a given day - e.g. a news title). BARB also classifies each programme title into one of a variety of programme categories or genres.

3.6 ADDITIONAL VARIABLES DEVELOPED FOR THIS RESEARCH

Supplementing the raw data outlined above, further variables were developed, derived, or adapted for use in this research:

(a) Behavioural Classifications of Individuals:

In addition to BARB's demographic classifications for panel members, individuals have been classified in this thesis along behavioural lines according to their weight of overall television viewing (i.e. Heavy, Average and Light viewers of television); to their average overall appreciation scoring patterns (i.e. High, Moderate and Low Scorers); and to the consistency with which they make use of various points on the audience appreciation scale (i.e. Consistent, Typical, and Varied users of the six point appreciation scoring scale). The development of these classifications will be described in detail in Chapter 4 ("Individual Viewers' Appreciation Scoring Patterns").

(b) Classification of Programme Types:

Within the Television Opinion Panel data being analysed, BARB has classified each programme into one of 29 principal programme types. Using further sub-categories (e.g. there are 45 possible types of sports), there are a total of 88 programme genre classifications. For the purposes of this thesis, and in line with its replication oriented nature (at least initially), this categorisation

structure is unwieldy. Therefore, a more workable classification, with a total of eight broad programme types, was adapted from earlier research in this area and is used throughout this thesis:

Table 3.2: Classification of Programme Types
Reclassification of the 29 principle BARB programme categories into three broad programme categories and eight programme types

Broad Category	Programme Type	Incorporates BARB Categories
Entertainment	Light Entertainment	<ul style="list-style-type: none"> • Variety • Situation Comedy • Other Comedy • Chat Shows • Quiz Shows and Panel Games • Cartoons / Animation • Family Shows • Contemporary Music Programmes
	Light Drama	<ul style="list-style-type: none"> • Long Running UK Drama Series • Long Running Non-UK Drama Series • Other Drama Serials/Series UK • Other Drama Serials/Series Non-UK
	Films	<ul style="list-style-type: none"> • Films Made for Cinema • Films Made for TV
	Sport	<ul style="list-style-type: none"> • Sports
Demanding	Drama/Arts	<ul style="list-style-type: none"> • Drama Single Plays • Arts Programmes • Special Events • Classical Music Programmes
	Information	<ul style="list-style-type: none"> • Documentaries and Features • Current Affairs: Political/Econ/Social • Current Affairs: Consumer Affairs • Current Affairs: Special Events • Hobbies / Leisure Pursuits
	News	<ul style="list-style-type: none"> • National News • Regional News
Other	Miscellaneous	<ul style="list-style-type: none"> • Children's Programmes • Religious Programmes • Party Political Broadcasts

This classification system was originally developed, with input from the BBC Broadcast Research Department, in a research monograph from the mid-1980's (Ehrenberg 1986). It involves recombining BARB's 29 principle programme types by broadly classifying each type as "demanding" or "entertaining" (with an additional "miscellaneous" category for a few programme types with unique audience appreciation patterns). The prior usage of this classification system will be discussed later in this thesis (especially in Chapter 6 "Programme Appreciation and Audience Size"). It reflects, however, one main outcome of the Ehrenberg Group's stream of research into patterns of audience appreciation; the definition of a term they called "demandingness". Such "demandingness" is determined by whether a programme "helped me to relax" or "made me think". Given this definition, programmes could be classified in two broad categories, labelled "demanding" and "entertainment". Demanding programmes tended to include news, current affairs, documentaries, and heavy drama, arts, and cultural programmes. Entertainment programmes tended to include light entertainment programming, light drama, films and sports programmes.

(c) Calculation of Programme Appreciation Scores:

As briefly outlined earlier, a programme's appreciation score (or AI score in Broadcasters' Audience Research Board terminology) is the average appreciation score it receives across all individuals who viewed the programme. This score can be calculated for each programme by summing the scores for all individual responses to that programme and then dividing this total by the total number of respondents.

In theory, AI's could be available for all programmes. For statistical reasons, however, such programme appreciation scores should only be calculated for programmes that attract at least a minimum number of individual responses. The statistical cut-off used in this thesis is the same as that used by BARB. Before calculating an overall average appreciation score for any particular programme showing, that programme must have had an audience of at least 25 or more individual panel members who provided audience appreciation responses. Given the nature of this diary research vehicle and the overall size of the Television Opinion Panel, this means that many programmes with small audiences do not receive an AI score. Of the 8,176 programmes shown in Table 3.1 that were actually broadcast, only 3,015 programmes will have met the minimum 25 response criteria from the 3,918 individual panel members who could possibly have been watching. Programmes for which appreciation scores are not calculated are most likely to be ones broadcast in off peak hours or on the smaller audience channels (e.g. BBC2 and Channel 4).

An appreciation score cannot be calculated for *all* programmes broadcast. In order to achieve sufficient (i.e. at least 25) responses for even the smallest audience programmes, this would require an extremely large (and therefore expensive) viewer panel with members numbering in the tens of thousands. While the statistical cut-off used in this thesis means that many of the smallest audience programmes are not included in the different analyses, this should not impact on the general patterns being researched here. As will be demonstrated, such patterns are mostly not all that startling or dramatic and there is little reason or theory to suggest why similar relationships should not continue to hold across very small audience programmes. In future research, the possibility remains that such small programmes may prove to be more distinctive. But small audience programmes are *small* - does this mean they

are also relatively unimportant? Also, if small programmes are in some way distinctive, then it may be that new research into their viewers' appreciation will require the development of specific new scales.

(d) Calculation of Audience Size Ratings:

Chapter 7 ("Programme Appreciation and Audience Size") examines the relationship between programme appreciation scores and programme audience size ratings. The definition of the audience size variable used here is based on traditional percentage (i.e. share of audience) rating points. Ratings therefore represent the number of 'viewers' (i.e. panel members who responded) for each programme as a proportion of the total viewing population (e.g. the number of members in the viewing panel that week). The audience size ratings in this thesis have therefore been calculated directly from the raw audience appreciation panel data based on the number of people giving a score to a particular programme. The fact that some other form of audience size ratings has not been used should not be seen as a weakness of this analysis. There is evidence that audience size estimates derived from BARB's Television Opinion Panel are credible and highly comparable to the standard television industry ratings derived from BARB's electronic peplemeters (Wober 1993).

3.7 SAMPLING FRAMES FOR THIS RESEARCH

The national Television Opinion Panel was continuously maintained by BARB to provide an "achieved" (reflecting the actual UK population) net sample size of 3,000 respondents per week. Up to 3,500 recruitment letters were therefore mailed out per month. Not all of this activity was necessary to

compensate for panel attrition. The panel maintenance process was also concerned with modifying the panel balance and with “decommissioning” panel members who failed to respond regularly, or who had reached the end of their two year panel membership. Reflecting this panel maintenance process and the fact that not all panel members will respond by filling in a diary each and every week, Table 3.3 illustrates the realities of panel member response rates over the five week period represented by this data sample.

Table 3.3: Breakdown of Individuals in the Panel According to Number of Diary Booklets Returned During this Five Week Period

Number of Weeks in which Diaries Returned	Number of Individuals
5 out of 5 weeks	1,888
4 out of 5 weeks	936
3 out of 5 weeks	525
2 out of 5 weeks	310
1 out of 5 weeks	259
Total	3,918

Where appropriate, the analyses in this thesis will make use of all the data collected over this period (e.g. from all 3,918 individuals as summarised above). In calculating, average *programme* appreciation scores, for example, it is important to utilise all available data in order to achieve sufficient responses for as many programmes as possible (given the 25 minimum response criteria outlined above). Certain *individual viewer* level analyses, however, are concerned with viewers’ behaviour patterns longitudinally over the course of this time period (e.g. repeat viewing analyses). In such analyses, and also elsewhere where there are no concerns about being able to achieve sufficient sample sizes, the data utilised will be restricted to those responses

collected from the 1,888 individual panel members who filled in a diary in each and every week of this five week period. All of the tables and analyses to follow in this thesis will clearly outline the particular sampling frame on which any patterns or conclusions have been based.

Table 3.4 provides demographic breakdowns for all 3,918 individuals who returned Television Opinion Panel diaries in at least one week of this five week period, and for the 1,888 individuals who returned diaries in each and every week during this period. BARB has designed the Television Opinion Panel to be representative of the UK population. When analyses in this thesis are then based on the smaller sample of individuals who returned diaries in each and every week, the main difference to note is with respect to age: older individuals are somewhat more likely to consistently appear in the sample from week to week over this five week period. In spite of these differences, however, various sensitivity checks carried out through the course of this research (exploring variations in the results that would be achieved using one or the other of these two still-very-large sample frames) have shown that the choice of a sample typically introduced very little or no bias to any of the main findings or conclusions.

Table 3.4: Demographic Breakdown of Final Thesis Database
5 Weeks, April 19/1993 through to May 17/1993

Panel Members Who Responded	Sex		Age						Social Class			
	M	F	12- 15	16- 24	25- 34	35- 44	45- 54	55+	AB	C1	C2	DE
In at least one week (n=3,918)	% 48	52	8	13	18	17	14	30	18	25	28	28
Each & every week (n=1,888)	% 49	51	4	10	17	19	14	36	17	26	27	30

3.8 SUMMARY AND SEQUENCE OF ANALYSES

This chapter has outlined the research philosophy underlying this thesis and provided a detailed description of the data that is now to be analysed over the next five chapters.⁴

As explained, the data reduction approach followed here is inductive in that it begins with empirical data analysis rather than formal theory. Much of this research is based on prior expectations (therefore it is not simple ‘data-dredging’), if not on formal hypotheses and ‘theoretical-then-empirical’ work. In many cases it is known from past research that something is there (even though that ‘something’ may not be a strong or definitively established prior result). For example, limited relationships have been demonstrated in the past between audience appreciation ratings and audience size ratings. Similarly there is the suggestion from prior research of a strong relationship between audience appreciation and viewers’ claimed frequency of viewing.

Given such prior expectations, analyses here should not necessarily therefore be based on standard statistical tests with null hypotheses that the relationships or patterns outlined are different from *chance*. Instead, do the

⁴**Technical Note:** Given its data intensive nature, it is important to outline here the significant technical and computer related decisions that had to be taken early in the data analysis process. In late 1992 and early 1993, the data sample under consideration here would have been considered extremely extensive (requiring roughly 500 megabytes of storage space for all the main data and analysis files). At the time, London Business School did not have networked computer facilities and PhD students were faced with very limited allocations of storage space on the School’s main frame computer. This led to an important trade-off decision, reflecting circumstances at the time but with resulting long term implications throughout the course of this research. Essentially, computer processing speed was sacrificed in exchange for increased memory and direct control. The analyses presented in this thesis were programmed and run on a (now already quite ‘ancient’ and painfully slow in comparison to current machines) Macintosh Colour Classic personal computer (16 megahertz, 68030 chip) with 10 megabytes of Random Access Memory and 750 megabytes of hard-drive storage space. The main statistical software package used throughout was STATISTICA for the Macintosh (currently version 4.1). This is a comprehensive and highly rated software package offering similar capabilities to the more well-known SPSS package.

patterns demonstrated here agree/disagree with the (non-zero) expectations, and crucially (as has already been stressed) do they hold up and generalise across varied subsamples (e.g. across different viewing weeks, across older and younger viewers, etc.)? Structured tables, comparing different conditions and telling a specific story (i.e. ‘hypothesis’) are the foundation of most of this thesis. Such tables follow a number of simple presentation rules such as using drastic rounding throughout, ordering the rows of a table by size, using row and column averages as foci, and developing layouts that should guide the eye and facilitate comparisons. These tables then provide rich data summaries that “can be readily used, interpreted and communicated” in light of both prior expectations and new conjectures (Ehrenberg 1991; Collins 1992; Ehrenberg 1992).

While this thesis is very much centred on the data reduction approach outlined above, there are a few cases where other statistical modelling techniques have also been incorporated for specific defined purposes. For example, Chapter 4 (“Individual Viewers’ Appreciation Scoring Patterns”) includes a sub-section that makes use of CHAID, a statistical segmentation modelling approach, for ‘storytelling’ purposes in order to provide an alternative perspective on some patterns already discussed. Similarly, Chapter 7 (“Programme Appreciation and Audience Size”) makes considerable use of simple linear regression techniques in order to replicate and generalise from similar prior research in this area.

Each of the following five ‘analysis’ chapters should be seen as quite self-contained, with each including additional detail where necessary concerning essential prior literature and various methodological issues. The chapters do, however, build on each other so their ordering is important. Chapter 4 (“Individual Viewers’ Appreciation Scoring Patterns”) first summarises

patterns in the ways different groups of *individuals* can vary in their overall audience appreciation scoring patterns. Chapter 5 (“Programme Appreciation Scores”) follows but is quite separate from Chapter 4. Here, basic overall patterns are summarised for how different *programmes* vary on average, and within various programme types, in terms of the aggregate programme appreciation scores they achieve when the scores of all individual viewers are summed and averaged.

Chapters 6, 7 and 8 must follow on from the previous two chapters since these three chapters then integrate and combine the findings of Chapters 4 and 5 in various ways. The actual ordering of Chapters 6, 7 and 8 is less clear-cut, however, and a few circular references will by necessity occur when, for example, an analysis in Chapter 6 might need to refer briefly to findings yet to come in Chapter 7. (As I believe would be typical of most academic research, analyses here were not necessarily carried out in the clear sequential order that seems to emerge in a structured final write-up. Indeed, multiple elements of this research programme were actually being carried out simultaneously and in close co-ordination with each other.)

Chapter 6 (“Programme Appreciation and Audience Composition”) looks at the general overall findings outlined in Chapter 5 by breaking analyses down in terms of the audience make-up of different programmes (e.g. the average audience composition attracted by different programme types and by programmes on different channels). In this sense, how do the individual level patterns outlined in Chapter 4 relate to the aggregated programme level patterns outlined in Chapter 5? Furthermore, are there specific programmes where an examination of programme appreciation scores given *by specific audience segments* (e.g. the ‘male’ programme appreciation score for a

programme would be the average of all individual male responses to the programme) might show exceptional patterns in line with such segments' average overall individual scoring patterns?

Chapter 7 ("Programme Appreciation and Audience Size") concentrates primarily on the analysis of programme level data (i.e. programme appreciation scores as summarised in Chapter 5) in order to establish the extent of any relationship between such programme appreciation ratings and more traditional ratings of audience size. Again, however, certain stages of this analysis must incorporate audience related factors (i.e. the individual appreciation scoring patterns outlined in Chapter 4) since such factors can influence programme appreciation.

Finally Chapter 8 ("Audience Appreciation and Repeat Viewing") incorporates both individual level data, as is represented in Chapter 4, and programme level aggregated data, as is represented in Chapter 5. How do *individuals* vary in their repeat viewing habits and does the appreciation score they give to a regular programme in one week relate to their frequency of viewing the programme in subsequent weeks? Do higher average *programme* appreciation scores then relate to higher average repeat viewing levels for such programmes?

Each of these chapters will conclude with a fairly brief summary. A broader summary overview and discussion of the findings established in Chapters 4 through 8 will then follow in Chapter 9 ("Discussion and Conclusion").

4. INDIVIDUAL VIEWERS' APPRECIATION SCORING PATTERNS

4.1 CHAPTER FOUR OVERVIEW

This chapter focuses on how television viewers vary in their audience appreciation scoring patterns. This is done by analysing response patterns and scoring levels across different groupings and categorisations of individuals. (This contrasts with Chapter 5 which will then focus on patterns of aggregated audience appreciation scores across different groupings and categories of television *programmes*.) Firstly, a variety of tables reconfirm and summarise already established individual level patterns in this type of audience appreciation data (e.g. how the six point appreciation scale tends to be used overall, and how scoring patterns vary within demographic viewer breakdowns). Several additional behavioural classification systems are then developed for use in categorising viewers further according to their viewing behaviour patterns (e.g. heavy or light viewers) or to their appreciation scoring patterns (e.g. viewers who tend to give high scores overall as opposed to viewers who tend to give lower scores). Overall, the tables presented in this chapter provide much of the groundwork for subsequent chapters, outlining the main dimensions on which individuals will be segmented and classified, and outlining basic audience appreciation patterns in the ways viewers differ along these dimensions.

4.2 VIEWERS LIKE WHAT THEY WATCH

Table 4.1 summarises almost 550,000 individual appreciation scoring responses to television programmes, dramatically illustrating the already well-

established pattern that “viewers mostly say that they quite like what they watch and they watch what they say they quite like” (Goodhardt, Ehrenberg and Collins 1975; Barwise and Ehrenberg 1988). As mentioned in Chapter 3 (“Methodology”), the six point BARB audience appreciation rating scale is asymmetric, with three positive options, two negative options, and one neutral option. Even with such an asymmetric scale, more than ninety percent of responses remain at the positive end of the scale (i.e. for analysis purposes this is largely a three or four point scale - the seeming ‘sameness’ of this data has already been commented on in Chapter 3 and it will be raised again under “Limitations” in Chapter 9). Such a pattern makes intuitive sense. People have a choice over what to view so why would they watch programmes that they do not find even the slightest bit enjoyable or interesting?

Table 4.1: The Range of Appreciation Scores

BARB Weeks 16-20 1993, N=548,834 individual responses

Point on Scale	Score	% of Responses
6	100	20
5	80	38
4	60	33
3	40	6
2	20	2
1	0	1
		100 %

4.3 DEMOGRAPHIC PATTERNS

Table 4.2 replicates and reconfirms earlier studies which have looked at demographic variations in programme appreciation scores. Here, the almost 550,000 individual responses have first been summed and averaged for each individual, and then averaged and broken down demographically across the

3,696 individuals in this five week sample who responded to a total of at least 25 different programme showings. The table shows that women give slightly higher scores on average than men, and that older viewers give slightly higher scores on average than younger viewers. (Note that the 12-15 age group is separated out here since this youngest group's scoring patterns are quite different. This pattern will be discussed and allowed for in later analyses.) Similarly, viewers from the lower income C2DE social classes tend to give slightly higher scores on average than do ABC1 viewers.

Table 4.2: Demographic Variations in Appreciation

5 Weeks, BARB Weeks 16-20 1993
N = 3,696 Individuals

	Number of Individuals	% of Individuals	Average Score
Male	1,784	48	70
Female	1,912	52	74
Age 12-15	288	8	77
Age 16-24	476	13	71
Age 25-34	640	17	70
Age 35-44	639	17	70
Age 45-54	510	14	71
Age 55+	1,143	31	74
Social Class AB	649	18	70
Social Class C1	933	25	71
Social Class C2	1,045	28	73
Social Class DE	1,069	29	74
Overall Average	3 696	100	72

Although the differences in average appreciation scores between these demographic segments are small, they are “highly significant” statistically (i.e. they exist in large samples and recur in different subsets or ‘samples’ of the

data). More especially, they are consistent with the results of independent past research (Gunter and Wober 1992).

4.4 BEHAVIOURAL CLASSIFICATIONS OF VIEWERS

In addition to the demographic classifications used above, individuals can also be categorised in terms of their actual viewing behaviour and in terms of their audience appreciation scoring behaviour. Three such categorisation systems are developed and presented here for use in analyses throughout subsequent chapters. In terms of *actual viewing behaviour* over the five week period of analysis, individuals are placed in one of three weight of viewing categories. In terms of *audience appreciation scoring behaviour*, individuals are then similarly categorised according to the average level of appreciation scores they give across all programmes viewed, and according to how consistent or varied they are in their use of particular points on the six point “interesting and/or enjoyable” rating scale.

For the purposes of developing and illustrating these behavioural classifications, the following tables present data only for those individuals in the BARB panel who returned diaries in each of the five weeks under consideration (see Table 2.2 in Chapter 2). This allows for additional checks on the robustness of the various classification systems (e.g. are “average viewers” generally “average viewers” from week to week, or would they be classified as “heavy viewers” in some weeks and “light viewers” in other weeks?). Individuals are first classified according to their viewing or scoring patterns across the entire five week period. This overall classification can then be compared to how consistently individuals would be similarly classified on a week by week basis.

The individual patterns to be outlined here are interesting in their own right, but they become especially relevant and important when considering patterns in *programme* appreciation scores (e.g. in Chapters 5 and 6). As will be outlined, for example, heavy viewers of television tend to give higher appreciation scores on average than do lighter viewers. Does one programme therefore tend to achieve a higher average appreciation score than another programme because it is inherently liked more by its viewers, or does this higher score simply reflect the fact that one programme might have a higher proportion of heavy television viewers in its audience?

(a) Light, Average, and Heavy Viewers of Television:

The BARB Television Opinion Panel data being analysed does include a variable for each individual panel member's weight of television viewing. Based on questioning at the time of recruitment into the panel, individual members are classified as being in one of three weight of viewing categories: light (0 to 17 hours of viewing per week), medium (18 to 31 hours), and heavy (more than 32 hours). Once identified as a light, medium, or heavy viewer, this categorisation may or may not reflect each individual's subsequent viewing behaviour over the full period in which he or she continues as a panel member. It was therefore decided to develop new weight of viewing categories based on actual viewing behaviour during the period under analysis.

Table 4.3 divides respondents into categories based on their average weight of viewing over the five week period under analysis. For replication purposes (i.e. in order to be able to compare results found in this thesis to earlier findings), this system is very much equivalent to the BARB categories

outlined above and used in some previous research. (Gunter and Wober 1992). Here, however, due to the data available, weight of viewing is measured in terms of *number of programmes* viewed rather than *hours* viewed per week. "Light" viewers have been classified as those who watched (i.e. gave an appreciation score to) 100 or less programmes over this 5 week period (i.e. about 20 or less programmes in a week). Those who viewed 200 or more programmes over this five week period are "Heavy" viewers (i.e. more than about 40 programmes in a week), while all other individuals have been classified as "Average" viewers. The actual average number of programmes per week viewed here by light viewers was 14; by average viewers 29, and by heavy viewers 60.

Table 4.3: Light, Average and Heavy Viewers of Television

Individuals' Weight of Viewing across this five week period
N = 341,252 Responses by 1,888 individuals

Weight of Viewing	Number of Individuals	% of Individuals	Average Score
Light (<100 *)	473	25	69
Average	750	40	71
Heavy (>200 *)	665	35	74
Total	1,888	100	72

* number of programmes viewed over this five week period

To check the robustness of this classification system, individuals' weight of viewing patterns were also classified on a weekly basis to compare to this five week overall classification. On average, individuals' weekly weight of viewing classifications matched their overall five week classification in 4.2 out of the five weeks. More than 75% of individuals could be very consistently categorised by viewing weight from week to week (e.g. they would be placed in the same weight of viewing category in at least four of the five weeks under consideration).

Overall, a clear pattern emerges in this table. Heavier viewers of television tend to be somewhat more generous on average in their audience appreciation scoring patterns than are lighter viewers. This provides a strong replication and reconfirmation of similar patterns that have been demonstrated in past research but were then based on smaller and less comprehensive data samples (Gunter and Wober 1992).

The weight of viewing classifications used here were not chosen haphazardly, nor was there a simple acceptance that something close to the accepted BARB weight of viewing classification would be 'best'. Instead, I also experimented (i.e. conducted sensitivity analyses) with a variety of other possible alternative classification systems. In first examining the quite normally distributed frequency distributions for viewers' weight of viewing in this sample, it became clear that there were no natural 'dividing lines' to provide obvious breaks between what one would classify as light, average, or heavy viewers. One option therefore was to divide the entire sample into three equal-sized (i.e. 33%) weight of viewing categories. Another option was to classify light viewers as those more than one standard deviation below the "mean" viewer, and heavy viewers as those more than one standard deviation above. Both these procedures would however lose their meaning when reapplied to different populations (e.g. older and younger viewers). A final option (the one chosen and applied in Table 4.3) was then to use a system that communicates well and would be easy and appropriate to reapply in any future research. In any case, this system gives something fairly close to a 33% division.

Sensitivity analyses at the time, using each of the above possible classification systems, did not lead to any radically varied findings (i.e. mostly quite similar average scoring patterns were found for light, average, and heavy viewers,

regardless of the classification system used, and the key main finding was highly consistent: heavier viewers of television tended to be somewhat more generous on average in their audience appreciation scoring patterns than were lighter viewers). The first option outlined above led to three equal-sized viewing categories, while the standard deviation classification method led to a larger number of average viewers (i.e. 68%) and smaller categories of light and heavy viewers (16% each). The final option led to three roughly but not quite equal sized categories (i.e. 25%, 39% and 35% as shown in Table 4.3). This chosen classification system (i.e. “less than 20 programmes” for light viewers and “more than 40 programmes” for heavy viewers) is easy to understand, however, and does not depend on the weight of viewing distribution and number of individuals in just this one particular data sample. This also makes it a simple system to reuse in classifying any new data samples (i.e. in replication or future research), therefore fitting in well with the general principles of data reduction and generalisability that underlie this research programme.

(b) Low, Moderate, and High Appreciation Scorers:

Using a similarly straightforward categorisation process, for most analyses individual panel members were classified in terms of their average levels of appreciation across all programmes each viewed over the five week viewing period under consideration. (A good deal of analyses of the more detailed data at the time showed that this produced no distortions in the results.)

In Table 4.4, “Low Scorers” are those individuals whose average level of appreciation was 65 or lower across those programmes for which they provided a response. “High Scorers” are those whose average level of appreciation for the programmes they viewed across this five week viewing

period was 80 or higher. All other individuals are classified as “Moderate Scorers”. As is evidenced in this table, individuals can vary significantly in the way they utilise the six point appreciation scale. High scorers naturally tend to consider the average programme viewed as being “very” or even “extremely” interesting and/or enjoyable (i.e they mostly tend to use points 5 and 6 on the six-point audience appreciation scale, resulting in an average appreciation score of 87 over all programmes viewed), while low scorers are less generous and tend to on average make more use of the “fairly” interesting and/or enjoyable evaluation (i.e. they use mostly points 4 and 5, but also make use of even lower scores, resulting in a much lower average appreciation score of 59 over all programmes viewed).

Table 4.4: Low, Moderate, and High Appreciation Scorers

N = 341,252 Responses by 1,888 individuals

Liking Category	Number of Individuals	% of Individuals	Average Score
Low Scorers (<60 *)	593	31	59
Moderate Scorers	805	43	72
High Scorers (>80*)	490	26	87
Overall	1,888	100	72

* cutoff points

Again, to further check the robustness, sensitivity, and consistency of this classification system, individuals had also been classified on a weekly basis in terms of these scoring patterns to compare to this five week overall classification. This showed that individuals could be very consistently categorised by their general scoring patterns from week to week.

In keeping with the initial replication oriented nature of this research programme, the starting point for developing this classification system derived directly from past research which classified viewers as “high likers” and “low likers”, using data from a precursor to the present BARB audience appreciation collection system (Wober 1976). As with the weight of viewing classifications outlined earlier, various cut-off points between what I would now term high, moderate and low scorers were then explored (i.e. sensitivity analysis). This was done to determine the relative sizes (i.e. numbers of individuals) of each of the resulting categories, as well as to determine how far such variations in cut-off points would lead to any dramatically different patterns in the average appreciation scores of the resulting high, moderate, and low scorer categories. With no evidence of such dramatic variation when using cut-off points ranging between 63 and 67 for low scorers and 83 to 87 for high scorers, the decision was taken to settle again on a classification system with more memorable numbers (65 and 85) that communicate well and can be easily reapplied to new data.

(c) Consistent, Regular, and Varied Appreciation Scorers:

Table 4.5 explores differences between those panel members who have quite consistent scoring patterns (i.e. they tend to almost exclusively use only a single point on the six point “interesting and/or enjoyable” scale) and those individuals who have more varied scoring patterns (i.e. they tend to allocate a wider range of scores to the different programmes they view).

In this table, a “Consistent Scorer” is an individual who has given exactly the same score to 80% or more of the programmes he or she has viewed over this

five week period. (On average, such consistent scorers have allocated an identical score to fully 90% of all programmes they viewed.) A “Regular Scorer” (the majority of average viewers) is one who has almost exclusively used only two points on the six-point scoring scale, allocating one or the other of these two scores to 80% or more of the programmes he or she has viewed. (On average, such regular scorers have used only two selected points on the scoring scale for 90% of all programmes viewed. For almost all of these individuals, these ‘most-used’ scores represent two adjoining points on the six-point scoring scale.) Finally, a “Varied Scorer” is an individual who has routinely allocated a wider range of three or more possible scores to the various different programmes he or she has viewed.

Table 4.5 shows that individuals classified as “Consistent Scorers” had an average level of appreciation of 78 (averaged for each individual across all programmes viewed over this 5 week period). “Regular” and “Varied” scorers show lower average levels of appreciation. Such patterns are an artifact of this categorisation process. As stated earlier, the overall pattern of individual appreciation scoring responses is that “viewers like what they watch”. Consistent scorers therefore tend to give consistently positive scores. Indeed, almost a third of these consistent individuals are those who almost exclusively gave scores of 100 to the programmes they viewed. In making somewhat fuller use of the six point appreciation scale, regular scorers and especially varied scorers are simply more likely to occasionally give a programme a neutral or negative score. This will bring average scores down for these viewer categories.

Table 4.5: Consistent, Regular, and Varied Appreciation Scorers
 Variation in the Range of the 6 Point Scoring Scale Used by Individuals
 N = 341,252 Responses by 1,888 individuals

Category	Number of Individuals	% of Individuals	Average Score
Consistent Scorers	306	16	78
Regular Scorers	1102	59	72
Varied Scorers	480	25	67
Overall	1,888	100	72

Note that this type of classification and analysis has not been carried out in any previous research on audience appreciation ratings. The definitions used for consistent, regular and varied scorers are therefore the result of much exploration into different possible ways of classifying the 'consistency' of viewers' appreciation scoring patterns. For example, it would be impractical to label a "Consistent Scorer" as only one who *always* gives the same score to all programmes viewed. There are very few such viewers. The key to developing this classification system, therefore, was to begin by looking at "average" appreciation scoring patterns (i.e. the majority of viewers can be classified as "regular" in that they typically and almost exclusively use only two points on the six point rating scale). Definitions for the remaining two categories (i.e. where to draw the dividing lines) could then be set by looking at how far individuals' audience appreciation scoring patterns varied at either extreme from the majority of "regular" viewers.

(d) Composition of the Different Behavioural Categories:

How do these different behavioural classifications relate to each other and who (in demographic terms) are the individuals they classify? For example,

what proportion of individuals who are classified as heavy viewers are also classified as high scorers? The following three cross-tabulation tables help to identify relations between the various behaviourally segmented audience groupings.

Table 4.6 looks at the individual composition of the light, average, and heavy viewer segments across the various demographic segments as well as across the two other behaviour-based segments (i.e. average scoring levels and scoring consistency). The table shows that there are demographic patterns with respect to weight of viewing. A somewhat higher proportion of women than men are heavy viewers. Panel members in the oldest age group (55+) are much more likely to be heavy viewers than are viewers in other age groups (once again the youngest 12-15 children's age group provides an exception to the general pattern and has been separated out in this table). Finally, viewers in the lower income social groups are also on average more likely to be heavy viewers than are viewers in higher income social groups.

In addition to such demographic patterns, there are clear behavioural patterns apparent in Table 4.6. For example, individuals who are heavy viewers of television are relatively more likely to also be high average appreciation scorers. Interestingly, this may imply that light viewers do not only watch less television than heavy viewers, but that they are also slightly more critical on average in evaluating those programmes they do choose to view. Alternatively, heavy viewers might simply be indiscriminate viewers who are more likely to 'enjoy everything they watch'. (The first of these two interpretations appears to be closer to the truth, as will be discussed in the next paragraph.) Such patterns are clear and consistent, but they are not all that strong. Across the 1,888 individuals being looked at here, the actual

correlation between individuals' weight of viewing (the number of programmes viewed over the five week period by each individual) and individuals' average scoring patterns (the average score given by each individual across all programmes viewed) is $r = 0.16$. (Note that this correlation is statistically highly significant. But even more importantly, a similar r greater than 0.1 arose in each of the five weeks. That cannot have happened by chance, i.e. this is an 'empirical' significance test.)

Table 4.6: Composition of Light/Average/Heavy Viewers

1,888 Individuals who returned Television Opinion Panel diaries each week during this five week period

			Light Viewers	Average Viewers	Heavy Viewers
No. of Individuals:	Total: 1,888		473	750	665
% of Individuals	Total: 100%	%	25	40	35
Sex:	Male	%	29	40	31
	Female	%	21	40	39
Age:	12-15	%	26	46	28
	16-24	%	31	49	20
	25-34	%	33	41	26
	35-44	%	33	40	27
	45-54	%	30	41	29
	55+	%	13	35	52
Social Class:	AB	%	37	40	23
	C1	%	34	42	24
	C2	%	22	42	36
	DE	%	14	35	51
High/Low Scorers	Low	%	30	41	29
	Moderate	%	27	41	32
	High	%	16	36	48
Consistent/Varied Scorers	Consistent	%	15	32	53
	Regular	%	24	42	34
	Varied	%	34	39	27

It is interesting to note that quite a high proportion (more than 50%) of viewers with consistent appreciation scoring patterns tend to also be heavy viewers. One might have expected that the reason an individual was classified as “consistent” was that he or she had only viewed a small number of programmes and had tended to give most of them the same score. Instead, it appears that many of these consistent scoring heavy viewers are watching a great deal of programming yet are giving most of these many programmes viewed an identical appreciation rating.

As outlined earlier, light viewers not only watch less programming but they also tend on average to be more ‘critical’ (i.e. they give lower programme appreciation scores on average than do heavy viewers). This can now at least partially be explained in that light viewers are shown to be somewhat more likely to make more use of the full range of the audience appreciation scale. This means that as a proportion of their total number of responses to those programmes viewed over this period, light viewers on average give a slightly higher proportion of relatively low scores.

What might explain the patterns outlined in the previous two paragraphs (i.e. that heavy viewers have a tendency to be more consistent in their scoring patterns while light viewers tend to be somewhat more varied)? Perhaps these patterns exist because, as outlined earlier, individuals choose to watch very few programmes that they do not like and this number does not necessarily need to grow proportionately with the total number of programmes viewed (i.e. light viewers may in effect watch a slightly higher percentage of programmes that they do not like, and therefore make somewhat more use of lower audience appreciation scores). Alternatively, these patterns may be evidence of a problem with the diary collection system. Heavy viewers have a

more time-consuming task in that they have to report scores for a larger number of programmes each week. Perhaps such viewers might therefore have more of a tendency to simplify (and in effect invalidate) this task by getting into the habit of repeatedly marking an identical score for most programmes viewed. If this were so (and it could only be proven through new primary research), then it would have real implications with respect to the reliability and accuracy of this data and its related ability to discriminate between programmes.

Table 4.7 now looks at the individual composition of the low, moderate, and high appreciation scorer segments. Here, there are similar but perhaps less pronounced patterns to those found in Table 4.6 for weight of viewing. A somewhat higher proportion of women than men tend to be high appreciation scorers. Likewise, a higher proportion of individuals in the older age groups tend to be high scorers when compared to individuals in the younger age groups (again with the exception of the 12-15 age group). The social class pattern is less pronounced than it was for weight of viewing, but again it suggests that a higher proportion of individuals in the lower income social groups will be high scorers.

Although the demographic patterns outlined in these two tables may appear quite similar, the "weight of viewing" section of Table 4.7 again demonstrates that high (or low) scorers and heavy (or light) viewers are by no means the same individuals. There is simply a tendency for a slightly higher proportion of low appreciation scorers to be light viewers of television, and for a slightly higher proportion of high scorers to be heavy viewers of television.

Table 4.7: Composition of Low/Moderate/High Appreciation Scorers

1,888 Individuals who returned Television Opinion Panel diaries each week during this five week period

			Low Scorers	Moderate Scorers	High Scorers
No. of Individuals:	Total: 1,888		593	805	490
% of Individuals	Total: 100%	%	31	43	26
Sex:	Male	%	37	42	21
	Female	%	26	43	31
Age:	12-15	%	17	49	34
	16-24	%	35	46	19
	25-34	%	41	40	19
	35-44	%	34	39	27
	45-54	%	36	39	25
	55+	%	24	46	30
Social Class:	AB	%	35	46	19
	C1	%	32	46	22
	C2	%	29	43	28
	DE	%	31	37	32
Weight of Viewing	Light	%	37	46	17
	Average	%	33	44	23
	Heavy	%	26	38	36
Consistent/Varied Scorers	Consistent	%	33	20	47
	Regular	%	26	67	17
	Varied	%	43	43	14

A stronger pattern is apparent between individuals' average overall scoring patterns and the consistency with which they use the six point appreciation scale. As discussed in Section 4.5 (c), however, this is likely an artifact of the categorisation process used. As shown in the Table 4.7, only 14% of varied scorers will be classified as high scorers with average overall appreciation scoring levels of 80 or higher. In contrast, 47% of individuals with consistent scoring patterns will be classified as high scorers.

Table 4.8 now looks at the demographic composition of viewers classified by their consistent, regular or varied appreciation scoring patterns.

Table 4.8: Composition of Consistent/Regular/Varied Appreciation Scorers

1,888 Individuals who returned Television Opinion Panel diaries each week during this five week period

			Consistent Scorers	Regular Scorers	Varied Scorers
No. of Individuals:	Total: 1,888		306	1,102	480
% of Individuals	Total: 100%	%	16	58	26
Sex:	Male	%	14	58	28
	Female	%	18	59	23
Age:	12-15	%	22	62	16
	16-24	%	11	51	38
	25-34	%	7	60	33
	35-44	%	14	60	26
	45-54	%	18	53	29
	55+	%	22	60	18
Social Class:	AB	%	11	60	29
	C1	%	13	57	30
	C2	%	16	58	26
	DE	%	23	59	18
Weight of Viewing	Light	%	10	56	34
	Average	%	13	62	25
	Heavy	%	24	57	19
High/Low Scorers	Low	%	17	48	35
	Moderate	%	8	67	25
	High	%	29	58	13

The patterns here suggest that a higher proportion of female viewers than male viewers will be consistent scorers. A higher proportion of older than younger viewers are similarly more likely to be consistent scorers (once again with the

exception of the 12-15 age group) and a higher proportion of viewers in the lower income social classes will have consistent scoring patterns. As already discussed under Table 4.6, light viewers of television are somewhat more likely to have varied scoring patterns while heavy viewers are somewhat more likely to have consistent scoring patterns. Finally, as also discussed earlier, Table 4.8 again evidences the artifact that a relatively higher proportion of high appreciation scorers will have consistent scoring patterns, while a relatively higher proportion of low scorers will make more varied use of the appreciation scoring scale.

4.5 A CHAID SEGMENTATION MODELLING ANALYSIS

As has been outlined in Chapter 3 (“Methodology”), this thesis is predicated on principles of data reduction. This approach is well-founded and reflects the data analysis philosophy of my supervisor. Nonetheless, it can lead to ‘non-typical’ PhD theses. As has been explained, this thesis differs from the more ‘standard’ thesis in marketing in that it represents a programme of research rather than a single study; replication is given a central role; statistical jargon is minimised (unless appropriate and necessary); and the data reduction analysis process, although very labour intensive, is conceptually quite ‘simple’ to understand. For the most part, this research is based on developing summary tables that structure, order, and ‘reduce’ the original data so as to discover and display patterns that may exist within this data and which generalise. In this sense, this research does not make extensive use of what some might term “advanced” or “black-box” statistical techniques that tend to summarise or represent the original data in the form of mathematical formulas or graphs. While I was comfortable and confident with this approach, I did at times become concerned that I should also be demonstrating a familiarity and ability with other possible methodologies and approaches.

This section of Chapter 4, through its use of advanced segmentation modelling software (CHAID - to be described in detail below), provides a brief exception to the general use of data reduction principles in this research programme. This came about from my own initiative, as my supervisor felt (quite correctly) that I could successfully complete my entire intended research programme through the consistent continued use of data reduction techniques. He did not object to me trying out other methods though, and the fact that I decided to do so here reflects the process I was going through in developing as a new researcher. In moving through the coursework component of the London Business School PhD Programme, I was of course exposed to many widely used multivariate data analysis techniques. In speaking with other students, I also became aware of a variety of even more 'exotic' methodologies that different students were utilising in their own research. I therefore resolved that, at an appropriate point in my research, I would at least explore the use of a suitable alternative analysis technique. This would enable me to demonstrate to myself, and to others, that I was capable of using such 'advanced' research methods. It would also give me the chance to compare and contrast the usefulness of such a technique with the data reduction approach.

(a) What is CHAID?

In this research programme, the appropriate opportunity to explore an alternative technique arose when I began considering the demographic and behavioural variations in individual scoring patterns that have been discussed so far in this chapter. As already outlined, these patterns will provide a basis for understanding and interpreting results in later chapters. At this point, a question that arose was that of which of the variables might contribute *most* to the prediction of individuals' average overall appreciation scoring levels.

To explore this, I began investigating different approaches to market segmentation and learned of CHAID (Chi-Squared Automatic Interaction Detector). CHAID is a tree-based classification procedure that detects interactions among a large set of nominal/ordinal scaled variables. The technique itself is relatively complex, but the output is in the form of an apparently clear visual tree diagram. The method goes back to the early sixties with the introduction of AID (Automatic Interaction Detector) and the later development of THAID (Theta Automatic Interaction Detector) (Kass 1980; Magidson 1993; Magidson 1995; Chaturvedi and Green 1995). Today, there are three different software packages available for tree classification applications - CHAID, TREE, and CART (Chaturvedi and Green 1995). I used CHAID because it is included as a module in the SPSS statistical package.

CHAID works by dividing a population into subgroups, using a categorical dependent variable as a criterion for forming these subgroups (e.g. whether an individual is a low, moderate, or high scorer). The CHAID branching process begins by first dividing the population into two or more distinct groups by choosing the 'best' predictor variable of whether individuals are low, moderate, or high appreciation scorers. (The categorical predictor variables available here are age, gender, social class, social class, weight of viewing, and scoring consistency). To determine the single 'best' predictor (which here turns out to be age), CHAID splits the tree on the predictor that has the lowest probability value, or p-value of having occurred just by chance, as long as the p-value is less than the significance level. (The CHAID programme's default significance level, as used in this research, is 0.05.)

Following this initial split or "branching", the CHAID process continues by splitting each of these groups into smaller subgroups based on other predictor

variables, until no more statistically significant predictors can be found. P-values are also used in this branching process to merge any predictor variables that do not discriminate between the categories of the dependent variable. (Magidson 1995; Chaturvedi and Green 1995).

(b) Results of a CHAID analysis:

Figure 4.1 displays the results of the CHAID analysis for the 3,696 viewer sample reflected in Table 4.2¹. This tree diagram shows the splitting process as the overall sample of viewers is split into 13 final subgroups which differ in their audience appreciation scoring patterns. The first split in this tree indicates that age is the most important predictor of an individual's appreciation scoring patterns. While 27% of the overall viewer population here are high appreciation scorers, for example, some 42% of the 12-15 age group are high scorers.

The second series of splits on this tree diagram indicates how different predictors are more important for different age groups. For example, gender is a more important predictor of older viewers' scoring patterns while weight of viewing is a more important predictor for the youngest age group.

We can note how CHAID merges categories of a predictor variable that are not significantly different. Although there are six possible age categories in this data, for example, the first level split by age shows that the 16-24 and 25-34 age groups have been merged together (i.e. not split). This is because these

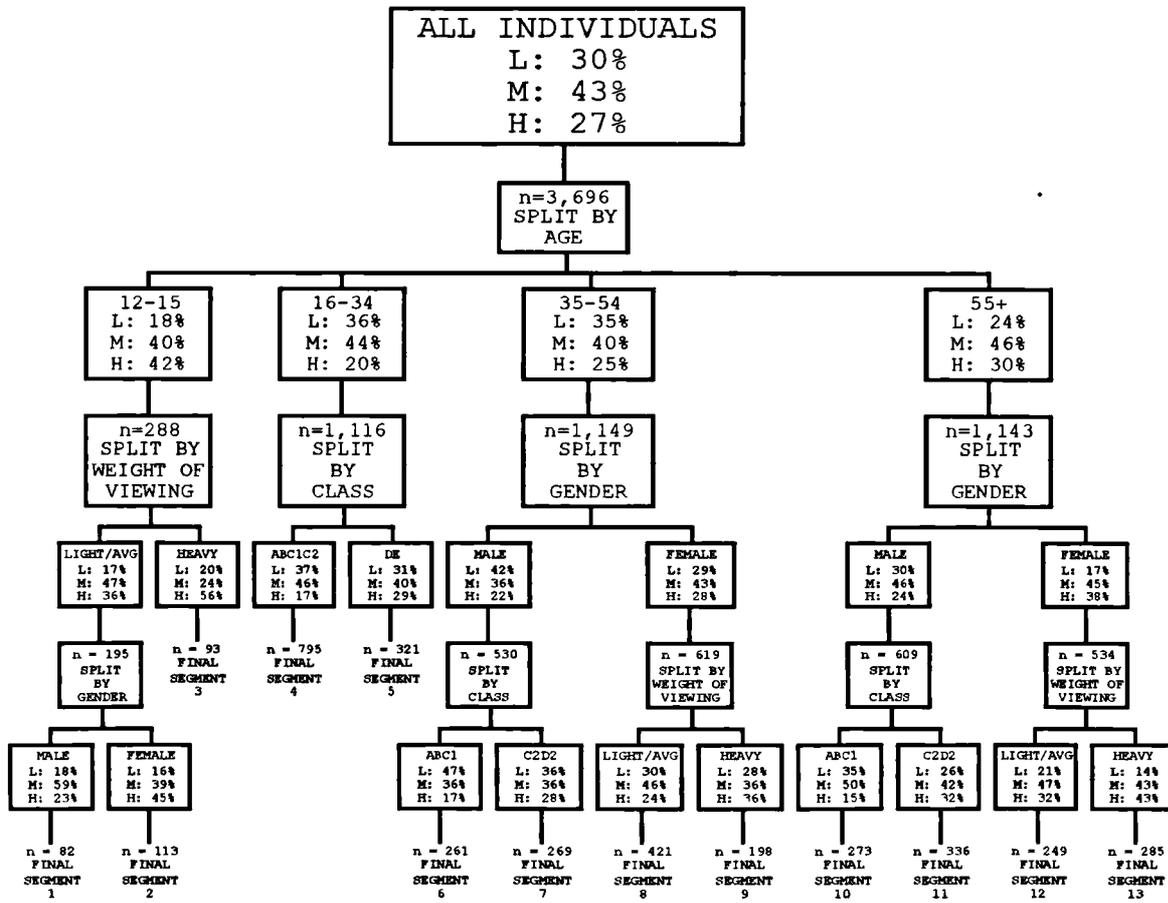
¹ To maximise sample sizes, this analysis uses the full 3,696 individual sample and not the 1,888 individual sub-sample which was previously used to develop and illustrate the various viewer categories in Section 4.4. This explains why the proportions of low, medium and high appreciation scorers in the "All Individuals" box of Figure 4.1 (e.g. 30%, 43% and 27%) differ slightly from the figures outlined in Table 4.4 (e.g. 31%, 43% and 26%).

two age groups do not differ statistically in their scoring behaviour. Similar merging has occurred for the 35-44 and 45-54 age groups.

Figure 4.1: Tree Diagram for Viewer Segmentation by Appreciation Scoring Patterns

n = 3,696 individuals who provided scores for 25 or more programmes during this five week period

L = % of individuals who are low appreciation scorers
 M = % of individuals who are moderate appreciation scorers
 H = % of individuals who are high appreciation scorers



The final 'segments' that CHAID derives are mutually exclusive and exhaustive. They do not overlap and each individual is contained in exactly one final segment. Since each segment is defined by a combination of

predictor variables, one could easily classify (or 'predict') each individual into his or her appropriate segment simply by knowing this combination of categories (e.g. final segment 9 in Figure 4.1 will include all females, aged 35-54, who are heavy viewers).

So what is the 'story' told by this CHAID analysis? Essentially the figures to compare are the percentages of low, moderate, and high scorers in each box on this chart. These confirm many of the patterns already outlined in this chapter. For example, for the first split in this figure by age, a higher proportion of older viewers than younger viewers are high scorers (with the exception of the youngest 12-15 age group as has already been demonstrated before). While 30% of the 55+ age group are high scorers, consistently only 20 to 25% of the 16 to 54 age groups are high scorers. Similarly, in later splits in this diagram, other of the already established patterns become apparent. For example, when female viewers aged 55+ are split by weight of viewing into final segments 12 and 13, it can be seen that a higher proportion of heavy viewers (43%) are high scorers than are light or average viewers (32%).

The next question is which specific audience segments can be identified as the most and the least generous in terms of their audience appreciation scoring patterns? The CHAID output does not highlight these segments as such. They would have to be identified through what in market research is often called 'eyeballing'. In Figure 4.1, the most generous overall appreciation scorers are final Segments 3 (heavy viewers aged 12-15) and 13 (female heavy viewers aged 55+). Some 56% of the individuals in Segment 3 are high appreciation scorers and 80% are classified as either moderate or high scorers. Within final Segment 13, 43% of individuals are high scorers and 86% of viewers are either moderate or high scorers. The least generous overall appreciation scorers would then appear to be final Segments 6 (male ABC1 viewers aged 35-54)

and 4 (ABC1C2 viewers aged 16 to 34). As many as 47% of individuals in Segment 6 are classified as low scorers, and 83% are either low or moderate scorers. Within Segment 4, 37% of individuals are low scorers and again a total of 83% are either low or moderate scorers. It is however not easy to see what more Figure 4.1 “says”.

Table 4.9 therefore re-presents the patterns identified above rather more clearly. The thirteen final Segments from Figure 4.1 are here profiled and arranged in order of their average appreciation scores.

Table 4.9: CHAID Segment Profiles In Decreasing Order by Segment Members' Average Overall Appreciation Scores

n = 3,696 individuals who provided scores for 25 or more programmes during this five week period

CHAID Final Segment	n Indivs	Avg Score	Sex		Age Group "1" = Age12 to 15 "6" = Age 55+						Social Class				Weight of Viewing				
			M	F	1	2	3	4	5	6	AB	C1	C2	DE	L	A	H		
3	93	79	M	F	1								AB	C1	C2	DE			H
13	285	78		F									AB	C1	C2	DE			H
2	113	78		F	1								AB	C1	C2	DE	L	A	
12	249	75		F									AB	C1	C2	DE	L	A	
1	82	74	M		1								AB	C1	C2	DE	L	A	
11	336	74	M												C2	DE	L	A	H
9	198	74		F				4	5				AB	C1	C2	DE			H
8	421	72		F				4	5				AB	C1	C2	DE	L	A	
5	321	72	M	F				4	5							DE	L	A	H
10	273	70	M										AB	C1			L	A	H
7	269	70	M					4	5						C2	DE	L	A	H
4	795	70	M	F			2	3					AB	C1	C2		L	A	H
6	261	68	M					4	5				AB	C1			L	A	H

As in Figure 4.1, the most discriminating predictor of appreciation scoring patterns throughout Table 4.9 is age. Individuals in the four highest scoring Segments are either in the oldest age group (aged 55+) or in the youngest age group (aged 12-15, which has repeatedly been shown, in Sections 4.3 and 4.4 of this chapter, to have unique appreciation scoring patterns). Individuals in the four lowest scoring Segments are, with the exception of Segment 10, in the four young to middle aged age groups (i.e. aged 16 through to 55).

Gender would be the next best predictor in Table 4.9 of whether individuals would be classified as high or low scorers. Thus, with the exception of Segment 3, which includes both men and women, the four highest scoring Segments are exclusively comprised of women. Similarly, with the exception of Segment 4, which includes both men and women, the four lowest scoring Segments are exclusively comprised of men.

Social Class and Weight of Viewing, although among the 'best' here, are less effective predictors of whether individuals are classified as high or low scorers. Social Class does not help distinguish the four highest scoring Segments at all (i.e. the first four lines in Table 4.9 include individuals from all social groups). For the four lowest scoring Segments, however, social class does play a role. Individuals in these low scoring Segments are likely to be from the higher income social groups. The opposite pattern is seen for weight of viewing as a predictor. The two highest scoring Segments are comprised exclusively of heavy television viewers, while the next two highest scoring Segments are light and average viewers of television. For the four lowest scoring Segments, however, weight of viewing does not play a distinguishing role (i.e. these Segments include individuals from all weight of viewing categories).

Overall, what has this exploratory use of CHAID segmentation analysis added to the general demographic and behavioural appreciation scoring patterns that have already been outlined in sections 4.3 and 4.4 of this chapter? Essentially, the value of this CHAID analysis lies in its simultaneous consideration of the various individual segment patterns outlined earlier in this chapter. The approach has identified which segmentation criteria are the best predictors of whether individuals are classified as high or low appreciation scorers. Most of the differences are slight (as shown rather clearly by the Segments' average AIs in Table 4.9) and in line with earlier results.

Furthermore, in combining these various segmentation criteria, it is possible to identify exceptions to the general rules established earlier in this chapter and in previous research. For example, in segmenting audiences unidimensionally by age, a higher proportion of viewers in the oldest age group (55+) will on average will be classified as high scorers. In Figure 4.1 and Table 4.9, however, CHAID final Segment 10 provides an exception to this general rule. It seems that aged 55+ *male* viewers in the higher income (ABC1) social classes have quite different (i.e. much lower on average) appreciation scoring patterns from all other individuals in this same age group (women and C2D2 men). I note that the CHAID analysis does not seem to help one evaluate all of this further.

(c) Comparing CHAID to the Data Reduction Approach:

CHAID is a visually and conceptually 'interesting' technique, especially from an exploratory research perspective (as will be outlined below). It has been utilised in this thesis to demonstrate an awareness of the potential use of other possible 'advanced' approaches for analysing audience appreciation data. As discussed, however, many of the findings of this CHAID segmentation

modelling exercise are readily apparent in earlier analyses using the data reduction approach (e.g. Table 4.7 describes the general composition of low, moderate and high appreciation scorers). Other patterns and exceptions that have been outlined could also have been derived through a more time-consuming continuation of the data reduction approach.

Perhaps the relative merits of these two segmentation analysis methodologies can best be illustrated, therefore, by explicitly contrasting 'my' approach - the data reduction approach - to the CHAID approach outlined above. In utilising the data reduction approach to complete a segmentation modelling exercise, I would proceed as follows:

1. I would first analyse the appreciation scoring levels in this data by each of the separate main effects (i.e. for gender, age, social class, and weight of viewing). This is essentially what was done early in this chapter.
2. Then, I would begin to look for at least some of the interactions (e.g. does the breakdown by weight of viewing in Table 4.3 then vary between the young and the old?).
3. As has repeatedly been outlined, I would not go much into significance because with large sample sizes most of these results are significant (i.e. they would be expected to recur with other such samples).
4. I would then check whether each such result is generalisable. For example, if women are higher appreciation scorers on average, do these same patterns occur separately in each of Weeks 1, 2, 3, 4 and 5 and do they generalise with the findings or other research? Similarly, do the interactions also generalise?

5. This data reduction approach involves the analyst picking out the large differences (and the 'no differences'), by systematically inspecting the data. Which is the 'biggest' relationship between demographics/weight of viewing and audience appreciation scoring patterns? Which is the next biggest (and so on) and which are the negligible ones? This is what 'eyeballing' should be - it is quite a deliberate process. These results could then be checked for interactions.
6. However, I would then check the results of such eyeballing in Week 1, or for men, for consistency against other data (e.g. Week 2, for women, etc.). An initial subjective-sounding process is therefore systematised.

The above approach can be contrasted (on a number by number basis) with the CHAID approach as follows:

1. CHAID concentrates on interactions from the start. It only gives *one* main effect in full (e.g. the first 'branch' - here for age). That can be very limiting - even the analyst (with the results from a straight CHAID analysis) does not know all of the main effects!
2. After this first branch, CHAID does look at interactions but it only looks at *some* ('most significant') of the interactions, as determined by its special selection process. Ordinary CHAID never looks (i.e. normally prints out or directly inspects) any of the *other* interactions. The irregular sequence of factors produced (the 'tree') is difficult to take in or 'eyeball'. And it is often very difficult to compare different CHAID analyses (e.g. UK versus US, or a year later, or in a sensitivity analysis as in looking at Week 5 only).

3. CHAID goes just by whether breakdowns of each kind are 'significant' (not even explicitly the 'largest'). This depends on (a) within-sample variability, (b) sample sizes, and (c) the actual size of the differences - which is what ultimately matters to scientists and practitioners.
4. CHAID, and other more familiar statistical techniques, are attractive to statisticians because of their 'simplicity' in the sense that a great deal of data can be edited and summarised relatively quickly through the application of 'automatic' rules in the 'black box' of a software package. The 'automatic' CHAID approach quickly produces a chart from which certain conclusions can be drawn. In contrast, the tabulating and interpreting even of just all the first order interactions is much more work in the data reduction approach. (Tables 4.2 through 4.5, for example, involved a painstaking process of repeatedly sorting the data, performing multiple sum and average tabulations, developing tables, and then interpreting and comparing all the resulting patterns.)
5. CHAID automates a particular selection of 'segments', but the selection process and the results are, in a deep sense, obscure. For example, the use of automatic cut-off rules, irrespective of the particular data set under analysis, raises the issue of 'how good is the best predictor?' At the various levels of this tree-structured analysis, the 'best' predictor determines each branch that is taken. But what if the next best predictor is almost or virtually as good (e.g. differing by only 1 or 2 percentage points in how well it predicts the categorical dependent variable)? Age may indeed be the 'best' overall predictor of whether an individual will be classified as a high or low audience appreciation scorer, but how much better a predictor is age in fact than gender? Could quite a different tree diagram emerge that summarises this data

virtually as effectively as the one leading to the 13 final segments outlined in Figure 4.1? (A similar “how good is best” issue will be raised in Chapter 7 - “Programme Appreciation and Audience Size” - in line with the use of standard simple linear regression tools).

In fact, with the relatively simple analysis presented in Figure 4.1, this is not really an issue as there is little scope for a dramatically different tree to emerge. More complex applications of CHAID, however, could conceivably involve large numbers of predictor variables, where these predictors and the dependent variable could each contain up to 31 possible categories! (This is the limit allowed by the SPSS software.)

Such “how good is the best predictor?” issues could then be partially addressed through sensitivity analyses using the CHAID programme’s manual mode, where the user gets to choose the predictor variable that will drive each split. Further sensitivity analyses could explore the variation in trees that might emerge under different significance levels. Such analyses might however quickly become quite complex and cumbersome, and much more so than the data reduction tabulations. Although it is possible to work interactively with this software, a key attraction of all classification tree techniques is the fact that they can be highly automated. It is well recognised that most users will simply “turn the software loose on the data” and then trust and depend on the built in ‘artificial intelligence’ of the CHAID programme to do a competent analysis (O’Brien and Durfee 1994).

6. The results in CHAID still need ‘interpretation’ (i.e. eyeballing). But my opinion now is that they are more difficult to understand and summarise than is the initial data. Moreover, it can be difficult to check the

quantitative CHAID results for consistency and generalisability against other data because not only the number of branches, but the branching sequence will tend to vary from case to case.

Overall, these approaches are two ways of looking at some of the effects of (here) gender, age, class, and weight of viewing on individuals' audience appreciation scoring behaviour. It would be surprising if each of these approaches failed to bring out any big factors in the data somehow. But CHAID does not, in itself, tell one in any straightforward way if there are only main effects (and what they are). Still, if this thesis had not been based on initial replication, and if there were no previous findings or 'expectations' to build on, then CHAID would have been a useful mechanised technique for partial early exploratory data analysis of main effects and interactions.

I would therefore recommend CHAID to other PhD students (with large data samples defined in suitable categorical variable terms), as an approach to try at a relatively early stage in their data analysis process. The technique is formalistic, but its 'automatic' nature means that it can provide a quick and simple (but highly selective or biased) means to reduce large volumes of data into a visual (but still not necessarily always easy to interpret) summary diagram. But I would never suggest relying only on a simple first-run CHAID analysis. In this thesis, CHAID was not utilised as a main data analysis approach and its exploratory use here has not added any significant additional insight into general patterns already outlined earlier.

(d) Comparing CHAID to other segmentation approaches:

Before leaving CHAID, it is worth briefly considering other more traditional approaches to market segmentation. How do methods such as Cluster

Analysis and Discriminant Analysis differ from CHAID and why were they inappropriate for use in the current analysis?:

- Traditional cluster analysis can also be used to find mutually exclusive groups of individuals who are relatively similar to each other (i.e. how people group 'naturally' on a number of characteristics such as demographics, weight of viewing, and audience appreciation scoring behaviour). Having determined these clusters, one could then further investigate how far these resulting groups *actually* differ from each other in a particular way (e.g. in their appreciation scoring behaviour). The problem is that the groups may not differ in this particular way, since the various clustering procedures require no a priori specification of independent and dependent variables. CHAID does require such prior specification and then seeks to group observations in a manner that minimises the variance of the dependent variable within each group. Since CHAID divides the population using the categorical dependent variable (i.e. the three categories of high, moderate, and low appreciation scorers), one knows up front that the segments from the resulting classification tree segmentation will differ on this *appreciation scoring behaviour* (Chaturvedi and Green 1995).

The outcome of a cluster analysis usually involves judgements. The decision on how many and which segments to choose for a final solution will depend on subjective factors such as how interpretable and statistically differentiated the groups are. The resulting cluster solution cannot be easily used subsequently to classify a new data sample into similar clusters. With CHAID, however, segments are more clearly defined (by combinations of the predictor variables) and they can therefore be used in principle to later apply classifications to a new sample (Magidson

1995). For example, one segment resulting from a CHAID analysis may be defined as including all older (age 55+) females who are heavy television viewers. It would be simple to reapply this classification to appropriate cases from a new data sample. (This would be a key advantage in terms of the replication oriented research philosophy of this thesis.)

- Unlike with cluster analysis, in discriminant analysis one *begins* with predefined 'segments' or groups. The technique essentially then aims to establish whether these groups differ from each other. (Actually, a common use of discriminant analysis in market research is for profiling the clusters identified from a cluster analysis.) For any one measure (e.g. height), the means of the different groups (e.g. different sets of patients from different hospitals) may differ. However, there may also be much overlap between the groups. Discriminant analysis therefore aims to combine a number of measurements (e.g. height, weight, age, etc.) as a weighted average to discriminate 'best' between the different groups (Ehrenberg 1991).

The discriminant analysis approach is used to understand group differences and to predict the likelihood that a new item will belong to a particular group. In this regard, discriminant analysis is like regression analysis except that the dependent variable consists of two or more separate categories (i.e. low, moderate, or high appreciation scorers). Also like regression analysis, however, discriminant analysis cannot make use of categorical independent variables (i.e. the age, gender, social class, and weight of viewing variables) and so the approach was not suitable for use here.

4.6 PATTERNS ACROSS PROGRAMME TYPES AND CHANNELS

This final analysis section of Chapter 4 provides a bridge to help begin framing the analyses to come in the next three chapters.

Firstly, without even considering audience appreciation patterns, the tables to follow examine how different segments of individuals allocate their viewing across different programme types and across the different channels. The findings reconfirm and illustrate a pattern that has been repeatedly established through research by the Ehrenberg Group (Ehrenberg 1986; Barwise and Ehrenberg 1988) but that nonetheless remains surprising to many. On average, across possible audience segments (e.g. by sex, age, and weight of viewing) the proportions of time allocated to the different programme types and channels on offer is much the same.

Secondly, once such viewing behaviour patterns have been illustrated, further tables outline how these different segments of individuals might vary in the audience appreciation scores they give to the different programme types and to the different channels. Again, with a few minor exceptions, there is little evidence of specific audience segments that particularly 'like' certain programme types or channels.

(a) Individuals' Viewing of Programme Types and Channels

Table 4.10 uses the programme type classification system outlined in Chapter 3 ("Methodology") to examine how different segments of individuals allocate their viewing across the different programme types.

Table 4.10: Individuals' Viewing of Different Programme Types

Aggregated Viewing Patterns of 1,888 Individuals
(Representing a total of 341,252 responses), BARB Weeks 16-20 1993

Average Percentage of Viewing spent on Different Programme Types, with
Individuals Segmented into Demographic (Sex, Age, Social Class), Weight of
Viewing, and Behavioural (Low/High Scorers, and Consistent/Varied Scorers)
Subgroups

	Entertaining				Demanding			Other	Total	
	Light Ent	Light Drama	Films	Sport	Drama Arts	Info/Docu	News	Misc	Ent	Dmd
All Adults	% 19	31	5	6	1	16	19	3	% 61	36
Male	% 19	26	6	9	1	16	20	3	% 60	37
Female	% 19	36	4	3	1	15	18	4	% 62	34
Age 12-15	% 22	40	5	5	1	10	4	13	% 72	15
Age 16-24	% 23	39	6	6	1	12	7	5	% 74	20
Age 25-34	% 20	35	6	5	1	16	13	4	% 66	30
Age 35-44	% 20	34	5	5	1	16	16	3	% 64	33
Age 45-54	% 18	29	5	7	1	17	21	2	% 59	39
Age 55+	% 18	25	3	7	1	16	27	3	% 53	44
Class AB	% 19	28	4	6	2	17	21	3	% 57	40
Class C1	% 19	31	5	7	1	16	19	2	% 62	36
Class C2	% 20	32	5	6	1	15	17	4	% 63	33
Class DE	% 20	32	4	6	1	14	19	4	% 62	34
Light	% 21	30	6	6	2	18	14	3	% 63	34
Medium	% 19	33	5	6	1	15	18	3	% 63	34
Heavy	% 19	29	4	6	1	15	22	4	% 58	38
Low	% 18	31	5	7	1	16	19	3	% 60	36
Moderate	% 20	31	5	6	1	15	18	4	% 62	34
High	% 20	33	4	5	1	15	18	4	% 62	34
Consist	% 18	32	4	6	1	14	22	3	% 60	37
Typical	% 19	32	5	6	1	15	18	4	% 62	34
Varied	% 20	30	6	6	1	18	16	3	% 62	35

Specific individuals can of course vary greatly in their viewing patterns, but Table 4.10 shows that when individual patterns are aggregated across a wide variety of possible segments, there is a very real lack of dramatic differences in the percentages of viewing time spent by each segment on the different types of programmes broadcast. This table may seem complex but the idea is to look down each column comparing the % figures shown for various audience segments to the “all adults” figure at the top of the column. With few exceptions, there is very little scatter in each column about the “all adults” figures. This is most powerfully illustrated in the “Total” columns for entertainment (“Ent”) and demanding (“Dmd”) programmes. Glancing down each of these two columns, only a few figures (i.e. primarily those for different age groups) differ by more than one or two percentage points from the “all adults” figure. Across most possible audience segments (e.g. by various demographic and behavioural criteria), approximately 61% of programmes viewed by the average individual are entertainment programmes while approximately 36% are more demanding programmes. (Note that the entertainment and demanding percentages do not sum to 100 since miscellaneous programmes are excluded from these classifications.)

As mentioned, the only strong exception to this general pattern is seen when audiences are segmented by age. Approximately 40% or more of all programmes watched by viewers in the two oldest age groups are more demanding (i.e. news and information) programmes. In contrast, for the two youngest age groups, well over 70% of all programmes viewed are entertainment programmes and only 20% or less of all programmes viewed are more demanding programmes. (For the youngest 12-15 age group, this pattern would be even more striking if the 13% figure in the miscellaneous column were added to the 72% figure in the entertainment column. This miscellaneous

column includes many children's entertainment programmes.) Breaking these broad entertainment and demanding categories down, it appears that older viewers watch a much higher proportion of "News" programmes while younger viewers tend to spend more time watching "Light Entertainment" and "Light Drama" programming.

In segmenting viewers by other criteria, differences in programme type viewing patterns are less noticeable. There are a few systematic exceptions, such as with women spending slightly more of their viewing on light drama or soap operas (e.g. 36% versus 26%) and with men spending slightly more of their viewing on sports and films. Similarly, heavy viewers view a higher proportion of news programmes than do light viewers (e.g. 22% versus 14%). But even these differences are not all that dramatic in relation to the total time people spend viewing.

Table 4.11 now carries out a similar analysis to examine how different segments of individuals allocate their viewing across the four channels (rather than across different programme types). Again, this table reconfirms patterns of viewing behaviour that have been repeatedly established by the Ehrenberg Group and can also be readily demonstrated using BARB's weekly audience size ratings (Barwise and Ehrenberg 1987). Across all audience segments, most individuals spend most of their viewing time (i.e. an average of just under 80%) on programmes broadcast on BBC1 and ITV. Looking down the columns for these two 'big' channels, and comparing the % figures shown for various audience segments to the 'all adults' figure at the top of the column, there is very little scatter among the different figures. Although these differences are not all that dramatic, it appears that women, people in the older age groups, and people in the lower income social classes would devote slightly more of their viewing to ITV.

Table 4.11: Individuals' Viewing of Different Channels

Aggregated Viewing Patterns of 1,888 Individuals
(Representing a total of 341,252 responses) 5 Weeks, BARB Weeks 16-20
1993

Average Percentage of Viewing spent on Different Channels, with Individuals Segmented into Demographic (Sex, Age, Social Class), Weight of Viewing, and Behavioural (Low/High Scorers, and Consistent/Varied Scorers) Subgroups

		Channel			
		BBC1	BBC2	ITV	C4
All Adults	%	37	11	41	11
Male	%	36	13	39	12
Female	%	38	8	43	11
Age 12-15	%	40	8	38	14
Age 16-24	%	36	10	38	16
Age 25-34	%	34	10	42	14
Age 35-44	%	35	11	42	12
Age 45-54	%	39	11	41	9
Age 55+	%	39	11	41	9
Class AB	%	42	13	34	11
Class C1	%	39	12	38	11
Class C2	%	36	10	42	12
Class DE	%	34	9	46	11
Light	%	38	16	33	13
Medium	%	38	11	40	11
Heavy	%	35	8	47	10
Low	%	38	11	40	11
Moderate	%	37	11	40	12
High	%	37	9	43	11
Consist	%	38	8	45	9
Typical	%	38	10	41	11
Varied	%	36	13	38	13

Across all audience segments, on average just over 20% of the programmes most individuals watch will be broadcast on the two smaller channels, BBC2 and Channel 4. Those viewers who spend relatively higher proportions of their viewing time on these two channels include men and light viewers of television. Interestingly, younger viewers tend to allocate a higher proportion (i.e. about 15%) of their total viewing to Channel 4 than do older viewers (i.e. about 9%), while BBC2 attracts a slightly higher proportion of older individuals' viewing (i.e. 11%) than of younger individuals' viewing (i.e. 9%). As with the programme type differences illustrated in Table 4.10, Table 4.11 suggests that different segments of individuals do allocate their viewing somewhat differently across the different channels. Again, however, these differences are not all that dramatic in relation to the total time people spend viewing television.

(b) Individuals' Appreciation of Programme Types and Channels

As has also been noted above in line with how viewers on average allocate their time, it is important to acknowledge that individuals can vary greatly in their appreciation scoring patterns. Tables 4.12 through to 4.15 now show, however, that when such individual patterns are aggregated across a wide variety of segments, there are few exceptions to the general scoring patterns already outlined within this chapter. For example, it has been shown that women tend on average to give slightly higher scores than men, and that older viewers tend to give slightly higher scores than younger viewers (again with the exception of the youngest 12-15 age group). With very few exceptions, these and the other individual scoring patterns (i.e. by social class, by weight of viewing, etc.) hold quite consistently across and within programmes of very different types and across and within the various channels.

Table 4.12: Individuals' Average Appreciation of Different Programme Types

Average scores of 1,888 individuals over 5 weeks, BARB Weeks 16-20 1993
By Demographic (Sex, Age, Social Class), Weight of Viewing, and Behavioural
(Low/High Scorers and Consistent/Varied Scorers) Subgroups

	n Indivs	Entertaining				Demanding			Other	Total		
		Light Ent	Light Drama	Films	Sport	Drama Arts	Info/ Docu	News	Misc	Ent	Dmd	All
Sex												
M	934	70	70	68	72	65	72	66	67	71	69	70
F	954	73	75	72	70	68	76	70	70	74	73	74
Age												
12-15	83	77	76	78	73	69	77	69	71	77	75	77
16-24	187	71	71	71	67	66	72	64	65	71	69	71
25-34	325	68	71	68	66	64	71	62	65	70	68	70
35-44	361	70	72	71	69	66	73	66	69	72	70	70
45-54	262	70	71	70	70	66	73	67	67	71	70	71
55+	670	73	73	71	75	69	76	73	74	74	74	74
Class												
AB	328	70	70	69	69	65	72	67	67	71	70	70
C1	485	70	71	70	70	66	73	66	67	71	70	71
C2	518	72	73	71	71	67	75	69	68	73	72	73
DE	557	72	74	71	73	68	74	70	71	74	72	74
Weight												
Light	473	69	69	69	67	63	72	63	67	70	69	69
Avg	750	71	72	70	71	66	73	66	67	72	70	71
Heavy	665	73	75	72	74	69	75	72	70	75	74	74
Scorers												
Low	593	59	60	60	59	56	61	55	56	60	58	59
Mod	805	72	73	70	72	67	74	68	68	73	71	72
High	490	86	87	84	86	79	88	85	83	87	87	87
Consist												
Consist	306	77	78	76	77	76	78	78	76	78	78	78
Regular	1,102	72	73	71	71	67	74	69	68	73	72	72
Varied	480	67	67	66	67	61	71	60	64	68	66	67
Overall	1,888	71	72	70	71	67	74	68	69	72	71	72

The next chapter ("Programme Appreciation Scores") will show in detail that average *programme* appreciation scores (i.e. when individual appreciation scores are summed and averaged for all viewers who watched each

programme) do vary between programme types and between the different channels. These programme patterns reflect the individual scoring patterns now seen in the "Overall" row of Table 4.12. Individuals tend to give higher appreciation scores to light entertainment, light drama, sports, and information programming. Somewhat lower scores are on average then given to films, drama and arts, and news.

The question of interest here is whether these general scoring patterns tend to hold across all audience segments or whether the different segments tend to each relatively 'like' and 'dislike' different programme types. For example, do women on average always tend to give higher audience appreciation scores than men or are there certain programme types where men might give higher scores than women? Even if women do tend to almost always give higher average appreciation scores than men, is the spread between average female appreciation scores and average male appreciation scores wider for some programme types and narrower for others?

If the general relative scoring patterns for each audience segment outlined in the "All" column of Table 4.12 held consistently for all programme types, then the remainder of this table should be highly predictable simply from the "Overall" row and the "All" column averages. For example, Table 4.13 compares 'predicted' average appreciation scores for male individuals for each programme type to the actual observed average male appreciation scores in the "M" row of Table 4.12.

Table 4.13: Observed and 'Predicted' Average Appreciation by Males of Different Programme Types

BARB Weeks 16-20 1993, n=1,205 programmes
which received a total of 100 or more responses

	Entertaining				Demanding			Other	Total		
	Light Ent	Light Drama	Films	Sport	Drama Arts	Info	News	Misc	Ent	Dmd	Overall
Sex											
M ^o	70	70	68	72	65	72	66	67	71	69	70
M ^p	69	70	68	69	65	72	66	67	70	69	70
M ^o -M ^p	1	0	0	3	0	0	0	0	1	0	
Overall	71	72	70	71	67	74	68	69	72	71	72

Notes: M^o = Observed male average appreciation scores
MP = Predicted male average appreciation scores
"Overall" figures are those from Table 4.12

The predicted scores in Table 4.13 are calculated by considering the fact that the overall average appreciation score given by all 1,888 individuals summarised in the overall row of this table is 72. The observed overall average appreciation score given by the 934 males in row M^o is 70. For each programme type, the predicted scores in Table 4.13 are simply then 2 points less than those observed in the "Overall" row.

Only in the "Sport" column, with a three point difference between observed and predicted, is this predicted score either not exactly the same as the observed score or more than a single point away. This very flat pattern is seen throughout Table 4.12, as outlined in Table 4.14 which repeats this exercise for all audience segments. In Table 4.14, most 'predicted' average individual appreciation scores for the various programme types are no more than one or two points away from the actual observed scores in Table 4.12. Overall, this

indicates that there is very little difference between the various audience segments in terms of which programme types get high and low scores (i.e. looking across individual rows in Table 4.12, all segments tend to give relatively low scores to drama and arts programming and relatively high scores to Information and Documentary programming).

There is little indication in these tables of particular programme types that are dramatically liked or disliked by specific audience segments in relation to those same segments' overall scoring patterns across all programme types. A few minor exceptions are apparent however. One is for sports programming, which is the only case where men tend to give higher average appreciation scores than women. As outlined in Table 4.12, the average male viewer here has given the average sports programme he viewed a score of 72. This compares to the average female viewer giving the average sports programme she viewed an appreciation score of 70. Table 4.14 shows that in line with the overall average scores given by males and females to "All" programmes viewed, these observed average individual scores for sports programmes are 3 points higher than might be expected for men and 3 points lower than might be expected for women.

Looking further down the column for "Sports" in Table 4.14, it appears that younger viewers also tend to give sports programmes lower than expected scores (i.e. observed scores for the youngest three age groups are 3 points lower than might be 'predicted' given these groups' overall scoring patterns), while the oldest 55+ age group tends to give somewhat higher than expected scores (i.e. 2 points higher) to this particular programme type. The only other exception with respect to age in Table 4.14 is for news programmes. Again, this programme type tends to appeal relatively less to the younger age groups (i.e. 2 or 3 point lower average appreciation scores than might be expected

given these groups' overall scoring patterns) and relatively more to the oldest 55+ age group (i.e. a 3 point higher average appreciation score).

Table 4.14: Difference between Observed and 'Predicted' Average Individual Appreciation Scores for Different Programme Types

Average scores given by 1,888 individuals, BARB Weeks 16-20 1993
By Demographic (Sex, Age, Social Class), Weight of Viewing, and Behavioural (Low/High Scorers and Consistent/Varied Scorers) Subgroups

	n	Entertaining				Demanding			Other	Total		
		Light Ent	Light Drama	Films	Sport	Drama Arts	Info/ Docu	News	Misc	Ent	Dmd	All
Sex												
M	934	1	0	0	3	0	0	0	0	1	0	70
F	954	0	1	0	-3	-1	0	0	-1	0	0	74
Age												
12-15	83	1	-1	3	-3	-3	-2	-4	-3	0	-1	77
16-24	187	1	0	2	-3	0	-1	-3	-3	0	-1	71
25-34	325	-1	1	0	-3	-1	-1	-4	-2	0	-1	70
35-44	361	1	2	3	0	1	1	0	2	2	1	70
45-54	262	0	0	1	0	0	0	0	-1	0	0	71
55+	670	0	-1	-1	2	0	0	3	3	0	1	74
Class												
AB	328	1	0	1	0	0	0	1	0	1	1	70
C1	485	0	0	1	0	0	0	-1	-1	0	0	71
C2	518	0	0	0	-1	-1	0	0	-2	0	0	73
DE	557	-1	0	-1	0	-1	-2	0	0	0	-1	74
Weight												
Light	473	1	0	2	-1	-1	1	-2	1	1	1	69
Avg	750	1	1	1	1	0	0	-1	-1	1	0	71
Heavy	665	0	1	0	1	0	-1	2	-1	1	1	74
Scorers												
Low	593	1	1	3	1	2	0	0	0	1	0	59
Mod	805	1	1	0	1	0	0	0	-1	1	0	72
High	490	0	0	-1	0	-3	-1	2	-1	0	1	87
Consist												
Consist	306	0	0	0	0	3	-2	4	1	0	1	78
Regular	1,102	1	1	1	0	0	0	1	-1	1	1	72
Varied	480	1	0	1	1	-1	2	-3	0	1	0	67
Overall	1,888	71	72	70	71	67	74	68	69	72	71	72

Otherwise, the only other slight exceptions in Tables 4.12 and 4.14 are with respect to behavioural segmentation criteria. There is a slight indication that news programmes appeal relatively more than might be expected to heavy viewers than to light and average viewers. Similarly, individuals who are low appreciation scorers on average tend to give relatively higher than expected (i.e. by 2 or 3 points) scores to films and drama/arts programmes, while individuals who are high appreciation scorers on average tend to be relatively less generous to these programme types. Finally, for the three demanding programme types, there are some interesting patterns for consistent and varied appreciation scorers that are difficult to interpret. Consistent appreciation scorers seem to give higher scores than might be expected to drama/arts and news programmes and lower than expected scores to information programmes. The opposite pattern is seen for varied appreciation scorers. As was outlined earlier in this chapter, viewers who are consistent appreciation scorers are slightly more likely to be female, older, lower income, and heavy viewers. None of the patterns for these various other segmentation criteria in Table 4.14 help to explain the unusual patterns here for consistent and varied viewers. The patterns nevertheless appear to cancel each other out in the aggregated demanding (“Dmd”) column for consistent, regular and varied appreciation scores.

Table 4.15 now concludes this chapter by looking at individuals’ average appreciation of different channels. In the “Overall” row of Table 4.15, it can be seen that individuals tend to give higher appreciation scores on average to programmes on BBC2 and Channel 4 (the ‘small’ channels) than to programmes on BBC1 and ITV (the ‘big’ channels). Within the various audience segments, and in relation to the overall average scoring patterns of these different segments, there is again little evidence of particular channels

that are dramatically liked or disliked by specific audience segments in relation to those segments' overall scoring patterns.

Table 4.15: Individuals' Average Appreciation of Different Channels

Average scores of 1,888 individuals over 5 weeks, BARB Weeks 16-20 1993
By Demographic (Sex, Age, Social Class), Weight of Viewing, and Behavioural (Low/High Scorers and Consistent/Varied Scorers) Subgroups

	Channel				
	BBC1	BBC2	ITV	C4	All
Sex					
M	68	74	69	72	70
F	73	76	73	75	74
Age					
12-15	74	78	75	80	77
16-24	67	73	69	74	71
25-34	67	72	68	72	70
35-44	69	74	70	73	70
45-54	69	74	71	73	71
55+	74	76	73	75	74
Class					
AB	68	74	69	72	70
C1	69	74	70	73	71
C2	72	75	72	74	73
DE	72	76	72	75	74
Weight					
Light	68	73	68	71	69
Avg	70	74	71	74	71
Heavy	74	76	74	76	74
Scorers					
Low	57	64	58	62	59
Mod	71	75	71	74	72
High	86	89	87	88	87
Consist					
Consist	78	78	78	78	78
Regular	71	75	71	74	72
Varied	65	72	65	71	67
Overall	71	75	71	74	72

A similar exercise to that shown in Tables 4.13 and 4.14 is presented in Table 4.16 to formally illustrate this relative lack of segmentation in the audience appreciation scores that different viewers give to the different channels.

Table 4.16: Difference between Observed and 'Predicted' Average Individual Appreciation Scores for Different Channels

Average scores of 1,888 individuals over 5 weeks, BARB Weeks 16-20 1993
By Demographic (Sex, Age, Social Class), Weight of Viewing, and Behavioural
(Low/High Scorers and Consistent/Varied Scorers) Subgroups

	Channel				
	BBC1	BBC2	ITV	C4	All
Sex					
M	-1	1	0	0	70
F	0	-1	0	-1	74
Age					
12-15	-2	-2	-1	1	77
16-24	-3	-1	-1	1	71
25-34	-2	-1	-1	0	70
35-44	0	1	1	1	70
45-54	-1	0	1	0	71
55+	1	-1	0	-1	74
Class					
AB	-1	1	0	0	70
C1	-1	0	0	0	71
C2	0	-1	0	-1	73
DE	1	-1	-1	-1	74
Weight					
Light	0	1	0	0	69
Avg	0	0	1	1	71
Heavy	1	-1	1	0	74
Scorers					
Low	-1	2	0	1	59
Mod	0	0	0	0	72
High	0	-1	1	-1	87
Consist					
Consist	1	-3	1	-2	78
Regular	0	0	0	0	72
Varied	-1	2	-1	2	67
Overall	71	75	71	74	72

This table shows only a few exceptions. For example, younger viewers of BBC1 tend to give slightly lower appreciation scores to BBC1 programmes than might be expected from such younger viewers' overall appreciation scoring patterns. Individuals who are in general low scorers tend to be relatively (but slightly) more generous than expected when it comes to programmes viewed on BBC2 and Channel 4, while the reverse is true for individuals who are in general high appreciation scorers. Similar slightly higher than expected patterns for BBC2 and Channel 4 are seen for varied appreciation scorers (who are on average lower appreciation scorers), while slightly lower than expected patterns for BBC2 and Channel 4 are seen for consistent appreciation scorers (who are on average higher appreciation scorers).

4.7 SUMMARY

This chapter has outlined and explored patterns in the way individuals differ in their appreciation scoring response styles. In addition to using demographic criteria, a number of structures have also been developed for use in classifying or segmenting different groups of individuals. The main findings are that:

- Women give higher appreciation scores on average than do men
- Older viewers give higher appreciation scores on average than younger viewers
- Individuals in the lower income social groups give higher appreciation scores on average than individuals in higher income social groups.
- Heavy viewers give higher appreciation scores on average than lighter viewers.

- The most important predictors of individuals' appreciation scoring patterns are age and gender.
- On average across possible audience segments, the proportions of time allocated to the different programme types and channels on offer is much the same.
- With few exceptions, there is little evidence of specific audience segments that particularly 'like' specific programme types or channels.

The patterns and structures outlined here will be referred to and will have important implications for the interpretation of results and findings in subsequent chapters.

5. PROGRAMME APPRECIATION SCORES

5.1 CHAPTER FIVE OVERVIEW

This brief chapter takes the individual television *programme* (rather than the individual *viewer*) as the unit of analysis and focuses on how such programmes vary in the appreciation scores they achieve. As with Chapter 4 (“Individual Viewers’ Appreciation Scoring Patterns”), this is intended to provide a foundation for the remainder of the thesis through replicating, reconfirming and updating past research findings. Programme appreciation scores are first described in general, and then analysed and compared across different possible categories or groupings of programmes (e.g. programmes categorised by programme type or grouped by channel of broadcast).

5.2 PROGRAMMES ARE MOSTLY QUITE LIKED

Chapter 4 summarised individual viewer’s appreciation scoring patterns with the phrase “viewers like what they watch”. As illustrated, most *individual* responses tend to be at the positive end of the six point appreciation scoring scale. In summing and averaging such individual responses to calculate *programme* appreciation scores, it is unsurprising that most programmes tend to be quite liked on average by their viewers. In theory, AI scores for different programmes could range from 0 to 100. In practice, they mostly fall within a fairly narrow range of from about 60 to 80.

Table 5.1 illustrates this pattern for more than 3,000 programmes (i.e. those meeting the minimum 25 respondent criteria) broadcast over the five week

period under analysis in this thesis. The average programme appreciation score is 72. With a standard deviation of 5, fully ninety-five percent of all programmes will achieve an appreciation score of between 62 and 82. This 60 to 80 range has been consistently found in Britain, Canada, the USA, Germany and elsewhere (Gunter and Wober 1992).

Table 5.1: Average Programme Appreciation Score for the Average Programme

BARB Weeks 16-20 1993, n=3,015 programmes
which received a total of 25 or more responses

	Mean Programme Appreciation Score	Standard Deviation
All Programmes	72	5.1

5.3 APPRECIATION FOR DIFFERENT PROGRAMME TYPES

As has been demonstrated in the past, some variation in programme appreciation scores can be found across different programme types (Barwise and Ehrenberg 1988; Gunter and Wober 1992). Table 5.2 uses the programme type classification system outlined in Chapter 3 (“Methodology”) to reconfirm and illustrate these broad programme appreciation patterns for the five week period under analysis here.

Looking only at the “Average Score” column, it can be seen that sports, information programmes (particularly documentaries), light drama, and some light entertainment programmes (e.g. situation comedy and quiz shows/panel games) generally achieve somewhat higher than average appreciation scores. Somewhat lower than average scores are then received by films, heavy drama and arts programmes, news, and much other light entertainment programming.

Under the “Miscellaneous” category, it is interesting to note how religious programmes tend on average to achieve especially high appreciation scores while party political broadcasts tend to achieve especially low appreciation scores. (Although there are only 16 programmes of each of these types summarised in this table, these average programme appreciation scores are very much in line with similar patterns that have been repeatedly noted in the past.)

As might be expected, looking now at the “Range” column in Table 5.2, the range of appreciation scores achieved by different programme types also varies. (Standard Deviations, such as those presented in Tables 5.1 and 5.3 would mostly show similar patterns, but for small n’s - e.g. <20 - the ranges are lower just because of that. Ranges are presented here instead just to illustrate the ‘extremes’ of how much the programme appreciation scores within different programme type categories can potentially vary. The largest ranges, for miscellaneous and children’s programming, also show standard deviations well above the 5.1 average in Table 5.1.) One news programme is typically not all that much different from another news programme, so the range of scores achieved by different news programmes is relatively narrow. In contrast, films are more heterogeneous as a programme type (e.g. in style and content) and different films will therefore achieve a wider range of programme appreciation scores. Again such patterns are consistent with and reconfirm the findings of past research (Barwise and Ehrenberg 1988; Gunter and Wober 1992).

Table 5.2: Appreciation Scores by Programme Type

BARB Weeks 16-20, n=3,015 programmes
which received a total of 25 or more responses

Programme Type	n	Avg AI	Highest Score	Lowest Score	Range
<u>ENTERTAINMENT</u>	1,254	73	87	50	37
Light Entertainment	437	73	85	51	34
Vanity	6	67	71	61	10
Situation Comedy	136	73	82	51	31
Other Comedy	34	69	80	56	24
Chat Shows	38	71	80	55	25
Quiz Shows/Panel Games	164	75	85	55	30
Cartoons and Animation	18	69	75	64	11
Family Shows	16	70	74	64	10
Contemporary Music	25	70	80	58	22
Light Drama	466	73	85	57	28
Long-running serials - UK	143	75	85	61	24
Long-running serials - non UK	168	71	80	57	23
Other series/serials UK	80	75	83	58	25
Other series/serials - non UK	75	73	85	62	23
Films	138	69	87	50	37
Made for Cinema	124	69	87	50	37
Made for TV	14	68	80	53	27
Sports	213	74	84	56	28
<u>DEMANDING</u>	1,464	72	85	49	36
Drama/Arts	36	68	78	49	29
Arts Progs + Special Events	25	68	78	52	26
Single Plays	11	67	73	49	24
Information	581	74	85	60	25
Documentaries and Features	199	75	85	61	24
Current Affairs	279	73	85	60	25
Hobbies/Leisure pursuits	103	73	82	61	21
News	847	71	81	61	20
National	340	71	81	61	20
Regional	507	72	79	62	17
<u>MISCELLANEOUS</u>	297	69	86	36	50
Children's	265	70	81	36	45
Religion	16	78	86	68	18
Party Political Broadcasts	16	54	67	41	26
All Programmes	3 015	72	87	36	51

5.4 PROGRAMME APPRECIATION BY CHANNEL

Table 5.3 compares average appreciation scores when programmes are grouped by channel of broadcast. Although such differences are not dramatic, the average programme on BBC2 or Channel 4 will achieve a higher programme appreciation score than will the average programme on BBC1 or ITV. Once again, these replicated results are completely consistent with past findings (Barwise and Ehrenberg 1988; Gunter and Wober 1992).

Table 5.3: Appreciation for Programmes on Different Channels

BARB Weeks 16-20 1993, n=3,015 programmes
which received a total of 25 or more responses

Channel	n Progs	Avg AI	Standard Deviation
BBC1	959	71	4.9
ITV	1,241	72	4.7
BBC2	362	74	5.5
Channel 4	453	73	5.7
All Programmes	3,015	72	5.1

Audience size issues will be dealt with in much greater detail in Chapter 7 (“Programme Appreciation and Audience Size”) but in order to help explain the patterns in Table 5.3 it should be noted here that a programme shown on BBC2 or Channel 4 (the ‘small’ channels) will invariably achieve a lower audience size rating than a programme showing at the same time on BBC1 or ITV (the ‘big’ channels). This means that a higher proportion of programmes on the big channels will meet the minimum 25 respondent criteria necessary before a programme appreciation score can be calculated. The “number of programmes” column in Table 5.3 therefore does not represent all programmes

broadcast on these channels over this five week period. Instead, these are all programmes broadcast which received audience appreciation responses from 25 or more individual panel members. One explanation for the higher overall averages for the small channels in Table 5.3 could therefore be that appreciation scores for many of the less popular programmes on these channels have not been included.

If programme appreciation scores could have been calculated for all programmes broadcast on each channel, and if the smaller channels still achieved higher appreciation scores than the big channels, this would apparently confound the Double Jeopardy pattern outlined in Chapter 2 (“Introduction to the Literature”). Double Jeopardy would suggest that smaller audience channels (comprised of smaller audience programmes) should be liked less overall by their fewer viewers than bigger audience channels are liked by their much greater numbers of viewers. Double Jeopardy, however, occurs when people have to choose between broadly similar items that differ in popularity (Ehrenberg, Goodhardt and Barwise 1990). Is the comparison in Table 5.3 between programmes on these small and big channels actually a comparison of like with like?

In fact, as outlined in Table 5.4, the types of programmes broadcast on BBC2 and Channel 4 do differ from those on BBC1 and ITV. (Note again, however, that the percentages outlined in Table 5.4 have been calculated to help explain the audience appreciation patterns in Table 5.3. Therefore only those programmes that received 25 or more responses are represented, not *all* programmes actually broadcast over this period on the different channels.) Although each channel is unique, the range and types of programmes broadcast on BBC1 and ITV are quite distinct from those on BBC2 and Channel 4. The smaller channels have a much greater proportion of

information programming (especially documentaries), sport, films and heavy drama. The bigger channels instead offer more news programming and light drama. Such differences could help to explain the lack of an expected Double Jeopardy pattern.

Table 5.4: Programme Types Broadcast on the Different Channels

BARB Weeks 16-20 1993, n=3,015 programmes
which received a total of 25 or more responses

Channel		Entertainment				Demanding			Misc
		Light Ent	Light Drama	Films	Sport	Drama & Arts	Info	News	
BBC1	%	11	15	3	5	1	17	33	15
ITV	%	12	20	2	4	1	13	39	9
BBC2	%	12	4	10	22	3	33	9	7
Chan 4	%	31	12	10	7	2	30	6	2
All	%	14	15	5	7	1	19	28	11

5.5 SUMMARY

This short chapter has outlined and explored patterns in the way programmes differ in the programme appreciation scores they receive. Programmes are mostly quite liked by those who view them, with the average programme achieving an overall appreciation score of 72. Around this mean, most programmes will achieve an appreciation score of somewhere between 62 and 82. Some small, but significant and generalisable, differences can be seen when programmes are classified and compared across different programme types and channels.

6. PROGRAMME APPRECIATION AND AUDIENCE COMPOSITION

6.1 CHAPTER SIX OVERVIEW

This chapter relates Chapters 4 and 5 to each other by addressing the interesting issue of whether programmes that achieve higher programme appreciation scores do so simply because they are attracting those audience segments which are more generous in their average overall scoring patterns. Essentially, this involves asking two questions. How does the audience make up of programmes (i.e. ‘audience composition’) vary in general and across different programme types and different channels? Do such audience composition patterns then have any implications in relation to the programme appreciation scores that different programmes or groups of programmes achieve?

The early portion of this chapter illustrates the relatively unsegmented nature of mass television audiences that was alluded to in Chapter 2 (“Introduction to the Literature”). Different programmes do of course attract different audiences. An afternoon soap opera will have a quite different audience profile from a prime-time drama programme. Nonetheless, these differences are smaller than is often supposed, especially when scheduling factors have been allowed for. For programmes shown at the same time on different channels, audiences will be largely unsegmented, simply reflecting the population composition of all those available to view at that time (Barwise and Ehrenberg 1988; Perry 1995).

The analysis framework in this chapter is again initially based on replication, albeit replication based on an extensive sample that could provide powerful

reconfirmations or possible refutations of earlier findings. The chapter concludes, however, with an entirely new contribution to knowledge. This involves looking much more closely at those ‘extreme’ programmes where audiences are relatively skewed or biased towards particular segments of viewers. Does an examination of the audience appreciation scores achieved by such programmes provide any evidence of programmes that reach quite highly targeted *and* especially involved audiences?

6.2 THE AUDIENCE COMPOSITION OF PROGRAMMES

How does the audience make up of programmes vary in terms of the audience segments they attract? The two middle columns in Table 6.1 summarise the audience composition of the average programme in this five week data set (e.g. 48% of the average programme audience is male and 52% is female; 6% of the average programme audience is aged 12 to 15, etc.), as well as the standard deviations of programmes from this mean. Given scheduling factors and the differences in audience members’ availability to view, it is unsurprising that the audience profile of different programmes can vary considerably across the 3,015 programmes in this sample.

It is interesting to compare the audience composition of the average programme to the actual profile of the individuals in the Television Opinion Panel (e.g. as discussed in the various tables in Chapter 4 “Individual Viewers’ Scoring Patterns”). The final column in Table 6.1, for example, shows that 31% of individual panel members are aged 55 and over. These individuals, however, account for 42% of the audience of the average television programme. This must reflect the fact that such older viewers are more likely than younger viewers to be available to view at different times of the day, and are therefore typically heavier viewers of television.

Table 6.1: The Audience Composition of Programmes

BARB Weeks 16-20 1993, n=3,015 programmes
which received a total of 25 or more responses

	% Composition of the Average Programme	% Composition Standard Deviation	% of Individuals in Panel
Male	48	14	48
Female	52	14	52
Age 12-15	6	9	8
Age 16-24	10	9	13
Age 25-34	14	7	17
Age 35-44	15	5	17
Age 45-54	13	5	14
Age 55+	42	19	31
Class AB	14	7	18
Class C1	22	6	25
Class C2	27	6	28
Class DE	37	9	29
Light Viewers	9	7	28
Average Viewers	30	11	40
Heavy Viewers	61	15	32
Low Scorers	30	6	31
Moderate Scorers	40	6	43
High Scorers	30	8	26
Consistent Scorers	20	7	16
Regular Scorers	56	6	59
Varied Scorers	24	8	25

Similar patterns are seen for the lower income social classes and for heavy viewers. While 29% of panel members are in the DE social class, these viewers account for 37% of the average programme audience. The 32% of panel members who are classified as heavy viewers account for fully 61% of the average programme audience. Note finally, however, that individuals with different scoring patterns (e.g. high and low scorers, and people who make consistent or varied use of the six point appreciation rating scale) are

represented in the average programme audience roughly in proportion to the individual make-up of the television opinion panel.

Table 6.2 now allows for scheduling effects by looking only at weekday programmes (i.e. Monday to Friday) within two specific 'starting time' bands.

**Table 6.2: The Audience Composition of Programmes
Within Two Starting Time Bands**

BARB Weeks 16-20 1993, Weekday programmes only
n=96 afternoon programmes beginning between 14:00 and 15:00 and
n=180 evening programmes beginning between 20:00 and 21:00

	Programmes beginning between 14:00 and 15:00		Programmes beginning between 20:00 and 21:00	
	Composition	Std. Dev.	Composition	Std. Dev.
Male	% 45	19	% 47	9
Female	55	19	53	9
Age 12-15	% 2	2	% 5	4
Age 16-24	7	6	9	5
Age 25-34	10	6	16	6
Age 35-44	15	5	15	4
Age 45-54	13	5	14	3
Age 55+	53	14	41	14
Class AB	% 11	5	% 16	6
Class C1	21	5	24	4
Class C2	26	5	27	5
Class DE	42	7	33	5
Light Viewers	% 4	4	% 14	5
Average Viewers	21	8	38	6
Heavy Viewers	75	10	48	8
Low Scorers	% 27	7	% 30	4
Moderate Scorers	39	6	42	3
High Scorers	34	9	28	4
Consistent Scorers	23	5	% 16	3
Regular Scorers	56	6	58	4
Varied Scorers	21	7	26	4

Especially for evening programmes, but also somewhat for afternoon programmes, much less variation in audience composition (e.g. lower standard deviations) is seen within programmes shown at the same time. This is despite the fact that such programmes will still be of quite varied types; will be broadcast on different channels; and will often have very different sized audiences. As one might expect for mid-afternoon programmes (i.e. starting between 14:00 and 15:00), higher percentages of women, older viewers, lower income social group viewers, and heavy viewers are seen in comparison to the figures in Table 6.1. (Heavy viewers in fact account here for fully 75% of the audience for the average mid-afternoon programme.) The audience profile for the average prime-time evening programme reflects more diversity, with higher percentages of light viewers and other segments which have less likelihood of being available to view at other times.

6.3 THE AUDIENCE COMPOSITION OF PROGRAMME TYPES

Table 6.3 now illustrates the audience profile for the average programme of different programme types. To allow for sufficient sample sizes within each 'segment' (i.e. the six segments for age), the analysis here has been restricted to some 1,200 programmes which received a total of 100 or more responses. This sample size issue becomes particularly important in calculating the appreciation scores in Tables 6.5 and 6.6. As a check on the patterns in Table 6.3, this table was later replicated for all 3,015 programmes with 25 or more responses. Almost identical percentage figures emerged, indicating that smaller programmes do not differ dramatically from larger programmes in terms of their average audience profiles.

Table 6.3: Audience Profiles for the Average Programme of Different Programme Types

BARB Weeks 16-20 1993, n=1,205 programmes which received a total of 100 or more responses

	Entertaining				Demanding			Other	Total		Overall
	Light Ent	Light Drama	Films	Sport	Drama Arts	Info	News	Misc	Ent	Dmd	
N	277	326	58	69	10	206	188	36	765	404	1,205
Sex	%	%	%	%	%	%	%	%	%	%	%
M	46	38	55	69	46	48	50	42	46	49	47
F	54	62	45	31	54	52	50	58	54	51	53
Age	%	%	%	%	%	%	%	%	%	%	%
12-15	8	8	6	4	5	6	2	15	7	4	6
16-24	13	14	14	9	10	10	4	11	13	7	11
25-34	15	17	17	11	16	16	11	12	15	14	15
35-44	16	17	18	12	17	16	15	14	16	16	16
45-54	12	13	15	13	12	14	15	9	13	14	13
55+	36	31	30	51	40	38	53	39	36	45	39
Class	%	%	%	%	%	%	%	%	%	%	%
AB	14	13	13	16	21	16	18	15	14	17	15
C1	23	23	23	24	25	24	23	20	23	24	23
C2	28	28	30	26	26	29	26	29	28	27	28
DE	35	36	34	34	28	31	33	36	35	32	34
Weight	%	%	%	%	%	%	%	%	%	%	%
Light	11	10	15	11	16	13	9	8	11	11	11
Avg	33	35	37	34	37	34	31	30	34	33	34
Heavy	56	55	48	55	47	53	60	62	55	56	55
Scorers	%	%	%	%	%	%	%	%	%	%	%
Low	28	28	32	32	30	30	31	27	29	30	29
Mod	41	40	43	42	43	41	42	40	41	41	41
High	31	32	25	26	27	29	27	33	30	29	30
Consist	%	%	%	%	%	%	%	%	%	%	%
Consist	18	19	15	19	14	17	22	18	18	19	18
Regular	56	57	56	59	58	57	57	57	57	57	57
Varied	26	24	29	22	28	27	21	25	25	24	25

For each row in this table, the idea is to look across the row for each programme type and to compare these figures to the "Overall" column. Again, even without considering the impact of scheduling factors and audience

availability to view, this table demonstrates that by and large there is little segmentation in the composition of the typical audiences attracted by very different programme types. The difference for the percentages of males attracted by programmes of different types, for example, is typically only two or three percentage points from the “Overall” average for males across all programmes. There are exceptions however. For example, men do form quite a high proportion (69%) of the audience for the average “Sports” programme while women form a relatively high proportion (62%) of the audience for the average “Light Drama” programme. Similarly, while the age 12 to 15 group forms 6% of the audience for the average programme overall, this group comprises a relatively large 15% proportion of the audience for the average “Miscellaneous” programme and a relatively low 2% proportion of the audience for the average “News” programme. (Note that the miscellaneous programmes category contains many children’s programmes.)

Again, it must be remembered that scheduling and audience availability factors can help to explain these apparent patterns and exceptions. For example, much light drama programming (soap operas, etc.) is shown in the mid-afternoon when the audience available to watch contains a higher proportion of female viewers. It is not therefore too surprising to see that the average light drama programme has an audience that is 62% female.

6.4 THE AUDIENCE COMPOSITION OF DIFFERENT CHANNELS

Table 6.4 now examines audience profiles for the average programme shown on each of the four channels.

Table 6.4: Audience Profiles for the Average Programme on Different Channels

BARB Weeks 16-20 1993, n=1,205 programmes
which received a total of 100 or more responses

	Channel				Overall
	BBC1	BBC2	ITV	C4	
N	449	154	399	203	1,205
Sex	%	%	%	%	%
M	47	59	43	46	47
F	53	41	57	54	53
Age	%	%	%	%	%
12-15	6	5	6	11	6
16-24	9	11	9	17	11
25-34	13	14	15	19	15
35-44	15	15	16	16	16
45-54	14	13	13	11	13
55+	43	42	41	26	39
Class	%	%	%	%	%
AB	17	18	12	14	15
C1	25	25	21	22	23
C2	27	26	28	30	28
DE	31	31	39	34	34
Weight	%	%	%	%	%
Light	11	17	8	12	11
Avg	34	37	30	36	34
Heavy	55	46	62	52	55
Scorers	%	%	%	%	%
Low	30	32	27	29	29
Mod	41	43	41	41	41
High	29	25	32	30	30
Consist	%	%	%	%	%
Consist	19	15	21	15	18
Regular	57	58	57	56	57
Varied	24	27	22	29	25

Although differences are again not all that dramatic, it can be seen that the average programme on either BBC2 or Channel 4 will attract a higher proportion of male viewers than will the average programme on each of the

two larger channels. Channel 4 is a 'younger' channel with the average programme attracting a relatively low 26% proportion of Age 55+ viewers in comparison to more than 40% for each of the other three channels. (The average age of viewers for the average programme broadcast on Channel 4 is roughly 37. This compares to an average age of 45 for BBC1, 49 for BBC2 and 44 for ITV.)

Looking across the rows in Table 6.4 for Social Class, the average programme on BBC1 and BBC2 will attract somewhat larger proportions of higher income AB viewers (17% and 18% respectively, in comparison to 12% for the average programme on ITV and 14% for Channel 4). Similar patterns are also seen for weight of viewing. The average programme on BBC2 attracts the highest percentage (18%) of light viewers when compared to the average programme on the other three channels. The average programme on ITV attracts the highest proportion (62%) of heavy viewers.

Overall, both Tables 6.3 and 6.4 demonstrate the relative lack of dramatic segmentation patterns where different channels and programme types would attract quite different average programme audience profiles. In relation to audience appreciation scoring patterns, there are even fewer exceptions to this general rule. The proportions of low, moderate, and high scorers are nearly the same for the average programme of each programme type (Table 6.3) and for the average programme shown on each channel (Table 6.4). Likewise, the proportions of viewers who are consistent versus more varied in their scoring patterns are nearly the same for the average programme of each type and channel.

6.5 PROGRAMME APPRECIATION & AUDIENCE COMPOSITION

The conclusion from this chapter so far must be that overall variations in average programme appreciation scores by programme type and by channel (as outlined in Chapter 5 “Programme Appreciation Scores”) are not due to these different programme groupings attracting average audiences with quite different appreciation scoring patterns (e.g. with viewers relatively skewed towards specific high or low appreciation scoring audience segments). For example, the audience composition of the average information/ documentary programme hardly differs from the audience composition of the average programme of any other type. Therefore such information/documentary programmes must achieve higher average programme appreciation scores only because they tend to be ‘liked’ slightly more on average by *all* individuals. Given the general individual viewers’ appreciation scoring patterns outlined in Chapter 4, there is little further indication of audience composition factors playing any role in determining average overall programme appreciation scores for the different programme types or channels.

Tables 6.5 through 6.8 confirm this conclusion by calculating programme appreciation scores overall and within various audience segments for the different programme types and different channels. Table 6.5, for example, appears very similar to Table 4.12 in Chapter 4 (“Individual Viewers’ Scoring Patterns”). In Tables 6.5 through 6.8, however, it is not average *individual* appreciation scores that are being compared (i.e. the average score each individual viewer has given to all the programmes he or she viewed over this five week period), but aggregated *programme* appreciation scores (i.e. the average of all individual responses to each particular programme).

Table 6.5: Appreciation Patterns for the Average Programme of Different Programme Types

BARB Weeks 16-20 1993, n=1,205 programmes which received a total of 100 or more responses

	Entertaining				Demanding			Other	Total		
	Light Ent	Light Drama	Films	Sport	Drama Arts	Info	News	Misc	Ent	Dmd	Overall
N	277	326	58	104	10	206	188	36	765	404	1,205
Sex											
M	72	70	68	74	66	74	69	67	71	71	71
F	75	75	71	75	68	77	73	71	75	75	75
Age											
12-15	74	80	73	79	66	76	69	67	77	72	75
16-24	71	72	68	68	66	72	63	59	71	67	69
25-34	69	72	68	72	63	71	64	63	70	68	69
35-44	71	73	69	71	65	74	68	65	72	71	71
45-54	73	73	69	72	66	74	67	68	73	71	72
55+	73	74	70	76	70	76	74	71	73	75	74
Class											
AB	71	71	67	72	65	73	69	67	71	71	71
C1	72	72	69	74	66	76	69	67	72	72	72
C2	74	74	70	74	67	75	70	69	74	73	73
DE	75	75	71	77	70	76	73	72	75	75	75
Weight											
Light	69	70	67	71	64	73	63	67	70	68	69
Avg	72	72	69	73	66	74	68	66	72	71	72
Heavy	75	75	71	76	69	76	73	71	75	75	74
Scorers											
Low	59	58	58	62	55	61	56	55	59	59	59
Mod	72	72	70	74	67	75	70	67	72	73	72
High	87	88	84	91	80	89	88	83	88	88	88
Consist											
Consist	79	80	76	79	77	78	79	77	79	79	79
Regular	74	74	71	75	68	76	71	70	74	73	73
Varied	69	67	65	71	61	71	62	61	68	67	67
Overall	73	74	70	75	67	75	71	69	73	73	73

For each of the 1,205 programmes under analysis, separate appreciation scores have been calculated in line with the various audience segments. For example, the AI for a programme's male viewers has been calculated by summing the scores given by each male viewer of the programme and then dividing this figure by the total number of male respondents to that programme. The 71 "Overall" figure in the "M" row of Table 6.5 then represents the average of a total of 1,205 of these male programme appreciation scores.

As was discussed in Chapter 4, individuals do give somewhat higher than average appreciation scores to some programme types (e.g. sports) and somewhat lower scores to other programme types (e.g. films). As was outlined in Table 5.2 of Chapter 5 ("Programme Appreciation Scores"), these individual patterns are reflected in the overall programme appreciation scores achieved by the average programmes of different programme types. Given these findings, the expectation for the 1,205 programme sample in Table 6.5 would be for quite predictable patterns to occur in the "Overall" column by audience segment, and in the "Overall" row by programme type.

Such expected patterns are apparent. Demographic and behavioural patterns by audience segment in the "Overall" column of Table 6.5 do fit quite well with the patterns established in Chapter 4 at the individual viewer level (e.g. average female programme appreciation scores are slightly higher than average male programme appreciation scores; average programme appreciation scores for older viewers are slightly higher than average programme scores for younger viewers - with the usual exception of the 12-15 age group; programme appreciation scores for lower income social groups are slightly higher on average than programme appreciation scores for higher income groups; etc.). Moreover, now quite predictable programme type patterns also hold in the "Overall" row of this table (e.g. sports and information

programmes tend to get relatively high average programme appreciation scores, while films and drama/arts programmes tend to get relatively low average programme appreciation scores).

The fact that such programme appreciation scores in the “Overall” row of Table 6.5 are slightly higher than the similar individual appreciation scores in the “Overall” row of Table 4.12 is due to two factors. As outlined in Table 6.1, higher scoring audience segments (i.e. older viewers, heavy viewers, consistent scorers) are all represented in higher proportions as a percentage of the average programme audience than as a percentage of the overall composition of the Television Opinion Panel. Furthermore, the programme appreciation scores in Table 6.1 also reflect the influence of audience size. In the next chapter (“Programme Appreciation and Audience Size”), a Double Jeopardy pattern will be outlined whereby programmes with large audiences tend to get slightly higher programme appreciation scores than programmes with small audiences. The individual appreciation scores in Table 4.12 have been calculated based on *all* programmes each individual viewed over this five week period. In contrast, the programme appreciation scores in Table 6.5 represent aggregated individual responses to only 1,205 large audience programmes.

The programme appreciation scoring patterns in the body of Table 6.5 are highly predictable simply from the overall row and column averages. This is formally illustrated in Table 6.6 using a similar exercise to that in Tables 4.12 through 4.14 to present deviations from a simple additive predictive model.

Table 6.6: Difference between Observed and 'Predicted' Average Programme Appreciation Scores for Different Programme Types

BARB Weeks 16-20 1993, n=1,205 programmes
which received a total of 100 or more responses

	Entertaining				Demanding			Other	Total		Overall
	Light Ent	Light Drama	Films	Sport	Drama Arts	Info	News	Misc	Ent	Dmd	
N	277	326	58	104	10	206	188	36	765	404	1,205
Sex											
M	1	-2	0	1	1	1	0	0	0	0	71
F	0	-1	-1	-2	-1	0	0	0	0	0	75
Age											
12-15	-1	4	1	2	-3	-1	-4	-4	2	-3	75
16-24	2	2	2	-3	3	1	-4	-6	2	-2	69
25-34	0	2	2	1	0	0	-3	-2	1	-1	69
35-44	0	1	1	-2	0	1	-1	-2	1	0	71
45-54	1	0	0	-2	0	0	-3	0	1	-1	72
55+	-1	-1	-1	0	2	0	2	1	-1	1	74
Class											
AB	0	-1	-1	-1	0	0	0	0	0	0	71
C1	0	-1	0	0	0	2	-1	-1	0	0	72
C2	1	0	0	-1	0	0	-1	0	1	0	73
DE	0	-1	-1	0	1	-1	0	1	0	0	75
Weight											
Light	0	0	1	0	1	2	-4	2	1	-1	69
Avg	0	-1	0	-1	0	0	-2	-2	0	-1	72
Heavy	1	0	0	0	1	0	1	1	1	1	74
Scorers											
Low	0	-2	2	1	2	0	-1	0	0	0	59
Mod	0	-1	1	0	1	1	0	-1	0	1	72
High	-1	-1	-1	1	-2	-1	2	-1	0	0	88
Consist											
Consist	0	0	0	-2	4	-3	2	2	0	0	79
Regular	1	0	1	0	1	1	0	1	1	0	73
Varied	2	-1	1	2	0	2	-3	-2	1	0	67
Overall	73	74	70	75	67	75	71	69	73	73	73

To illustrate, the observed minus 'predicted' deviation of "1" for males and light entertainment programmes in Table 6.6 is calculated as follows.

- (1) The observed Male average programme appreciation score for Light Entertainment programmes in Table 6.5 is 72.
- (2) A 'predicted' score can be calculated from the overall row and column averages in Table 6.5. The overall average programme appreciation score for all viewers of all programme types is 73 (as seen in the lower right hand corner cell of Table 6.5). The overall average male programme appreciation score for all programme types is 71 (as seen in the "overall" cell of row "M"). $73 - 71 = 2$. For each programme type, the 'predicted' male programme appreciation scores are then simply two points less than those scores in the "Overall" row of Table 6.5. For Light Entertainment programmes, this means a 'predicted' score for males of $73 - 2 = 71$.
- (3) The observed minus 'predicted' deviation for Male programme appreciation scores and Light Entertainment programmes in Table 6.6 is therefore $72 - 71 = 1$.

The average programme appreciation score patterns in Table 6.6 mostly show similarly small deviations to the individual appreciation score patterns in Table 4.14. For the most part, there are few if any average programme appreciation scores, calculated within programme types for particular audience subgroups, that are surprisingly high or low given overall average programme appreciation scoring patterns within these audience subgroups.

Limited exceptions occur largely where they would now be expected given earlier individual viewer level analyses. "Female" programme appreciation scores are lower for sports programmes than might be expected given womens' scores on average for all programmes. Programme appreciation scores for news programmes, calculated for the oldest 55+ age group, are

slightly higher than one might expect given the overall average programme appreciation scores calculated for all programmes viewed by such viewers. The opposite is true for the younger age groups. Similarly, “Light Viewer” and “Average Viewer” programme appreciation scores for news programmes are relatively low while “Heavy Viewer” programme appreciation scores for such news programmes are slightly higher than might be predicted.

Table 6.7 concludes this section by calculating and presenting programme appreciation scores overall and within various audience segments for the different channels. Given the findings outlined earlier for different segments’ programme appreciation scores by programme type (i.e. programme appreciation patterns in Table 6.5 that closely reflect the individual level patterns that were illustrated in Table 4.12), one would expect a table very similar to Table 4.15 (“Individual’s Average Appreciation of Different Channels”) to emerge. This is the case, although again, for the reasons outlined above with respect to Table 6.5, such *programme* appreciation scores in the “Overall” row and “Overall” column of Table 6.7 are slightly higher than the average appreciation scores calculated by channel for *individuals* in the “Overall” row and “Overall” column of Table 4.15.

Table 6.7: Appreciation Patterns for the Average Programme on Different Channels

BARB Weeks 16-20 1993, n=1,205 programmes
which received a total of 25 or more responses

	Channel				Overall
	BBC1	BBC2	ITV	C4	
N	449	154	399	203	1,205
Sex					
M	70	75	70	74	71
F	73	76	74	77	75
Age					
12-15	74	78	76	74	75
16-24	66	71	69	75	69
25-34	67	73	68	73	69
35-44	70	73	70	74	71
45-54	70	74	71	77	72
55+	73	75	74	74	74
Class					
AB	69	73	70	74	71
C1	70	75	71	76	72
C2	72	75	73	76	73
DE	73	77	74	77	75
Weight					
Light	67	73	67	75	69
Avg	70	74	71	76	72
Heavy	73	77	74	77	74
Scorers					
Low	57	64	57	62	59
Mod	71	75	71	76	72
High	87	90	87	90	88
Consist					
Consist	78	79	79	79	79
Regular	72	75	73	77	73
Varied	64	72	66	73	67
Overall	71	75	72	76	73

The remainder of Table 6.7 is very 'flat' as is illustrated in Table 6.8 which once again formally presents 'observed' versus 'predicted' row and column comparisons. There are few limited exceptions where programme appreciation

scores calculated by channel within particular audience subgroups would be relatively high or low in comparison to overall programme appreciation scoring patterns for such subgroups.

Table 6.8: Difference between Observed and 'Predicted' Average Programme Appreciation Scores for Different Channels

BARB Weeks 16-20 1993, n=1,205 programmes
which received a total of 25 or more responses

	Channel				Overall
	BBC1	BBC2	ITV	C4	
N	449	154	399	203	1,205
Sex					
M	1	2	0	0	71
F	0	-1	0	-1	75
Age					
12-15	1	1	2	-4	75
16-24	-1	0	1	3	69
25-34	0	2	0	1	69
35-44	1	0	0	0	71
45-54	0	0	0	2	72
55+	1	-1	1	-3	74
Class					
AB	0	0	0	0	71
C1	0	1	0	1	72
C2	0	0	1	0	73
DE	0	0	0	-1	75
Weight					
Light	0	2	-1	3	69
Avg	0	0	0	1	72
Heavy	1	1	1	0	74
Scorers					
Low	0	3	-1	0	59
Mod	1	1	0	1	72
High	1	0	0	-1	88
Consist					
Consist	1	-2	1	-3	79
Regular	1	0	1	1	73
Varied	-1	3	0	3	67
Overall	71	75	72	76	73

Where exceptions do emerge in Table 6.8, they are limited and largely in line with the individual level patterns outlined in Chapter 4. “Male” programme appreciation scores for BBC2 are higher than might be expected given such “male” programme appreciation scores across programmes shown on all channels. Programme appreciation scores for Channel 4, calculated for younger viewers only (with the exception of the youngest 12-15 age group), are slightly higher than might be expected or ‘predicted’ given younger viewers’ overall average programme appreciation scores across all channels. The opposite is true for the oldest 55+ age group, where programme appreciation scores on Channel 4 are slightly lower than might be expected given these older viewers’ programme appreciation scores across all channels. Finally programme appreciation scores for light viewers and varied scorers are slightly higher than might be expected for the two smaller channels (BBC2 and Channel 4), while programme appreciation scores for consistent scorers are slightly lower than might be expected for these channels.

To provide an interim summary conclusion to this section and to this chapter so far, television audiences for different programme types and channels are relatively unsegmented, especially in comparison to other print and radio broadcast media where different types of magazines and radio stations often attract highly targeted audiences along various segmentation criteria (e.g. sex, age, income, etc.). The programme appreciation score patterns outlined in Chapter 5 and in this chapter reflect this lack of overall audience segmentation. Chapter 4 showed that there are demographic and behavioural patterns in the way different audience segments make use of the audience appreciation scale overall (i.e. women tend to give higher scores than men, heavy viewers tend to give higher scores than lighter viewers, etc.). As outlined in this chapter (and in the individual level patterns outlined in Chapter 4), such patterns do not seem to vary further in terms of the average programme appreciation scores achieved by different types of programmes and by different channels.

Consider now, however, the relatively few exceptions to these overall patterns. For example, in Table 6.3 it was shown that men form a relatively high proportion of the average audience for sports programmes (i.e. 69%). In Table 6.5, programme appreciation scores for male viewers were also slightly higher than might be expected or ‘predicted’ given male viewers’ overall programme appreciation scoring patterns. Similarly, in Table 6.4, it was shown that the average proportion of older viewers (age 55+) in the audience of the average programme on Channel 4 was smaller than for programmes on the other three channels (i.e. such viewers comprise 26% of the average programme audience for Channel 4 in comparison to 41 to 43% of the average programme audience for the other three channels). In Table 6.7, the average programme appreciation score calculated for these age 55+ viewers for the average Channel 4 programme was lower than would be predicted given the programme appreciation scores for such viewers across all programmes on all channels. Do these limited exceptions begin to provide a suggestion that there are programme types (or channels) that not only attract somewhat more segmented or ‘targeted’ audiences, but that such ‘targeted’ viewers are also relatively more involved with these programmes than are other viewers who also happen to be watching? The remainder of this chapter will explore this question and begin to consider the nature of ‘narrowcasting’.

6.6 PROGRAMME APPRECIATION AND ‘NARROWCASTING’

Despite the fact that research by the Ehrenberg Group into patterns of television audience viewing behaviour has repeatedly shown television to be a largely unsegmented medium (as discussed in Chapter 2 “Introduction the the Literature” and in Barwise and Ehrenberg 1988), there are sufficient

variations in programme audience profiles to allow for savings in targeting advertising to specific audience segments.¹ Media planners therefore remain greatly concerned over questions of television audience segmentation. And what of the future? As the television medium fragments, there is often talk of the rise of 'narrowcasting', with channels and programmes that are expected to reach small, well-defined, and highly involved audiences. Even today, cable and satellite stations would love to be able to demonstrate to advertisers that their programming delivers highly targeted and especially attentive audiences.

It is interesting to note here that much of the pilot work for the Television Audience Assessment research discussed briefly in Chapter 2 ("Introduction to the Literature") was supported by US cable interests. These supporters hoped that this research into the development of possible syndicated measures of audience reaction (e.g. similar measures to the UK's audience appreciation index) would provide evidence to support claims that cable's specialised programming would involve viewers and deliver especially attentive, premium audiences. Instead, TAA's research (TAA 1983b) demonstrated that cable subscribers did not find cable programming to be any more appealing or involving than broadcast programmes. In fact viewers made little distinction between cable and broadcast programming, and any variation in viewer response seemed to be directly due to the programme itself.

¹ Indeed, media planners often highlight such differences through the use of indices. For example, 14% of the average programme audience in Table 6.1 is comprised of what many companies would consider highly desirable higher income social class AB viewers. A rare programme that attracts an extremely high proportion of such AB viewers (i.e. two standard deviations above the average proportion for all programmes) would have an audience comprised 28% of AB viewers. If programmes were classified in terms of their 'AB index', the average programme would be placed at 100 while this extreme programme would then have an one hundred percent higher index of 200. When advertising placement decisions are made through the use of such indices, it is often easy to forget the reality that 86% of the viewers of the average programme are *not* social class AB viewers. Even for the 'extremely targeted' programme here, a full 72% of the audience is comprised of viewers who are not in the highest income AB social group.

The data in this thesis only covers the four main UK terrestrial broadcasters so specific issues with respect to smaller audience cable and satellite programmes and channels cannot be directly investigated. Nevertheless, within the limits of this four channel data set, there are possible approaches that can be taken to look more closely at whether there are examples of programmes that reach quite targeted and highly appreciative audiences. It could even be argued that in repeatedly concentrating attention on the ‘average’ programme (e.g. the average programme overall, the average programme of a specific type, or the average programme shown on a particular channel), earlier analyses in this thesis may have glossed over some of the variation that does exist across specific individuals and across specific programmes.

The following section of the thesis now represents an entirely new contribution to knowledge. It examines issues concerning the extent to which some programmes do attract quite highly ‘targeted’ or segmented audiences, and how audience appreciation scores given to such programmes may be related to this ‘targeting’.

6.7 PROGRAMME APPRECIATION AND AUDIENCE BIAS

Are programmes that attract biased audiences (e.g. heavily skewed towards male viewers) more likely to be enjoyed by such viewers in comparison to other viewers who happen to be watching? A common sense hypothesis might be that ‘the better targeted the programme, the more likely it is to be enjoyed by the target group’.

(a) Framing the Research Question

This is an interesting issue and it is one that first arose with respect to this thesis in late 1993. At that time, I was in the pilot stage of early analyses based on an initial one week sample of aggregated BARB summary audience appreciation data. I was directly in the midst of replicating some past studies on the relationship between audience appreciation and audience size (to be discussed in detail in Chapter 7) and I was also beginning to consider relative programme audience bias and its relationship to programme appreciation scores calculated within specific audience segments. On 7 October 1993, it seemed like quite a coincidence when a short article of direct relevance and extreme interest to me appeared in the “Media Watch” column of *Marketing* magazine. This column is compiled each week with the help of CIA Media, a large London media buying firm. The crux of this brief article lay in the following table (labelled Table 6.9 for the purposes of this thesis) and direct quotation :

Table 6.9: Enjoyment Ratings: 16 to 34 year olds

Programme	Date/Day	16-34 Index	Enjoyment Rating (%)
1. Look Who's Talking	Sept 23 (Thurs)	119	78
2. Tango and Cash	Sept 18 (Sat)	105	75
3. Soldier Soldier	Sept 21 (Tues)	100	78
4. Cocoon: The Return	Sept 20 (Mon)	100	70
5. The Upper Hand	Sept 23 (Thurs)	95	69
6. Coronation Street	Sept 20 (Mon)	87	72
7. The Bill	Sept 21 (Tues)	84	71
8. Blind Date	Sept 18 (Sat)	83	63
9. Freddie Star	Sept 22 (Wed)	67	68
10. Emmerdale	Sept 21 (Tues)	60	63

“This week’s Media Watch looks at some recent top-rating ITV programmes, and checks out how much the important 16 to 34 age group claims to enjoy them. We have also shown the relative bias of the audience to this age group for each programme. The table shows a

close correlation between the two rankings, suggesting that the better targeted the programme, the more likely it is to be enjoyed by the targeted sub-group. If you believe that TV environment is relevant, it underlines the importance of careful programme linked schedule construction.” (Marquis 1993)

The following are key quotations and a table (labelled Table 6.10 for the purposes of this thesis; a rearranged version of Table 6.9) from a letter I wrote to the Editor of *Marketing* in reference to the above column:

“But how much does your example merely reflect a relationship between audience size and audience enjoyment amongst the population you are looking at (i.e. 16 to 34 year olds)? By looking at the relative proportion of 16-34 year old viewers in the audiences for different programmes, you obscure any audience size effects. Based on your 16-34 Index and on BARB’s consolidated total audience figures, the following table provides derived estimates of the size of each programme’s 16-34 age group audience. The programmes are listed in order of audience size, and your 16-34 and enjoyment indices are provided.

**Table 6.10: Enjoyment Ratings: 16 to 34 year olds,
with Table Rearranged by Audience Size**

Programme	Date/Day	16-34 Index	Audience Size (*)	Enjoyment Rating (%)
			%	
1. Coronation Street	Sept 20 (Mon)	87	26	72
2. Look Who's Talking	Sept 23 (Thurs)	119	25	78
3. Tango and Cash	Sept 18 (Sat)	105	21	75
4. The Bill	Sept 21 (Tues)	84	20	71
5. Cocoon: The Return	Sept 20 (Mon)	100	19	70
6. Soldier Soldier	Sept 21 (Tues)	100	18	78
7. The Upper Hand	Sept 23 (Thurs)	95	18	69
8. Blind Date	Sept 18 (Sat)	83	17	63
9. Freddie Star	Sept 22 (Wed)	67	13	68
10. Emmerdale	Sept 21 (Tues)	60	10	63

(*) Source BARB, % Ratings = Number of Viewers aged 16-34 for this programme as a percentage of total UK viewing population aged 16-34

The pattern in the last two columns of this table suggests that audience enjoyment is a function of audience size. This is in line with the widespread phenomenon known as Double Jeopardy, often cited by Professor Andrew Ehrenberg of South Bank Business School - not

only are less popular programmes watched by fewer people, but they are also liked somewhat less by those who choose to watch them.

Although the correlations between audience size and enjoyment and between your 16-34 Index and enjoyment are similar (0.7 and 0.8), the audience size effect is well established and should be accounted for..... This question of the effect of targeted programmes on enjoyment by audience subgroups offers great potential for more research, but perhaps the conclusion from this data should simply be that *the greater the number of viewers in the audience for a programme, the more likely it is to be enjoyed.*" (Carrie 1993)

The above discourse is interesting, not only because it helped to frame the analyses to come in the remainder of this section, but also because the original article demonstrates the need to understand general patterns in one's data before drawing conclusions from it.² As mentioned earlier, the Double Jeopardy relationship between audience size and audience appreciation will be discussed in detail in Chapter 7 ("Programme Appreciation and Audience Size"). For the purposes of this chapter, two main conclusions from Chapter 7 should be summarised in advance here. Firstly, overall (for all programmes on all channels broadcast at any time of the day) little relationship can at first be demonstrated between audience size and audience appreciation. Secondly, however, when allowances are made for scheduling and channel effects, a definite but limited Double Jeopardy pattern does then emerge (i.e. smaller audiences lose out in two ways; not only are less popular programmes watched by fewer people, but they are also liked somewhat less by those who do choose to watch). The programmes chosen by CIA Media to illustrate the above example were all top rating, prime-time evening programmes broadcast

² For example, what if one had done a similar exercise to the above in order to compare a 16-34 Index to a 55+ Index for these same ten programmes and then to look at relative enjoyment levels amongst the 16-34 age group and amongst the 55+ age group? Without a knowledge of general patterns in this data (i.e. older viewers tend to give higher appreciation scores on average than younger viewers), spurious conclusions could quite easily be drawn.

on ITV. Such a sample allows for channel effects on audience size by looking at only one channel (i.e. ITV only), and it also allows broadly for scheduling effects on audience size by looking at programmes broadcast at a similar time of day (i.e. prime time evening programmes only). Given such allowances, the expectation would be for a Double Jeopardy audience size relationship to occur as outlined in the above letter to the Editor of *Marketing*.

Building on the relative audience bias concept and the related notion of programme ‘targeting’, and realising the largely unsegmented overall nature of television programme audiences (as outlined earlier in this chapter), it was decided to look much more closely at those ‘extreme’ programmes where audiences are quite skewed or biased towards particular segments of viewers. Does an examination of the audience appreciation scores achieved by such programmes provide any evidence of programmes that reach quite highly targeted and especially involved (i.e. appreciative) audiences?

Again, the proposition or hypothesis to be investigated here is that ‘the better targeted the programme, the more likely it is to be enjoyed by the target group’. The use of the word ‘targeted’ here does not imply that anything is known about the actual viewership a programme producer or broadcaster is aiming at for each particular programme. Such detailed information is not available within this data set. Instead, the relative audience composition of different programmes is used as a proxy for how targeted a programme’s content is towards a particular demographic (e.g. gender, age, and social class) or behavioural (e.g. heavy and light viewers) audience segment. Programmes with audiences that are highly biased towards a certain audience segment(s) are said to be targeted at this segment(s). One must be aware of the reasons why such audience bias does occur (e.g. scheduling factors), while also realising that programme makers and broadcasters are quite aware of such

‘availability to view’ factors and are designing programmes to satisfy the appropriate ‘target’ audience. The mid-afternoon weekday viewing audience on all channels, for example, is relatively skewed towards female viewers and the content of many mid-afternoon programmes reflects this. Since female viewers comprise a large proportion of the audiences for many (typically small total audience size) afternoon programmes, are these female viewers also enjoying such programmes more than other viewers (e.g. men) who happen to be watching?

Section 6.7 (b) now provides a detailed example of the analysis approach taken to formally look at the relationship between programme appreciation scores and the relative audience bias of programmes by *gender*. Section 6.7 (c) then provides a summary overview of findings from similar analyses that were repeated to look at the relationship for programmes with highly biased audiences by *age*, *social class*, and *weight of viewing*. Finally, Section 6.7 (d) concludes with an overall summary discussion on programme appreciation and audience bias.

(b) Detailed Analysis for Programme Audience Bias by Gender

Table 6.11 shows that there are specific programmes that are highly biased towards male viewers (i.e. 80 to 94% of all viewers watching are male) and that these male viewers tend to like such programmes more than other viewers who are watching (i.e. women). The programme appreciation scores calculated here for male viewers are on average four points higher than those for female viewers (an average score of 71 versus 67). In 19 out of the 20 programmes here, the ‘male’ programme appreciation score calculated for the programme is higher than the ‘female’ programme appreciation score. On average, given all

the patterns presented so far in this thesis, higher programme appreciation scores would otherwise have been expected for female viewers than for male viewers. This would be the case even though almost all of the programmes in Table 6.11 are sports programmes, where the difference between male and female appreciation scores is on average narrower than for other programme types (e.g. see Table 6.5).

Table 6.11: The 20 Programme Showings Most Biased Towards Men

5 weeks, BARB weeks 16 to 20 1993 (April 19 to May 17)
Most biased audience programmes from a sample of all programmes (n=2,023 programmes) over this five week period which achieved 50 or more responses

Title	Genre	Showing* (Wk/Day/ Channel/Hour)				Aud Size	% Male	Male AI	Female AI	Overall AI
		W	D	C	H					
Gazzetta Football Italia	Sport	19	6	4	11	51	94	73	47	71
Football Italia	Sport	19	7	4	14	126	91	67	53	66
Football Italia	Sport	16	7	4	14	125	87	71	64	70
Standing Room Only	Info/Docu	18	1	2	18	88	86	77	70	76
Football Italia	Sport	18	7	4	14	100	86	68	64	68
Sportscene	Sport	17	3	2	19	67	85	56	54	56
Cricket	Sport	19	2	2	15	200	85	76	76	76
Football Focus	Sport	18	6	1	12	195	85	73	72	73
Three Women in Love	Film	16	5	2	24	53	83	65	60	65
The London Match	Sport	18	7	3	14	52	83	82	71	80
Cricket	Sport	19	2	1	13	103	83	73	72	73
Football Italia	Sport	20	7	4	14	72	82	68	65	68
The Central Match - Live	Sport	19	7	3	14	53	81	72	82	74
Sportsnight	Sport	18	3	1	22	348	81	70	69	70
Golf	Sport	18	5	2	14	89	81	76	69	75
Sportsnight	Sport	19	3	1	22	336	81	72	66	71
Sunday Grandstand	Sport	19	7	2	14	248	81	71	68	70
Racing	Sport	19	5	2	14	82	80	75	74	75
The Match	Sport	17	7	3	14	82	80	73	71	73
Rugby Special	Sport	16	7	2	17	97	80	71	74	72
Average						128	84	71	67	71

Note: * Week (W): BARB Weeks 16 through 20 (This five week period)
Day (D): 1 = Monday; 7 = Sunday
Channel (C): 1=BBC1; 2=BBC2; 3=ITV; 4 = Channel 4
Hour (H): Hour in which the programme started (24 hour clock)

These ‘most biased towards men’ programmes all have relatively small audiences in relation to all programmes broadcast over this five week period. The statistical cut-off for the smallest programme included (“Aud Size” in Table 6.11 = audience size) was 50 responses, to allow for the calculation of AI scores. There are still sample size issues here, however, in some of these female appreciation scores (i.e. for the smallest audience programmes) are being calculated across as few as three to five viewers. This issue is addressed somewhat through using larger samples of programmes in later analyses (i.e. in Tables 6.10 and 6.11). Also, as a check, Table 6.11 was later replicated with nearly identical average results using a statistical cut-off for programme audience size of 100 responses. The biased audience programmes in Table 6.11 tend to be ones broadcast in off-peak times (mid afternoon or late evening on weekdays, or mid-afternoon on the weekend). Seventy-five percent of these programmes were broadcast on the two smaller channels (BBC2 and Channel 4). The same titles show up repeatedly in this table because these are individual *showings* of a programme title over this five week period. (While a film title will only have been broadcast once during this five week period, television schedules are highly structured and repetitive with different episodes of the same title shown on the same day and time from week to week.)

In repeating this analysis for the 20 programme showings with an audience composition most biased towards women, similar but less definitive patterns are seen. A full table is not presented here because the results can be quickly summarised. A wider range of programme types occurs, dominated by light drama (mostly soap operas broadcast in the mid-afternoon hours on ITV) and by mid-morning news and information programmes (e.g. “Top of the Morning” and “This Morning”). These are all weekday programmes. The

mid-morning and mid-afternoon broadcast times suggest that a key factor behind such female biased audiences is simply the greater availability of women on average to view television at these times. The female proportion of the audience for these 20 programmes ranges from 87% down to 79%, and the average female programme appreciation score is 74 in comparison to the average male programme appreciation score of 68 (again there are sample size issues with these male programme appreciation score calculations). This 6 point spread in average programme appreciation scores would seem to suggest that programmes that are more effectively 'targeted' towards women are more likely to be enjoyed by that target group than by others (i.e. men) who also happen to be watching. However, this conclusion is less definitive than in the findings outlined above for the twenty programmes most biased towards men. For example, women on average are expected from the start to give higher scores than men. For light drama, news, and information programmes (e.g. in Table 6.5) one would already expect at least a 3 to 5 point spread in average programme appreciation scores for men and women before even beginning to consider any particular programmes that might be highly biased or targeted towards women.

Table 6.12 now summarises and compares the above patterns using a slightly broader sample of 40 'male biased' and 40 'female biased' programmes (to begin alleviating some sample size concerns).³ Table 6.13 then compares programme appreciation patterns for these still 'extremely' biased programmes to a more comprehensive overview of the relative audience bias by gender of more than 2,000 programmes broadcast over this five week period.

³ It is interesting to note that for men, even this larger sample is still dominated by sports programmes (29 out of 40 programmes) although a few more information/documentary programmes and films do also begin to appear. On a lighter note, this seems encouraging for the state of the average male mind until one begins to look closer at the titles of such films and information programmes that are attracting highly male audiences. For example, these programme titles include the documentary "Videos, Vigilantes, and Voyeurism." and the film "La Lectrice". It seems there is some grain of truth in the stereotype that the average man is relatively sports and sex obsessed.

Table 6.12: Programme Appreciation and Those Programmes with Most Biased Audience Composition By Gender

5 Weeks, Weeks of April 19/1993 through to May 17,1993
All Programmes with 50 or more responses (n=2,023 programmes)

	Average Rating		Male Avg %	Female Avg %	Male Avg AI	Female Avg AI	Overall Avg AI
	%	Std Dev					
All Programmes <i>n = 2,023</i>	8	7	47	53	70	74	72
40 "Most Male" Progs	4	1	81	19	72	69	72
40 "Most Female" Progs	3	3	21	79	68	75	74
<i>Difference From 'Predicted' for All programmes</i>					<i>Men:</i>	<i>Women:</i>	
40 "Most Male" Progs					+2	-5	
40 "Most Female" Progs					-4	-1	
Sports Programmes <i>n = 153</i>	6	4	69	31	74	75	75
40 "Most Male" Progs	4	1	81	19	72	69	72
<i>Difference From 'Predicted' for Sports programmes</i>					<i>Men</i>	<i>Women:</i>	
40 "Most Male" Progs					+1	-3	

At first glance, Table 6.12 may appear confusing. The key figures to note, looking first at the centre section of the table ("All Programmes"), are the "Difference from Predicted" figures for men and women. These are calculated in a similar manner to now familiar 'observed' versus 'predicted' row and column comparisons first presented in Section 4.6 (b) of this thesis. As has been noted, the programmes most biased towards viewers of different genders represent quite different mixtures of programme types (e.g. the most biased 'male' programmes include sports, information and film while the most biased 'female' programmes include light drama, information and news). Hence a comparison of average programme appreciation scores across these two

samples of 40 programmes is not strictly a comparison of 'like' with 'like' since it is known that programme appreciation patterns do vary by programme type.

The "Difference from Predicted for All Programmes" figures here are an attempt to broadly allow for such programme type effects. This assumes that the relative male and female programme appreciation scores in the "All Programmes" row of Table 6.10 hold constant for all rows in this table. For the 40 most female biased programmes shown in the third row of Table 6.12, for example, the "Overall Average AI" (i.e. 72) is two points higher than the "Overall Average AI" in the "All Programmes" row (i.e. 70). Two points are therefore added to both the "Male Avg AI" and "Female Avg AI" scores in the "All Programmes" row to arrive at 'predicted' scores one might expect overall for men and women for a mixed sample of programmes with an average programme appreciation score of 74. The "Difference from Predicted" figure of -4 for Men is therefore the observed score in the "40 Most Female Programmes Row" (i.e. 68) minus the 'predicted' Female Average AI score (i.e. $70 + 2 = 72$).⁴

What do these "Difference from Predicted" scores then imply? For the forty programmes most biased towards men, the conclusion is clear. The (statistically significant) average male appreciation score for these programmes is two points *higher* than would be expected based on the average male programme appreciation score for all programmes. The average female appreciation score is then a full five points *lower* than would be expected given the average female programme appreciation score for all programmes. Given the limited range of programme appreciation scores overall, as demonstrated throughout

⁴ For the "40 Most Male Programmes" shown in the second row of Table 6.12, the "Overall Average AI" in this row (i.e. 72) is the same as the "Overall Average AI" in the "All Programmes" row. This simplifies any calculation as the "Difference from Predicted" figure of +2 for Men is simply the 72 observed figure in the "40 Most Male Programmes" row minus the 70 'predicted' figure in the line for "All Programmes".

this thesis, these differences are quite large. However, it is known that almost seventy-five percent of these male biased programmes are sports programmes. The bottom portion of Table 6.12 therefore recognises that it would be fairer to base 'predicted' scores for these forty programmes on patterns across all *sports* programmes only. *For a sample of sports programmes with an average overall AI of 72*, the results are still splendidly clear. The male average programme appreciation score is one point higher than might be expected while the female programme appreciation score is three points lower than would be predicted.

Results for this limited sample of the most 'extreme' male biased programmes would therefore confirm an initial proposition that 'the better targeted the programme, the more likely it is to be enjoyed by the target group.' For sports programmes on average, the programme appreciation score given by all female viewers (75) is one point higher than the programme appreciation score given by all male viewers (74). As audiences for particular sports programmes become better 'targeted' at men, programme appreciation scores given by male viewers actually begin to exceed those given by female viewers. A key qualifying phrase in the previous paragraph however was "*for a sample of sports programmes with an average overall AI of 72.*" Sports programmes that are better 'targeted' at men may be liked *relatively* more by men than by women, but it appears that such programmes are getting lower overall programme appreciation scores than the average sports programme. Both the average male programme appreciation score and the average female programme appreciation score are lower for these 'targeted' programmes than for the average sports programme.

Two factors would have to be considered here in trying to explain why these programmes on average achieve lower programme appreciation scores than might be expected for sports programmes. Firstly, while twenty-nine of these forty male biased programmes are sports programmes, remember that the remaining eleven programmes are a mixture of films and information programmes with lower programme appreciation. Secondly, as will become clearer on reading the next chapter (“Programme Appreciation and Audience Size”) a limited Double Jeopardy audience size effect could be operating here. The nature of this sample selection has controlled for many factors that might otherwise obscure any relationship between audience size and average programme appreciation (e.g. most of these 40 programmes are of a similar programme type and were broadcast at similar off-peak hours on the two smaller channels). Since these forty male biased programmes on average reach a smaller audience (average rating of 4%) than the average sports programme (6%) or the average programme overall (8%), they might well be expected on average to achieve lower programme appreciation scores regardless of how biased or unbiased their audience composition is.

Returning now to the forty most female biased programmes in Table 6.12, the pattern illustrated is more ambiguous. For a sample of all programmes with an average overall AI of 74, the observed average male programme appreciation score (68) is four points *lower* than might be predicted. The observed female programme appreciation score (75), however, is also *lower* by one point than might be predicted. Again, this pattern does suggest that programmes which are better ‘targeted’ at women will be liked *relatively* more by female viewers than by male viewers in relation to male and female programme appreciation scoring patterns for all programmes. Female viewers are not, however liking these highly targeted, and typically smaller audience, programmes more than all the other less targeted programmes they also watch.

Table 6.13 now concludes this analysis by looking beyond the limited number of 'extreme' cases summarised in Table 6.12. Here, 2,000 programmes are divided into five 20% bands according to how biased an audience each programme attracts. For male viewers on average, the "Difference from Predicted" figures in this table provide far less dramatic, but still supporting, evidence for the concept of narrowcasting. These differences may appear small but they are based on the averages of a large number of programmes and are highly statistically significant. They do therefore exist. The real importance or relevance of such differences, however, lies in the fact that they provide exceptions to established general scoring patterns and in that they support - i.e. they generalise from - the patterns illustrated in Table 6.12.

Programmes which might be targeted towards (or at least are viewed more by) men will also tend to be liked relatively more by such targeted viewers than by others who may also be watching (e.g. women). The average male AI for the 20% most male biased programmes is 1 point higher than might be predicted given male overall programme appreciation scoring patterns. For the 20% least male biased programmes (i.e. the 20% most biased towards female viewers), the average male AI is 2 points lower than might be predicted.

Patterns in the "Female Average AI" column of Table 6.13 provide no real evidence of a similar pattern. For example, female programme appreciation scores for the 20% most female biased programmes are no higher than might be predicted across all programmes in this sample.

**Table 6.13: Programme Appreciation and
Programme Audience Bias By Gender**

5 Weeks, Weeks of April 19/1993 through to May 17,1993
All Programmes with 50 or more responses (n=2,023 programmes)

	Average Rating		Male Avg %	Female Avg %	Male Avg AI	Female Avg AI	Overall Avg AI
	%	Std Dev					
All Programmes	8	7	47	53	70	74	72
20% "Most Male" Biased	5	6	66	34	71	73	72
20% "Quite Male"	8	10	52	48	71	74	73
20% "Even Male/Female"	10	9	46	54	72	76	74
20% "Quite Female"	9	5	40	60	70	75	73
20% "Most Female"	6	4	30	70	68	74	72
<i>Difference From 'Predicted'</i>					<i>Men:</i>	<i>Women:</i>	
20% "Most Male" Biased					+1	-1	
20% "Quite Male"					0	-1	
20% "Even Male/Female"					0	0	
20% "Quite Female"					-1	0	
20% "Most Female"					-2	0	

(c) Results from Similar Analyses for Other Segmentation Criteria

The next stage was to repeat similar analyses with respect to other segmentation criteria (i.e. age, social class, weight of viewing). The calculation and compilation of these various tables led mostly to null results similar to those above for female viewers. Particular programmes (or groups of programmes) that attract a high proportion of viewers from a certain age group, social class, or weight of viewing classification, are not on average enjoyed relatively more by these 'targeted' viewers than would be predicted given established overall programme appreciation scoring patterns by each audience segment. For age and weight of viewing, only two identified exceptions to this rule will therefore be briefly discussed in the next paragraphs. For social class (an important variable for advertisers) no further

discussion is necessary. Although relatively 'targeted' or biased programmes by social class (i.e. AB or CD biased programmes) could be found, there was no indication at all that such programmes were appreciated any more than might normally be expected by the targeted social group.

In line with the patterns outlined in Tables 6.12 and 6.13 for men, programmes with audiences biased towards either the youngest (12-15) or the oldest (55+) age groups are appreciated relatively more than might be predicted by such 'targeted' viewers than by others who are watching. In the case of the 40 most extremely biased programmes towards the youngest age group, this evidence of 'narrowcasting' was not too surprising since these programmes were exclusively mid-afternoon childrens' programmes and cartoons. There is no doubt that such programmes are developed and broadcast to directly appeal to this youngest age group and one would expect audiences for such programmes to be skewed towards children.⁵ One would also expect such specialised programmes to be appreciated relatively more by children than by older viewers who are also watching.

One other limited 'narrowcasting' exception highlighted in this analysis was for light viewers, especially for the forty programme sample of programmes most biased towards this viewing group (i.e. where 20 to 45% of the audience was comprised of light viewers in comparison to overall patterns where light viewers on average make up about 11% of the average programme audience). Such 'light viewer' biased programmes included a varied mixture of films, light

⁵ As outlined, in Table 6.1, the 12-15 age group makes up about 8% of the average audience for all programmes broadcast. (Viewers under the age of 12 are not included in this data set.) Even for the forty programmes with audiences most biased towards children, therefore, such children comprise only between 34 and 54% of the viewing audience. Older viewers from all age groups comprise a significant proportion of the audience even for childrens' programmes. This reiterates the point that television is a far less segmented mass medium than print, for example, where examples could readily be found of magazines that are read virtually exclusively by quite narrowly defined target audiences.

entertainment, information, and drama/art, and were ones exclusively broadcast on either BBC2 and Channel 4. Again, light viewers on average gave these 'targeted' programmes higher relative appreciation scores than would be expected given light viewers' overall appreciation scoring patterns. (While programme appreciation scores for heavy viewers were still higher than for light viewers, the normal five or so point spread between these weight of viewing categories was narrowed to two points for this particular light viewer 'targeted' selection of programmes.)

(d) Summary

Overall, this analysis of programme appreciation and audience bias has provided tantalising but very limited support for a 'narrowcasting' thesis. There are programmes that attract biased audiences and that are then liked more (in comparison to average appreciation scoring patterns) by the 'targeted' viewing segment than by others who are also watching. But is this the 'narrowcasting' pattern that media planners would like to see? Where is the evidence, even in extreme cases, of any programmes that are watched by very targeted and *highly* appreciative audiences? Only in the case of the programmes most highly targeted towards men, were particular examples found that more than slightly differed from the overall programme appreciation scoring patterns by audience segment first outlined in Table 6.5. For such highly male biased programmes, male programme appreciation scores were actually *higher* than female programme appreciation scores, even though women on average tend to give higher appreciation scores than men. When other exceptional patterns were found (i.e. for 'extreme' programmes biased towards light viewers), these only saw a narrowing in the normal range between the average programme appreciation scoring patterns of different

segments (i.e. a slightly narrower than normal difference between the average programme appreciation scores given by light and heavy viewers).

Although this analysis has been based on only the four mass audience UK channels, and not on supposedly more highly targeted cable and satellite channels, it does at least call into question much of the hype surrounding the future of television. Will today's mass audience channels be supplanted by highly targeted 'narrowcast' channels that are expected to survive because they reach 'premium' small, highly defined and highly involved audiences? Could a 'UK Woman' speciality channel, for example, be expected to attract dramatically more targeted audiences than those represented by the 40 highly biased female programmes outlined in this chapter? Even if such a channel did reach a narrow and highly targeted audience, does the existence of such targeted audiences necessarily also have to imply 'high involvement' or 'high appreciation'?

6.8 PROGRAMMES MOST LIKED BY DIFFERENT SEGMENTS

An exploratory follow-up to the above investigation took a slightly different approach by looking at the programmes 'most liked' by various possible audience segments (i.e. those with highest appreciation scores). Are such highly appreciated programmes (e.g. those programmes most liked by men) also equally liked by other viewers who are watching (e.g. women)? What patterns are there between the programmes 'most liked' by a particular audience segment and the 'audience bias' of these same programmes towards that particular audience segment.

Table 6.14: Programme Appreciation and Programmes "Most Liked" By Men

5 Weeks, Weeks of April 19/1993 through to May 17,1993
All Programmes with 50 or more responses (n=2,023 programmes)

	Avg Rating %	Male Avg %	Female Avg %	Male Avg AI	Female Avg AI	Difference Female AI - Male AI
All Programmes	8	47	53	71	75	4
40 Progs "Most Liked" by Men	11	46	54	82	82	0
40 Progs "Least Liked" by Men	4	43	57	53	63	10
20% "Most Liked" By Men Progs	10	49	51	78	79	1
20% "2nd Most Liked" By Men	8	48	52	74	76	2
20% "3rd Most Liked" By Men	7	48	52	71	74	3
20% "4th Most Liked" By Men	7	46	54	68	73	5
20% "Least Liked" By Men	8	43	57	62	70	8

Table 6.14 highlights typical results from this investigation across the various segmentation criteria, in this case summarising patterns for the programmes most appreciated by men. The patterns are clear but not all that strong, and these are the most dramatic patterns seen for any similar audience segmentation by gender, age, social class, or weight of viewing. Overall, programmes that are especially appreciated by one viewing segment also tend to be especially liked by other viewing segments. In this example, the 20% band of programmes most appreciated by men (i.e. an average male AI of 78) are also highly appreciated by women (i.e. an average female AI of 79). The 20% of programmes with the lowest male programme appreciation scores (i.e. 62) likewise get relatively low female appreciation scores (i.e. 70). In line with the general pattern that women tend to give slightly higher appreciation scores on average than men, even those programmes most highly appreciated by men still get either the same or slightly higher female appreciation scores on

average. As might be expected, the spread between average male and average female appreciation scores does increase as one moves down the “Difference Female AI - Male AI” column from those programmes most appreciated by men to those least appreciated.

In terms of the audience composition of these most and least male appreciated programmes, a slightly lower proportion of men are on average in the audiences of those programmes that achieve the lowest male programme appreciation scores. Even so, the audience composition of even the forty programmes most liked by men and the forty programmes least liked by men is not all that much different from the average audience composition of 47% men and 53% women for “All Programmes”.

6.9 SUMMARY

This chapter has investigated the relationship between the audience composition of different television programmes and the programme appreciation scores that they achieve. This has been done on several levels, both through looking at overall averages across different groupings of programmes by type and channel, and then by looking in more detail at particular sub-samples of programmes with quite ‘biased’ or skewed audiences.

To answer the questions first asked in the initial overview of this chapter, the audience make-up of programmes does vary in general and across different programme types and different channels. Such variation is more likely to be seen in terms of gender and age than of social class. On average across programmes these audience composition differences are smaller than is often

supposed. Given this relatively unsegmented nature of television audiences, it is established that little of the variation in programme appreciation scores outlined in Chapter 5 (“Programme Appreciation Scores”) is due to audience composition effects. Programmes that are highly appreciated by one segment of viewers also tend to be highly appreciated by all other segments of viewers who watch. An examination of ‘extreme’ programmes, where audiences are highly skewed towards particular segments of viewers, finds little overall evidence of programmes that reach quite highly targeted *and* especially appreciative audiences.

7. PROGRAMME APPRECIATION AND AUDIENCE SIZE

7.1 CHAPTER SEVEN OVERVIEW

In considering audience appreciation data as a possible supplementary media planning currency to traditional ratings, it is important to establish the extent of any relationship between these two types of data. This chapter looks in detail at the question of how far audience appreciation ratings for television programmes might be related to ratings of audience size. The results of previous research on this issue are first broadly summarised. A methodology for a comprehensive series of new replications and analyses is then outlined. Finally, results are presented in the 'step-by-step' sequence of the analyses undertaken. Finer details on the findings of earlier research are outlined and discussed wherever appropriate in the context of these new analyses.

7.2 MAIN FINDINGS OF PAST RESEARCH

The fact that BARB and its precursors (e.g. the Independent Broadcasting Authority's earlier AURA *Audience Reaction Assessment* system) have long collected both audience size and audience appreciation data immediately suggests questions about the relationship, if any, between the two. It is therefore perhaps surprising that comparative analyses using such UK data have been quite rare (Barwise, Ehrenberg and Goodhardt 1979; Menneer 1987a). Furthermore, the results found in these analyses have not always been completely consistent. Nevertheless, a limited Double Jeopardy relationship between audience size and programme appreciation has repeatedly been demonstrated: programmes with large audiences tend to be liked on average slightly more by their viewers than programmes with small audiences are liked

on average by their (fewer) viewers (Ehrenberg, Goodhardt and Barwise 1990). As will be outlined (the analyses to come in this chapter begin by directly replicating these main earlier studies), the key to identifying and isolating such relationships has been to make necessary allowances for the many factors that can have an independent influence on either audience size or on audience appreciation ratings for a programme. These factors include, for example, scheduling effects, channel effects, and programme type effects.

Beyond the above UK research based on BARB data, few other programme likeability studies have involved sufficient and appropriate data to explore this issue in any detail. Without accounting for various influencing factors, most other research over the years has simply concluded that there is little relationship between audience size and programme appreciation: programmes which achieve high audience size ratings audiences do not on average get especially high appreciation scores nor do they get especially low appreciation scores (Goodhardt, Ehrenberg and Collins 1975; Windle and Landy 1996). This general finding is typified in research carried out in the US by Television Audience Assessment (TAA 1983a). Here, an "Appeal Index" similar to the UK's audience appreciation measure was developed to provide a summary index of a programme's entertainment or enjoyment value. Looking across several hundred programmes, no significant relationship was found between the size of a programme's audience and the amount of satisfaction viewers derived from the show. With more limited sample sizes, quite contradictory results have even been found. Another U.S. study looked at the relationship between audience size and programme appreciation within a sample of only seventeen quite varied programmes and concluded that "programmes that have small audiences are more appreciated than programmes which have large audiences" (Hoffman 1984).

7.3 METHODOLOGY

The ‘a priori’ expectation in this thesis, based on the prior BARB-related research outlined above, is that when various influencing factors are isolated and accounted for, there will be a limited relationship between audience size and programme appreciation. The question is to what extent the results of any particular prior studies will be reconfirmed or refuted. Furthermore, in building on such past work, will the results found here suggest a more powerful relationship between audience size and programme appreciation that perhaps goes beyond the established Double Jeopardy phenomenon?

(a) Definition of Variables

The linear regressions to be presented in this chapter deal with a single dependent variable (programme appreciation) and a single independent variable (audience size).

The programme appreciation variable is referred to throughout by the now familiar shorthand of “AI” for Audience Index. For analyses based on various samples and subsamples of programmes (e.g. by programme type, by channel, by time, etc) such AI scores are calculated by summing the scores on the six point scale for all individual responses to that programme and then dividing this total by the number of respondents. In later analyses, looking at appreciation scores for particular audience segments across different programmes, the AI score will reflect the segmentation variable under consideration in each specific analysis, as clearly outlined at the head of each table. For example, the AI for a programme’s ‘female’ viewers will be

calculated by summing the individual scores of each female viewer of the programme and then dividing this total by the number of female respondents to that programme.

In previous studies looking at the relationship between audience size and programme appreciation, audience size has been presented either as the total number (i.e. millions) of viewers watching the programme (Menneer 1987a; Menneer 1987b) or as the percentage of the audience that is viewing (Goodhardt, Ehrenberg and Collins 1975; Barwise, Ehrenberg and Goodhardt 1979; Barwise and Ehrenberg 1982). The definition of the audience size variable used here, referred to in the following tables and equations as the “R” or “Rating,” is based on traditional percentage (i.e. share of audience) rating points. As was outlined in section 3.5 (d) of Chapter 3 (“Methodology”), ratings therefore represent the number of ‘viewers’ (i.e. panel members who responded) for each programme as a proportion of the total viewing population (e.g. the number of members in the viewing panel that week). In analyses looking at ratings for particular audience segments (e.g. ‘men’), the rating will be calculated as the number of ‘men’ viewing that particular programme divided by the number of ‘men’ overall in the viewing panel that week.

(b) The Analysis Approach Taken

With a much larger and more comprehensive data sample than has been utilised in the past to address the relationship between programme appreciation and audience size, this chapter replicates and extends considerably on previous analyses. The ‘step-by-step’ approach taken is designed to be illustrative, clearly presenting and accounting for patterns in

the ways a variety of programme and audience factors can exert influence on the relationships found between audience size and programme appreciation. Essentially, this represents a large-scale exercise in simple linear regression, where many hundreds of 'best-fit' lines have been plotted and systematically compared for programme and audience sub-samples that are progressively restricted along ever more narrowly defined criteria.

The various factors which might influence audience size ratings or programme appreciation scores will be outlined in detail as this chapter progresses. The important thing to note here, when looking for any supposed general relationship between audience size and programme appreciation, is that each of these factors may only have a primary impact on one side of the equation (e.g. on "audience size" or on "programme appreciation"). For example, a programme shown in the prime evening hours will obtain a higher rating or audience size than a programme broadcast in the early afternoon, regardless of how much each of these programmes is appreciated by its viewers. Likewise, programmes shown on the 'big' channels (BBC1 and ITV) will invariably achieve larger audiences than programmes shown on the 'small' channels (BBC2 and Channel 4). Beyond such *programme related* factors, Chapter 4 ("Individual Viewers' Appreciation Scoring Patterns") illustrated that there are *audience related* factors which could perhaps influence the programme appreciation side of the equation. For example, even though two programmes might have the same total audience size or 'rating', a programme with a high proportion of female viewers might still be expected to achieve a higher average appreciation score than a programme with a high proportion of male viewers.

On reading through the analysis and results sections to come, the step-by-step nature of this methodological approach will become readily apparent. In

essence, it is the basic equivalent of multiple regression, in that it accommodates multiple additional variables that can influence the key relationship between audience size and programme appreciation. Rather than developing a single complicated multiple regression equation through the creation of many new 'fitting' coefficients, however, this method follows a more laborious but otherwise straightforward approach to data reduction. (This is done by comparing the simple linear regressions - e.g. for men and for women - instead of by forcing a common slope or equation in a multiple regression analysis.)

The relationship between audience size and programme appreciation is first explored via a simple linear regression across all programmes in the sample. This leads to the inconclusive 'no relationship' result typified by most past research on this issue. From this starting point, subsequent simple linear regression analyses aim to explore the relationship between audience size and programme appreciation on a progressively more and more narrow basis aimed at comparing as far as possible 'like with like'. Increasingly narrow samples of programmes are first taken into consideration to allow for various programme effects (e.g. scheduling, channel, programme type, etc.). The relationship is then further explored by looking only within particular defined audience segments to try to allow for audience differences in the way viewers score programmes (i.e. the individual appreciation scoring patterns outlined in Chapter 4).

This step-by-step simple linear regression approach can be justified here for a number of reasons. Firstly, this research replicates and then builds directly on the findings of previous research. There are therefore clear prior expectations in many cases for the results that will be found. This supports the methodical approach followed here, rather than a more 'shotgun' type of

technique (i.e. multiple regression) that might otherwise have been employed in more exploratory primary research. Also, although audience appreciation scores for different programmes do vary, they mostly fall within a fairly narrow range of from about 60 to 80. If such audience appreciation scores were to be the dependent variable in a regression analysis involving many independent variables and dummy coefficients, the resulting multiple regression equation would be both complex and difficult to interpret. In contrast to stepwise regression for example, the approach followed here allows for richer explorations of the impact of each influencing factor.

(c) A Methodological Note on Best Fit:

The key methodological issue that arises in this chapter concerns the question of “How Good is the Best Fit?” (Ehrenberg 1982; Ehrenberg 1991). In essence, the analyses underlying this chapter involve calculating and comparing a great many ‘best fit’ least squares regression equations that have been fitted to multiple sets of data (e.g. different data subsets representing specific channels, programme types, audience segments, etc.). For each equation, it makes common sense that the scatter of individual data points should determine the best choice of the equation to be fitted. In practice, however, the fit of the best fitting line in a least squares regression analysis is little better than the fit of many other lines one could reasonably draw. Furthermore, in a thesis concerned with systematic and generalisable patterns, the accuracy of fit within a single set of data is secondary to considerations about how far a general result might hold across different groups of data.

The nature of Between Group Analysis will be discussed shortly. First, however, consider in greater detail the effectiveness of regression analysis at

picking out a ‘best’ solution for a single set of data. Professor Ehrenberg outlined his view on this issue in one of our recent thesis-related discussions:

“Regression equations in my view are pretty arbitrary. The precise coefficients depend for example on what fitting method is used; least squares is traditional but still arbitrary. For example, the great Laplace 200 years ago co-invented regression analysis, but with least ‘mean deviations’. The even greater Gauss co-invented it using least ‘squared deviations’. He justified the least squares approach by writing to Laplace: ‘At least my method is no more arbitrary than yours’. Wow!”

In plotting y against x , the least squares approach will provide a ‘best’ solution that minimizes the squared *vertical* distances between each data point and the regression line. (Quite different ‘best’ solutions could also be found by minimising the squared *horizontal* distances or by looking at a regression of x on y rather than y on x .) The minimum variance for this ‘best’ solution, however, may hardly be any smaller than that found for a variety of other ‘near-best’ equations with markedly different slopes. This is even more the case when, as in the analyses to be presented here, the correlation coefficients for the data are relatively low (i.e. correlations of 0.2 up to 0.6 between audience size and audience appreciation). With such data, even quite widely differing equations will hardly differ in the sizes of their residual scatter.

As outlined earlier, the step by step analysis process in this chapter involves looking at progressively narrower sub-samples of data to compare as far as possible like with like and to see if doing so leads to a stronger demonstrated relationship between audience size and audience appreciation. This raises the statistical issue of whether the many ‘best fit’ equations found for these increasingly narrowly defined data sets actually do differ from the earlier equations found for larger and more heterogeneous data samples (i.e.

heterogeneous in terms of both programme and audience related factors). But how does one compare different equations between groups to see whether they ‘really’ (i.e. statistically) differ from each other? Indeed, there are two basic questions to consider when comparing two (or more) straight line equations: Are the slopes of the lines the same or different, and are the intercepts the same or different? Depending on the answers to these questions, it may be determined that the lines are the same, that they are parallel, that they intersect, or that they are not parallel but also do not intersect.

There are several methods involving complicated test statistics that one could use to answer the questions outlined above and to compare different straight line regression equations (Kleinbaum, Kupper and Muller 1988). Given the realities of the least squares regression approach, however, is there really a need to test rigorously whether one ‘arbitrary’ coefficient in the data ‘really’ differs from another? In comparing two ‘best fit’ equations, for example, one might just as well be comparing two quite different equations which would also have virtually the same claim to being called ‘best fit’ solutions. The emphasis in this chapter does not therefore lie on the detailed accuracy or fit of any individual equation to a single set of data. Indeed, as with the data presented in tables throughout this thesis, rounding will be used extensively to facilitate visual comparisons across the slopes and intercepts of multiple equations. Rather the emphasis here is on Between Group Analysis (or BGA), which involves seeing whether or not there is one equation that might hold (at least to a close degree of approximation) for many sets of data.

This chapter will not present the reader with the hundreds of detailed regression equations that underlie this analysis. Instead, the aim will be to use

Between Group Analysis to simplify the presentation of results and the identification of important patterns and differences. To illustrate the BGA approach that has been followed, the early tables and findings in the following “Analysis and Results” sections will be outlined in detail. Later findings will then be presented in a more summarised format.

7.4 ANALYSIS AND RESULTS

Figure 7.1 shows a scatter diagram of the relationship between audience size and programme appreciation for more than 3,000 programmes broadcast during the five consecutive week period under consideration in this thesis.

Figure 7.1: Audience Size and Programme Appreciation

n = 3,015 programmes with 25 or more responses

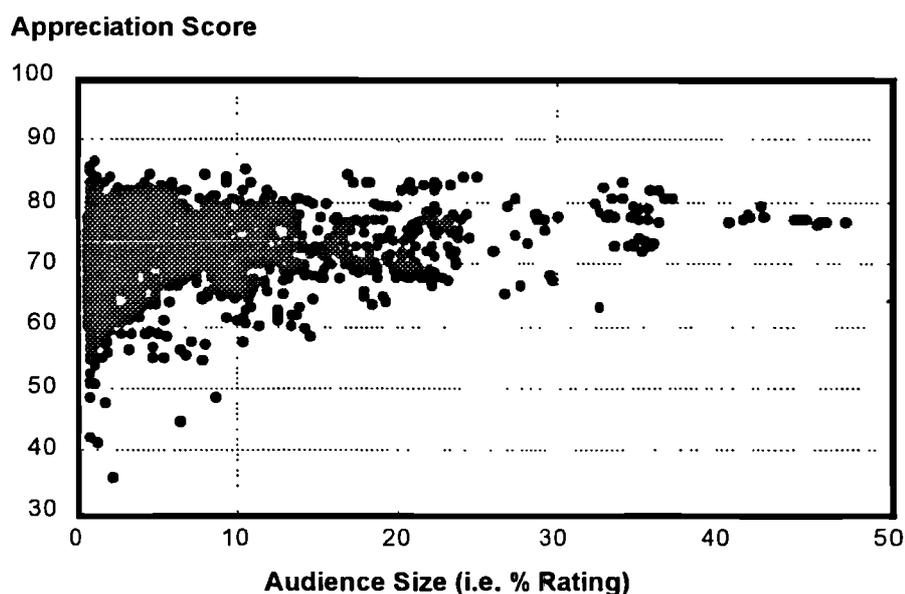


Table 7.1 then summarises a simple linear regression ‘best-fit’ equation based on this data sample. As has repeatedly been noted in the past, the high degree of scatter (correlation of 0.2) and near zero slope suggests that there is little

relationship whatsoever: programmes with large audiences do not on average get especially high appreciation scores, nor do they get especially low ones.

Table 7.1: The Overall Relationship

n = 3,015 programmes with 25 or more responses

	n Progs	Avg AI	Std Dev	Avg R	Std Dev	r	Equation
All Programmes	3015	72	5	5	7	0.2	$AI = 0.1R + 71$

Note: "AI" = Appreciation Score "R" = % Rating "r" = correlation

With the exception of the few earlier UK studies based on BARB data, this is the stage of analysis where most prior research looking at the relationship between programme appreciation and audience size has concluded. Most such studies have, however, been based on substantially smaller data samples.

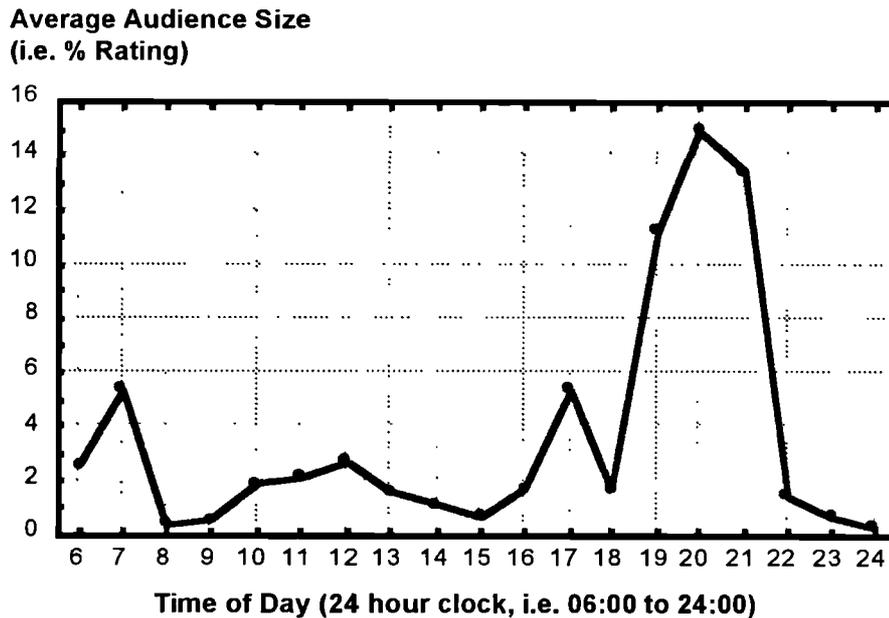
(a) The Relationship Allowing for Scheduling and Channel Effects:

It has been suggested that scheduling factors such as time of day and channel effects should be allowed for because they can greatly affect audience size without necessarily affecting programme appreciation (Barwise, Ehrenberg and Goodhardt 1979). One should therefore compare programmes shown at much the same time of day on the same channel, or at least on a channel of similar size and character (e.g. programmes on BBC1 and ITV, in general, attract larger audiences than do programmes on BBC2 and Channel 4).

Figure 7.2 illustrates the average shape of the 'viewing curve' for weekday audiences on BBC1 and ITV over the five week period under analysis. ("Time of Day" here represents programme starting times.)

Figure 7.2: Shape of the Viewing Curve Over the Day

n = 4,433 programmes shown on weekdays (Monday through Friday) on BBC1 and ITV during 5 consecutive weeks in April and May 1993



As illustrated, the number of viewers watching varies through the day. There are small peaks in viewing during the early morning, through lunchtime, and in the late afternoon (e.g. after-school). Viewing then drops off around dinner time (e.g. between 6:00 and 7:00 p.m.) before building to peak prime-time viewing hours between about 7:00 and 10:00 p.m. This viewing curve demonstrates that regardless of programme appreciation ratings, programmes shown during the evening will attract larger audiences on average than programmes shown during the morning, afternoon, or late-night hours.

Table 7.2 allows for the influence of such scheduling effects on the relationship between programme appreciation and audience size. To allow for time of day effects, analysis is first restricted from the more than 3,000 programmes summarised in Table 7.1 to the 965 programmes that were shown in prime time evening hours. To allow broadly for channel effects, programmes

shown on BBC1 and ITV are then separated out from those shown on BBC2 and Channel 4.

Table 7.2: Allowing for Time and Channel Effects

n = 965 prime-time evening programmes (with starting times between 7:00 p.m. and 10:00 p.m.) with 25 or more responses

Evening Programmes	n Progs	Avg AI	Std Dev	Avg R	Std Dev	r	Equation
All	965	72	6	9	9	0.3	$AI = 0.2R + 71$
BBC1 & ITV	595	72	5	12	11	0.4	$AI = 0.2R + 70$
BBC2 & Chan 4	370	73	6	5	4	0.4	$AI = 0.6R + 71$
BBC1	235	72	6	12	7	0.3	$AI = 0.2R + 69$
ITV	360	72	5	11	12	0.4	$AI = 0.2R + 70$
BBC2	190	74	6	5	4	0.4	$AI = 0.6R + 71$
Channel 4	180	73	6	5	4	0.5	$AI = 0.6R + 70$

Note: "AI" = Appreciation Score "R" = % Rating "r" = correlation

Given the fact that ratings of audience size are clearly otherwise influenced by audience members' availability to view at different times of the day, restricting this analysis to prime-time programmes only (e.g. row "All" in Table 7.2) does lead to a slightly improved correlation between programme appreciation and audience size than that shown in Table 7.1. (The correlation between audience appreciation and audience size shown earlier in Table 7.1 was 0.2. By now accounting for scheduling factors, and looking only at prime-time programmes, this correlation has increased to 0.3.)

The most apparent pattern in this table, however, is the distinction in rows two and three of Table 7.2 between programmes shown on the two bigger channels (BBC1 and ITV) and those shown on the two smaller channels (BBC2 and C4). As already discussed in Chapter 5 ("Programme

Appreciation Scores”), and as is now illustrated in the “Average AI” and “Average Rating” columns, programmes on the smaller channels on average achieve smaller audiences but higher appreciation scores than do programmes on the bigger channels. Although there is now slightly less scatter (i.e. a correlation of 0.4), allowing for such channel effects does not improve the still very flat relationship between audience size and audience appreciation for the bigger channels from that shown for “All Programmes”. For programmes on BBC2 and Channel 4, however, there is a significantly steeper slope (even though a great deal of scatter remains) indicating a stronger positive relationship on these channels between audience appreciation and audience size. A 10 point ratings difference between programmes on the smaller channels translates on average to a 6 point audience appreciation score difference. For the larger channels, a 10 point ratings increase will on average only lead to audience appreciation scores that are 2 points higher.

The final four rows in Table 7.2 next examine each channel separately. This confirms that the smaller channels should be considered separately from the bigger channels, but that the two lines for BBC1 and ITV are otherwise pretty much the same, as are the two lines for BBC2 and Channel 4.

(b) The Relationship Within Programme Types

After allowing for scheduling and channel effects, earlier research by the Ehrenberg Group next looked at the impact of programme type effects on the relationship between audience size and audience appreciation (Barwise, Ehrenberg and Goodhardt 1979). As with the programme classification system outlined in Table 3.4, programmes were broadly classified into two main types, *information* and *entertainment*. The findings were intriguing. For different

programmes of the same type, higher appreciation scores tended to go to those programmes with larger audiences. Between these two types, however, there was a negative correlation since information programmes tended to have smaller audiences but, for a given audience size, higher appreciation scores than entertainment programmes. Table 7.3 outlines two equations that summarised the relationships found through this research, which in this case was based on prime-time evening programmes that were broadcast on BBC1 and ITV.

Table 7.3: Earlier Research Allowing for Programme Type (1)

Barwise, Ehrenberg and Goodhardt (1979)
Data collected through the Independent Broadcasting Authority's AURA system

Programme Type	Equation
Entertainment Programmes	$AI = 1/3 R + 60$
Information Programmes	$AI = 1/3 R + 70$

Note: "AI" = Appreciation Score "R" = % Rating

These results led Barwise, Ehrenberg and Goodhardt to define a term they called "demandingness", which is determined by whether the programme "helped me to relax" or "made me think". The suggestion is that highly demanding (i.e. information) programmes require a commitment from the viewer but also give back more in return. Less demanding (i.e. entertainment) programmes do not require as much effort or deliver as much impact, and so would not be as highly appreciated. The theoretical interpretation is that the more demanding a programme is, the more interesting and/or enjoyable it must be (relative to more relaxing programmes) before people will watch it.

Peter Menneer of the BBC Broadcasting Research Department later replicated this analysis on further data from BARB's then relatively new Television Opinion Panel collection system (Menneer 1987a; Menneer 1987b). (As has been outlined, the BARB system which has been running since 1983 has a panel size of approximately 3,000. The Independent Broadcasting Authority's earlier AURA system, from which the Barwise, Ehrenberg and Goodhardt study above drew its data, was based on bi-weekly diaries that were sent to a panel of about 500.) Menneer also found that entertainment programmes on average had lower appreciation scores than information programmes. For entertainment programmes, he found the expected positive correlation between audience appreciation and audience size. For the demanding or information programmes, however, Menneer found a *slight* tendency for AI's to be lower amongst those programmes that generate large audiences. Table 7.4 outlines the two equations that summarised the relationships found in this analysis, again centred on prime-time evening programmes broadcast on BBC1 and ITV.

Table 7.4: Earlier Research Allowing for Programme Type (2)

Menneer (1987)

Data Collected through the BARB Television Opinion Panel

Programme Type	Equation
Entertainment Programmes	$AI = 0.65R + 68$
Information Programmes	$AI = -0.31R + 78$

Note: As usual, "AI" = Appreciation Score but the "R" or Rating here refers to "millions of viewers" rather than % share of audience.

Table 7.5 now replicates these earlier studies based on the five week sample of programmes being analysed in this thesis.

**Table 7.5: Allowing for General Programme Types
BBC1 & ITV**

n – 595 prime-time evening programmes (with starting times between
7:00 p.m. and 10:00 p.m.) with 25 or more responses

Evening Programmes BBC1 & ITV	n Progs	Avg AI	Std Dev	Avg R	Std Dev	r	Equation
All Programmes	595	72	5	12	11	0.4	$AI = 0.2R + 70$
Entertainment	295	72	5	17	11	0.5	$AI = 0.2R + 68$
Demanding	283	72	4	6	6	0.4	$AI = 0.2R + 71$

Note: “AI” = Appreciation Score “R” = % Rating “r” = correlation

The two general equations shown for “Entertainment” and “Demanding” programmes in Table 7.5 can be compared to those shown in Tables 7.3 and 7.4. In this regard, these findings would seem most similar to those established in the Barwise, Ehrenberg, and Goodhardt study. The lines for “Entertainment” and “Demanding” programmes in Table 7.5 are roughly parallel within the relevant range of possible ratings (the slopes in this table have been rounded and technically do differ very slightly), but with a higher average score for demanding programmes than for entertainment programmes. The slopes here, however, are somewhat less than the 1/3 shown for both equations in Table 7.3 and there is only a three point rather than a ten point spread between the intercepts for entertainment and demanding programmes. There is no evidence of the quite flat line Menneer found for demanding programmes which would have suggested that audience appreciation scores for such programmes do not vary with the ratings the programmes achieve.

It is of note that this is the third major study that has used BARB related UK audience appreciation data to look at the relationship between audience size and programme appreciation. All in all, Table 7.5 is consistent with the

conclusion in both earlier studies that demanding programmes will on average achieve higher appreciation scores than will entertainment programmes. The much smaller three point spread between the intercepts in this new research is not however consistent with prior results. Given the fact that all three of these studies have led to quite different findings, the most that the patterns outlined so far in this chapter seem to suggest is a limited general Double Jeopardy effect. Programmes with small audiences are liked slightly less by their relatively few viewers than more popular programmes are liked by their greater numbers of viewers. A sub-pattern to this is that demanding programmes appear on average to achieve slightly higher programme appreciation scores than entertainment programmes.

At this point, the direct replication oriented portion of this chapter is complete. In moving forward from here, analyses continue the step by step process of isolating programme and audience factors that might be disguising any stronger relationships that could exist between programme appreciation ratings and audience size ratings. Given the lack of definitive conclusions so far, after allowing for important scheduling factors and broad general programme types (e.g. entertainment and demanding), it would be folly to expect that strong and conclusive relationships between programme and appreciation and audience size will now suddenly appear. Instead, the expectation is for a great many null results, with occasional insights into areas and exceptions where relatively stronger or weaker patterns and relationships seem apparent. Between Group Analysis begins to play a significant role here in helping to identify such exceptions and in simplifying the presentation of results from what would otherwise be a great many slightly varied regression equations.

In continuing now with this examination of the relationship between programme appreciation and audience size, the question is whether allowing for broad programme types (e.g. entertainment and demanding) is enough to isolate the effect of programme type on the average appreciation scores that programmes achieve. As outlined in Chapter 5 (“Programme Appreciation Scores”), such average programme appreciation scores do vary across the programme types that comprise the general entertainment and demanding categories used above (e.g. the “Entertainment” category includes both relatively low scoring films and relatively high scoring light drama programmes). In trying to isolate the general relationship between programme appreciation scores and audience sizes, perhaps such differences should be allowed for. Table 7.6 therefore restricts the analysis to a finer breakdown of the entertainment and demanding categories by specific programme types.

**Table 7.6: Allowing for Specific Programme Types
BBC1 & ITV**

n = 595 prime-time evening programmes (with starting times between 7:00 p.m. and 10:00 p.m.) with 25 or more responses

Programme Type	n Progs	Avg AI	Std Dev	Avg R	Std Dev	r	Equation
Entertainment	295	72	5	17	11	0.5	AI = 0.2R + 68
Light Entertain	107	70	5	15	7	0.3	AI = 0.2R + 67
Light Drama	140	74	5	23	13	0.5	AI = 0.2R + 70
Films	22	71	5	11	8	0.6	AI = 0.4R + 67
Sport (*)	26	71	3	5	5	0.3	AI = 0.2R + 70
Demanding	277	72	4	6	6	0.4	AI = 0.2R + 71
Information	117	74	5	8	7	0.4	AI = 0.3R + 72
News	160	71	2	5	5	0.2	AI = 0.1R + 70

Note: “AI” = Appreciation Score “R” = % Rating “r” = correlation
 (* = relationship is not statistically significant at the p=.01 level)

With the exception of the “Sport” programme category, all of the ‘best fit’ relationships summarised by the individual equations in Table 7.5 are highly statistically significant. (Note that a line has not been fitted for Drama/Arts programmes under the “Demanding” heading since there was only an insufficient sample of 6 programmes of this type.) Concern here, however, is not so much with the fit of any particular line to its data set as with comparing these different lines across multiple data sets (e.g. across all the different programme types). With almost all of these slopes being 0.2 or 0.3, the most straightforward way to summarise this table is to say that regardless of programme type, a 10 point ratings increase would on average translate into only a 2 or 3 point higher programme appreciation score. (With a UK population of some 55 million, remember that a 10 point difference in ratings points between two programmes is large and implies an audience size difference of 5.5 million.) Otherwise, there are expected differences in intercepts that roughly reflect the programme type differences that were earlier outlined in Table 5.2. For example, the intercepts for light drama, sport, and information programmes are higher than the intercepts for light entertainment and films. The line for “News”, with its low correlation and near flat slope, is of special interest here as it suggests a very limited relationship between programme appreciation and audience size for such news programmes. This is more in line with Menneer’s findings for “Demanding” programmes as whole.

But do all these separate equations actually add ‘insight’ into the relationship between audience size and programme appreciation? Do these equations really (e.g. statistically) differ from each other or might one overall equation provide an effective summary for these different sets of programmes classified by type? Referring back now to earlier discussion in this chapter about ‘best

fit' simple linear regression equations and Between Group Analysis, consider the residual standard deviations for each of the lines that have been fitted in Table 7.6. In column (1) of Table 7.7, these residual standard deviations are calculated for each of the 'best fit' lines representing the relationship between programme appreciation and audience size for the six programme types in Table 7.6.¹ The residual standard deviation about the 'best fit' line for "Light Entertainment", for example, is 4.9. This says that most of the observed appreciation (AI) scores for light drama programmes lie within about 5 points on either side of this 'best fit' line. The "Average" line column (1) then gives the average of the residual standard deviations for these six individual regression lines (3.9).

Table 7.7 Residual Standard Deviations (rsd) for the Linear Regression Equations Outlined in Table 7.6

Programme Type	(1)	(2)
	All Equations	Fitting Avg Equation $AI = 0.2R + 70$
	rsd	rsd
Light Entertain	4.9	5.4
Light Drama	4.3	4.3
Films	4.2	4.6
Sport (*)	2.9	2.9
Information	4.9	5.8
News	2.3	2.3
Average	3.9	4.2

Note: "AI" = Appreciation Score "R" = % Rating "r" = correlation
 (* = relationship is not statistically significant at the $p = .01$ level)

Column (2) of Table 7.7 next follows a Between Group Analysis approach in presenting an equation for a common line ($AI = 0.2R + 70$) that is simply the

¹ Note that these residual standard deviations are based on the actual underlying data and 'best fit' equations, before any of the standard deviations, correlations, slopes and intercepts were rounded for clarity of presentation in Table 7.6.

average of the six individual lines calculated for the different programme types.² Note how the residual standard deviations for the common line in column (2) are mostly only marginally higher than the residual standard deviations for the ‘best fit’ lines in column (1). The average of the residual standard deviations for each programme type about the common line is 4.2, which is only about eight percent more than the 3.9 average of the residual standard deviations for each of the six different ‘best fit’ lines. All of this suggests that for all practical purposes, this single common line is a ‘near best’ fit across all programme types. Rather than trying to interpret and synthesize a relatively complicated table of equations like Table 7.6, the relationship between programme appreciation and audience appreciation for different programmes can now be briefly summarised as: ‘For all programme types, a 10 point ratings increase will translate into approximately a 2 point higher programme appreciation score.’ If necessary, to make some further comment about the very limited exceptions where the fit of the common line in Table 7.7 is less close to the ‘best fit’ line, one could add that ‘for a given audience size, information programmes will on average achieve a slightly higher appreciation score (e.g. by one or two points) than the average programme type, while light entertainment programmes will achieve a slightly lower appreciation score than the average programme type’.

(c) The Relationship Allowing for Audience Differences

This section now concludes this chapter by briefly summarising the largely null results resulting from an extensive series of similar Between Group Analysis investigations to that outlined above. These were aimed at

² The slope coefficient is calculated from the extreme means of the Appreciation Scores and the % Ratings for these seven data sets, and the intercept coefficient is calculated by making the line go through the overall means of the data [Ehrenberg, 1991 #317].

summarising a wide variety of simple linear regression equations (similar to those in Table 7.6) which narrowed further the examination of the relationship between programme appreciation and audience size through allowing for the potential effect of audience differences. Where two programmes with the same audience size rating obtain different programme appreciation (AI) scores, this may be caused by viewers liking one programme better than the other. Alternatively (or in combination), different programme AI scores may be due to programmes attracting somewhat different audiences with different scoring patterns (e.g. a programme may attract a high proportion of heavy viewers who tend to score programmes more highly).

The analyses carried out here again had the aim of comparing as closely as possible 'like with like'. Audience effects were therefore allowed for by continuing with the simple linear regression approach, but by investigating the relationship between audience size and audience appreciation exclusively within specific audience *segments* (e.g. by gender, age, social class, and weight of viewing). For example, two versions of Table 7.6 could be constructed with the first presenting (for the various programme types) equations summarising the relationship between 'male' programme appreciation scores and 'male' audience sizes, and the second presenting similar equations for 'female' programme appreciation scores and 'female' audience size ratings. Six versions of the table would then be needed to look at whether patterns varied by age, and so on for social class (four tables) and weight of viewing (three tables). With scheduling effects, possible programme type effects, and possible audience differences all taken into consideration, such a restricted analysis would permit the purest possible exploration of how programme appreciation might vary with audience size.

Overall, even in allowing for so many possible influencing factors, almost identical and very limited Double Jeopardy relationships between audience size and appreciation were found within different genders, social classes, and weight of viewing groups. Only with respect to age, did a hint of an interesting pattern did emerge. The ‘best fit’ linear regression approach suggested a somewhat stronger relationship (less scatter, with correlations of up to 0.6 and steeper regression lines), for younger viewers within more demanding programmes and for older viewers within light entertainment programme types. For demanding programmes, for example, ‘Age 25-34’ programme appreciation scores and ‘Age 25-34’ audience size ratings were quite highly correlated, while Age 55+’ programme appreciation scores and ‘Age 55+’ audience sizes showed little correlation and a near flat regression line.

7.5 SUMMARY

This chapter suggests that audience appreciation does measure something different from audience size (i.e. the two measures are in no way perfectly correlated). In the most closely defined attempts to fit a single simple linear regression equation to compare ‘like with like’ (e.g. looking at a particular time, on a particular channel, within a particular programme type, and amongst similar categorisations of viewers), a correlation of up to about 0.6 can be found between audience size and audience appreciation. The highest correlations and the most steeply sloped ‘best fit’ lines are mostly found for “Light Entertainment” programmes. The lowest correlations and the least sloped ‘best fit’ lines are found for more demanding and homogeneous programme types like “News.” There is however a wide degree of variation

(scatter) in this programme appreciation data and AIs fall within quite narrow limits (mostly varying between about 60 and 80). Given these realities, the Between Groups Analysis approach has demonstrated that a common line can often be fitted with ‘near best fit’ accuracy to summarise what at first glance might seem to be quite different relationships (e.g. as just outlined for “Light Entertainment” and “News” programmes).

Overall, a definite Double Jeopardy pattern does emerge (e.g. bigger programmes are liked more) when one makes allowances for scheduling differences, but this is not all that strong a relationship (i.e. a substantial 10 point increase in audience size ratings for a programme will only lead on average to a 2 point increase in programme appreciation scores). Little further is added to this limited Double Jeopardy pattern when allowances are made for programme differences and then for audience differences. Such generally weak patterns may help to explain why repeated studies over the years have often arrived at slightly different conclusions and have not been able to establish (beyond Double Jeopardy) any stronger, generalisable relationships between programme appreciation and audience size.

8. AUDIENCE APPRECIATION AND REPEAT VIEWING

8.1 CHAPTER EIGHT OVERVIEW

This brief chapter focuses on repeat viewing patterns for regular television series. A repeat viewer is someone who sees a programme in one week and then also watches it again the following week. Do individuals who are frequent viewers of a particular programme give it higher appreciation scores on average than infrequent viewers? How do average repeat viewing levels for television programmes relate to the appreciation scores achieved by these programmes? The tables presented here are replication oriented, and the prior expectation is that this chapter will broadly confirm patterns that have been established in the past. Nevertheless, this analysis is important and relevant. The extensive individual-level data available here allows for a more direct and in-depth investigation of certain repeat viewing relationships than was possible given the more limited data samples available for many past studies. Even if the patterns outlined in this chapter are not ‘new’, they are a powerful confirmation of more tentative findings suggested in earlier research.¹

8.2 MAIN FINDINGS OF PAST RESEARCH

The concept of repeat viewing was touched on briefly in Chapter 2 (“Introduction to the Literature”). In relation to general television audience

¹ The brief nature of this chapter is in keeping with recent arguments encouraging the core marketing journals to give space to publishing replication studies (Monroe, 1992). The write-up of replication and extension oriented results does not need to be lengthy, especially if the findings of earlier research do broadly ‘replicate’. Such results “can be reported extremely briefly as a one to two page research note: this will help us to establish EGs (Empirical Generalisations) at minimal cost in terms of journal space” (Barwise, 1995).

viewing behaviour, the Ehrenberg Group's research has repeatedly established that despite the repetitive nature of television programming, viewers are not highly loyal to specific programmes from week to week. Typically far less than half the viewers watching a prime-time series in one week, will watch it again the following week. Repeat viewing is nonetheless related to audience size in that there is a Double Jeopardy relationship where programmes with higher ratings achieve *somewhat* higher repeat viewing levels.

Beyond these well established general repeat viewing patterns, intermittent research by the Ehrenberg Group and by other researchers has explored the relationship between repeat viewing and audience appreciation for programmes. Early research by Aske Research Ltd. on behalf of the Independent Broadcasting Authority examined 11 programmes which were broadcast at regular times from week to week. Viewers were sampled in one week and then again two weeks later (due to the nature of the panel collection system used at that time). Repeat-viewers were defined as those who had seen a programme in the first week, and who also saw the episode two weeks later. For each programme, the general pattern found was that the higher the AI score given by an individual, the more likely the person was to be a repeat viewer (Aske Research Ltd. 1973a). A follow-up to this study looked at further situations, such as where different episodes of the same programme were shown at different times or on different days of the week (e.g. Coronation Street). A similar repeat viewing pattern was apparent. Viewers who saw both episodes gave them higher appreciation scores than viewers who saw only one episode (Aske Research Ltd. 1973b).

Some substantiation for the above conclusion was found in Television Audience Assessments (TAA's) development and testing of an 'Appeal' index

in the United States (TAA 1983a; TAA 1984b). Although repeat viewing could not be explicitly measured in this laboratory study (based on 470 female household heads), participants who gave a programme a high appeal score were more likely to *say* that they usually made a special effort to watch the show and that they had frequently watched the series before. TAA therefore saw the Appeal Index as being a very useful construct to advertisers and programmers due to its association with people's loyalty to the programmes they watch.

Further research by the Ehrenberg Group was based on two parallel one-week experimental field studies (a diary study and a questionnaire study). These studies collected liking and viewing information both for specific episodes of regular series and for the series themselves (Barwise and Ehrenberg 1982; Barwise and Ehrenberg 1987). A limitation to this research was that where frequency of viewing was measured (in the questionnaire survey) it measured *claimed* and not actual viewing. For example, viewers were asked to recall how many of a series of five episodes they may have watched. For various reasons, the nature of this type of questioning led to respondents severely overestimating their frequency of viewing by a factor of as much as 2.5 times. Despite such overclaiming, consistent patterns with earlier analyses were found. How much individuals liked a particular *series* correlated with how often they said they saw it. In the eight day period covered by the diary study, there were twenty cases where more than one *episode* of the programme was shown. Here, there was a consistent tendency for those who saw more episodes to give them higher average liking scores.

As BARB Television Opinion Panel members should only give an appreciation score to a programme if they have actually watched it, the data in this thesis does reflect a direct individual level measure of actual frequency of repeat

viewing over this five week period. Furthermore, as this diary data is collected on a weekly basis and respondents often fill in the diaries on a day-by-day basis, there are none of the overclaiming problems associated with asking people to recall their viewing patterns over a series of weeks. Given the size of this data sample, it is possible to track repeat viewing patterns for many hundreds of programmes and, especially in the case of programmes with relatively large audiences, there are quite large samples from which to generalise.²

8.3 APPRECIATION & FREQUENCY OF VIEWING

Chapter 4 (“Individual Viewers’ Appreciation Scoring Patterns”) showed that individual viewers can vary greatly in the appreciation scores they give to a particular programme. Table 8.1 now highlights the strong relationship between how much different viewers say they appreciate a programme and how often they watch it.

² As noted in Chapter 3 (“Methodology”), in light of the fact that this is panel data where the same individuals are responding week after week, it is surprising that the opportunity of tracking such individual level patterns in this data has not been exploited to any degree in the past. Instead, most prior audience appreciation studies based on BARB data have used only already aggregated programme level data.

Table 8.1: Appreciation and Individual's Frequency of Viewing

n=35 regular series, with one weekly episode in
each week of this five week period
BARB Weeks 16-20 1993

	Frequency of Viewing (Out of 5 weekly episodes)				
	1/5	2/5	3/5	4/5	5/5
999	79	81	82	86	85
Wildlife 100	78	79	81	82	84
Taggart	75	79	80	82	84
The Fresh Prince of Bel Air	74	80	82	81	85
Sunday Grandstand	73	74	73	72	79
Casualty	73	76	76	80	80
Public Eye	73	67	80	76	86
Grandstand	72	72	70	77	78
Desmond's	72	76	74	81	79
Highway to Heaven	71	75	75	79	81
The Cosby Show	71	78	75	79	84
Tomorrow's World	71	71	72	73	79
Watching	70	73	77	76	81
40 Minutes	70	72	71	79	83
Question Time	70	72	71	76	79
Birds of a Feather	70	71	74	77	82
The Addams Family	70	72	75	75	78
London's Burning	69	74	74	78	80
The Chief	69	73	77	79	79
Crime Stalker	69	70	77	74	76
Masterchef	69	72	75	78	83
The Money Programme	69	72	72	68	75
Growing Pains	67	73	76	81	79
Standing Room Only	67	72	78	77	84
Points of View	66	67	69	73	74
TV Heroes	65	67	66	67	72
Living Islam	65	67	79	78	77
Gingerbread Girl	63	68	69	67	74
Jim'll Fix It	62	67	72	71	83
The Comic Strip Presents...	62	66	67	68	74
Surprise Surprise	62	65	72	75	78
Conjugal Rites	61	67	68	72	75
Just a Gigolo	60	66	69	71	73
Top of the Pops	59	59	60	63	66
Cluedo	57	61	64	71	68
Average Appreciation Score	68	71	73	75	79

The programmes in Table 8.1 are a sample of 35 regular series, with episodes that were broadcast once each week, at the same time and on the same day, during this five week period. Viewers who saw only one episode of the programme gave it an average appreciation score of 68 while those who saw five out of five episodes gave it an average score of 79. Similar analyses were repeated for many hundreds of other programmes, including programmes with several multiple episodes during a single week (i.e. Coronation Street or other soap operas) and programmes that are shown regularly but at different times of day. In all cases, frequency of viewing is strongly correlated with individuals' appreciation.

Table 8.2 shows the same sample of 35 series as Table 8.1. This time the programme titles are arranged in decreasing order by the number of individuals who saw the programme at least once during this five week period. This table illustrates the reality that for a regular series of say 10 or 20 episodes over a season, few people will have seen all or nearly all of them. On average only six percent of the viewers of any regular series in this table will have seen all five out of five consecutive episodes. Forty-eight percent of the individuals who watched the average series at least once over the course of this five week period will have seen it only the one time.

Another pattern that is quite clear in Table 8.2, in comparing series in the first half of the list to those in the last half, is that popular television series (i.e. with large audiences) have more frequent viewers than less popular series. This is yet another example of Double Jeopardy. Smaller audience programmes lose out not only in that they have smaller numbers of viewers, but also in that those viewers will watch the programme less often and will tend to give it a lower appreciation score on average than will viewers of larger programmes.

Table 8.2: Individuals' Frequency of Viewing

n=35 regular series, with one weekly episode in each week of this five week period
BARB Weeks 16-20 1993

Programme Title	N who saw at least once	Frequency of Viewing (% who viewed one out of five episodes. two out of five episodes, etc.)					
		1/5	2/5	3/5	4/5	5/5	
Taggart	1691	% 20	17	18	21	24	
Birds of a Feather	1613	% 46	22	16	10	6	
London's Burning	1448	% 40	22	16	14	8	
999	1444	% 39	22	16	13	10	
Surprise Surprise	1440	% 38	19	17	15	12	
The Chief	1324	% 29	20	20	19	13	
Conjugal Rites	1288	% 33	20	18	16	13	
Cluedo	1254	% 43	22	15	12	8	
Casualty	1225	% 42	22	17	12	7	
Masterchef	1186	% 38	21	18	14	9	
Watching Wildlife 100	1030	% 49	22	14	9	6	
Gingerbread Girl	1022	% 46	21	14	12	7	
Jim'll Fix It	1016	% 40	20	14	15	11	
Growing Pains	1001	% 50	24	14	8	4	
Top of the Pops	961	% 49	21	14	9	6	
Just a Gigolo	949	% 47	23	17	12	2	
Sunday Grandstand	801	% 49	20	13	12	6	
Tomorrow's World	794	% 66	22	9	3	1	
The Comic Strip Presents...	768	% 54	17	11	11	8	
TV Heroes	695	% 53	23	15	8	2	
Grandstand	609	% 63	21	10	4	3	
40 Minutes	598	% 59	19	12	8	3	
Points of View	591	% 72	19	6	2	1	
The Fresh Prince of Bel Air	574	% 56	20	11	9	4	
Question Time	519	% 35	26	17	13	9	
The Addams Family	488	% 44	22	15	14	5	
The Cosby Show	373	% 42	25	14	12	7	
Desmond's	283	% 56	19	15	6	4	
Standing Room Only	266	% 58	21	10	7	5	
The Money Programme	259	% 51	20	15	9	6	
Living Islam	216	% 56	22	12	7	2	
Public Eye	182	% 59	19	10	8	3	
Highway to Heaven	176	% 68	18	10	3	2	
Crime Stalker	173	% 52	24	11	9	4	
Crime Stalker	156	% 55	15	15	12	3	
Average			48	21	14	11	6

8.4 APPRECIATION AND REPEAT VIEWING

Reflecting the frequency of viewing patterns outlined in Tables 8.1 and 8.2, Table 8.3 considers the relationship between repeat viewing and audience appreciation for the average of these 35 regular series. The “Week 1” and “Weeks 1 & 2” columns suggest typical repeat viewing patterns that have been found in the past. For the average programme title in this sample, Forty-eight percent of the viewers who watched in Week 1 will watch the programme again the following week. The average appreciation score given by all viewers who watched the first episode of the programme in Week 1 was 73. Those viewers who repeat-viewed the following week were individuals who gave higher than average scores (75) to the first episode. The remaining three columns of this table continue to follow the repeat viewing patterns of these Week 1 viewers over subsequent weeks. Fourteen percent of Week 1 viewers went on to watch the programme again in each of the following four weeks. These heavy repeat viewers had given very high average appreciation scores of 79 to the first week’s episode.

Table 8.3: Programme Appreciation and Repeat Viewing

n=35 regular series with one episode each week over this five week period
(Summarising a total of 33,002 individual appreciation responses to these 35 programmes)

	Viewers of the Average Programme in Week One Who Also Repeat Viewed in:				
	Week 1	Weeks 1 & 2	Weeks 1 to 3	Weeks 1 to 4	All Five Weeks
Average Repeat Viewing Patterns	% 100	48	30	22	14
Average Appreciation Score Given by Viewers in Week 1	73	75	77	78	79

8.5 SUMMARY

The patterns outlined in this chapter show a marked relationship between audience appreciation and frequency of viewing or repeat viewing. These replication oriented analyses both summarise and provide a powerful confirmation of the more tentative and limited findings of past research. Such a general relationship between appreciation and repeat viewing may seem intuitive (i.e. one would expect viewers to more frequently view programmes that they like more than other programmes). This chapter has helped quantify the extent of such a relationship, however, and has also served to demonstrate the extent to which audience appreciation ratings can discriminate such viewing behaviour patterns. In this regard, appreciation ratings can be considered as a predictor of repeat viewing patterns. For media planning purposes, advertisers may wish to have their ads carried by programmes which achieve higher appreciation scores. This is because viewers of such programmes are somewhat more likely to be in the audience from week to week (and are therefore more likely to be repeatedly exposed to any advertisements that might be rebroadcast).

9. DISCUSSION AND CONCLUSION

9.1 CHAPTER NINE OVERVIEW

This chapter provides a short integrated summary overview of the audience appreciation patterns established and outlined in the previous five analysis chapters. Following this general discussion, the appendix to this thesis is introduced, the limitations of this thesis research are raised, and some potential directions for future research are discussed. Finally, a summary conclusion closes the thesis.

9.2 SUMMARY OF OVERALL FINDINGS

In today's rapidly evolving television environment, there is a great deal of current media and advertising industry interest in possible 'qualitative' television audience research measures that might complement or add insight to more traditional ratings of demographics and audience size. This empirically based thesis research has been about identifying and describing patterns of *audience appreciation ratings* for television programmes. The research approach taken has involved a varied combination of replication work, new extensions to prior research, and completely new analyses.

The overwhelming challenge in this research has been in dealing with the general 'sameness' of the data. As has been discussed in detail, the audience appreciation scale is extraordinarily coarse, especially since 'viewers mostly say they quite like what they watch' and so only three or four points on the six point scale ever get much use. Many of the analyses carried out here have

led to either limited or null results. This was not at all unexpected and as the old adage goes, “null results are not necessarily dull results”. The confidential nature of this data has led to a great deal of speculation, discussion, and inferred ‘truth’. There is therefore a real need for more general understanding on how to evaluate and interpret this type of attitudinal data, particularly from perspectives beyond the current internal usage of this data by the broadcasters who commission it.

The main findings and conclusions from this research can be summarised as follows:

1. Replication analyses have confirmed that people do hold opinions about the television programmes they watch and that specific individuals can vary considerably in their average audience appreciation scoring patterns. New analysis suggests that the most important predictors of an individual’s average scoring patterns will be age and gender. Older viewers and female viewers are more generous in their scoring patterns on average than younger viewers and male viewers. Social class and weight of viewing are less effective predictors of whether individuals will be classified as high or low appreciation scorers.
2. As has also been demonstrated in previous research, television programmes mostly tend to be quite liked by their viewers, with most programme appreciation scores falling within a fairly narrow range of from about 60 to 100 on the possible scale of 0 to 100.
3. Most of the variation in programme appreciation scores is due to the programmes being liked differently (e.g. light drama programmes are

liked relatively more by *all* viewing segments) and not to their attracting differently scoring audiences. On the whole, audience composition does not vary much across different programmes, programme types, and channels. Where the audience profiles of programmes do differ, they are more likely to show variation in age and gender than in social class or weight of viewing. In developing the high/low scorer and varied/consistent scoring pattern categories in Chapter 4, and then applying them in Chapter 6, an important new contribution of this thesis was to illustrate in detail how little programmes tend to differ in terms of the overall average appreciation scoring patterns of the different audiences they attract.

4. Another entirely new contribution of this thesis was its exploration of relatively 'extreme' programmes which do attract audiences that are quite biased towards particular audience segments. Although there were a great many null results, the findings show limited support for a 'narrowcasting' thesis in that there are programmes that attract 'targeted' or relatively biased audiences (e.g. heavily skewed towards male viewers). Furthermore, the particular audience segment attracted by such a programme (e.g. men) will also like the programme relatively more on average than other viewers who are watching (e.g. women). Such examples are very rare, however, and tend to be noteworthy only at the extremes (e.g. for the relatively few programmes that are very heavily biased towards male or female viewers).
5. Audience size varies a great deal across programmes. There is only a weak positive Double Jeopardy relationship between audience size and programme appreciation, however, as has also been demonstrated in past research through controlling for other factors (e.g. scheduling,

type of programmes). The extensive new and replicated analyses underlying this thesis have aimed at an even finer controlling for the various factors that can independently influence either audience size or programme appreciation. Again, many null results are found in that the relationship between programme appreciation and audience size still remains weak, with much scatter. There is certainly no evidence, however, that there are programmes with low ratings which are especially liked by all those watching them. Overall, on average, a very large ten point increase in average audience size ratings (i.e. from a 5% rating to a 15% rating, equivalent for the UK to an increase in total audience size from about 3 million viewers to 8 million viewers) will correspond with only about a 2 point average increase in programme appreciation ratings (e.g. say from an average score of 70 to 72).

6. There is a very clear relationship between how often someone watches a programme and how much they say they appreciate it. Given the large and varied data samples available in this thesis, this finding (illustrated by typical examples in Chapter 8) provides strong confirmation of past findings which were often based on quite limited data and on self-reported measures of viewing frequency that were subject to severe overclaiming. Programmes that achieve higher appreciation scores will on average have more frequent viewers than will less appreciated programmes.

What implications do the findings in this thesis have for advertisers and media planners? Is there a potential role for appreciation ratings as a supplementary media planning measure to traditional audience size ratings? To begin answering these questions, it is important first to think about something that has not been directly measured or referred to in this thesis: advertising

effectiveness. Traditional television ratings provide estimates of total audience size and of the audience's demographic profile, but little indication as to the proportion of this potential audience that is actually looking or listening. Should not the real target for advertisers be those people who are actually present *and* paying attention when a commercial is broadcast? In this regard, do highly involving programmes result in greater advertising effectiveness than relatively uninvolved programmes? To what extent can measures of audience appreciation be used as an indication of viewers' involvement when viewing a programme?

The above questions are outside the scope of this thesis but this research does have implications in line with how such constructs (e.g. involvement or appreciation) have been measured or defined in other research studies that have looked at programme context effects on advertising effectiveness. For interested readers, the appendix to this thesis therefore provides a literature review, surveying the long history of marketing research into the relationship between programme involvement and advertising effectiveness. This frames and leads into a discussion of some further implications of this thesis research in line with audience appreciation data potentially being interpreted and used for media planning purposes.

9.3 LIMITATIONS

In some ways, the biggest limitation with this thesis might seem to be the relative 'sameness' of the data that has been analysed. As discussed in Chapter 3 ("Methodology"), however, much prior research lies behind the development of this audience appreciation scale and this thesis has therefore

not involved any new scale-development work. Instead the emphasis has been on trying to better understand and interpret patterns in the data that this scale collects.

Nonetheless, I have spent much time over the course of my research pondering how one might get more discrimination in the top part of the scale (the only part that respondents really use). Despite much thought, however, the fact is that I just don't know. Having immersed myself in this data for a number of years, my *opinion* now is that the limited 60-80 range found for most programme appreciation scores should not imply a critique of the scale (i.e. this 'sameness' in audience appreciation data is simply what one should expect - it does not necessarily mean that the scale is not teasing out some wider variability that actually exists in viewers' appreciation).

The key to resolving the 'limitation' of this scale then is to accept why these data patterns must be so whatever scale one might use to measure appreciation or liking. Audience appreciation data is all about viewers' attitudes to programmes they freely *choose* to watch. In this regard, it is natural that viewers should mostly 'quite like' what they have chosen. Why would viewers freely choose to watch programmes that they do not like, and why should many, if any, programmes therefore be expected to achieve low appreciation scores (i.e. much below 60)? Furthermore, in a generally low involvement media, it would be unlikely for all that many programmes to achieve very high (far over 80) appreciation scores. The fact that systematic (i.e. replicable or generalisable) patterns can still repeatedly be found in such relatively 'flat' data suggests that even very small differences in programme appreciation scores are real and significant (i.e. such as the one or two point differences between a score of 72 and seemingly similar scores of 70, 71, 73, and 74). In this sense of establishing significant and real differences through

replication, the whole process of generalisability (which underlies much of the analysis carried out here) can be seen as a form of sensitivity analysis.

I stress once again that the data I have examined here is what UK industry has routinely collected at great cost for many years. While this research has taken the measurement scale as a 'given', I doubt whether anything much better (e.g. more sensitive) could be devised and yet be usable for routine practical application. The insensitivity of the AI results is the nature of the beast (although this is not widely accepted). It is not just due to an insensitive measurement device, but that there is almost certainly nothing much there to measure. That in itself is very important.

There are data collection issues with the Television Opinion Panel diary collection process, however, that could have implications for the findings contained in this thesis. These diaries are distributed and collected on a weekly basis. There is therefore no understanding of exactly when they are filled out by individual respondents. What proportion of respondents carefully fill in these diaries on a daily or nightly basis? What proportion then leaves this chore until the end of the week, when the reliability of what they might be reporting could be subject to memory or frequency of viewing over-claiming effects? The audience size estimates that can be derived from these audience appreciation ratings suggest that such over-claiming must be relatively minimal. Nevertheless there is potential for automatic reporting of a programme that "I always watch" when this may or may not have been so for that particular week. Similarly, perhaps some of the lack of variation in programme appreciation scores could come from boredom effects, with respondents repeatedly checking the same score for multiple programmes in order to speed their diary filling chore.

A final limitation of this thesis, especially in looking at issues of ‘narrowcasting’ and relative audience bias, is that the data here was only for the four main UK broadcast channels. As has been repeatedly demonstrated in this thesis, television broadcasting remains a quite unsegmented medium. If audience appreciation data could have been analysed for relatively more narrowly targeted cable and satellite channels (e.g. MTV), would more conclusive results have been found in this “audience appreciation and audience bias” investigation? (Note that even today’s smallest and most highly ‘targeted’ cable channels will still be relatively unsegmented in relation to the highly targeted readerships that can be reached by print media options. Note also that collecting audience appreciation data for the smallest channels and smallest audience programmes would be a very costly undertaking as a very large panel of respondents would need to be surveyed in order to find a sufficient sub-sample who would have watched each particular small-audience programme.)

9.4 DIRECTIONS FOR FUTURE RESEARCH

In many ways, this thesis represents a ‘snapshot in time’ of what will be an ongoing investigation, both through continued analysis of this data set and through investigations based on new data and theory. Some currently planned avenues for future research based on this data set include:

1. The week by week individual level response data available here still leaves opportunities for continued research looking at repeat viewing patterns. In particular, an interesting issue I am beginning to investigate is to look at the effects of competing programmes and viewers’ availability to view.

If a viewer watches a programme one week and then does not repeat view the following week, is this because he or she chose to watch something else, or is it just because he or she was simply not available to watch television at all at the same time the following week? The possibility that a viewer was simply unavailable to view the following week does not necessarily imply that this was because he or she did not 'like' the programme. In choosing to watch something else the following week, however, this could likely mean that the viewer switched to something he or she 'liked more'.

While the discussion in the above paragraph may seem intuitive, it is not at all obvious what impact allowing for these two possibilities might have on the strength and extent of the patterns established between audience appreciation and repeat viewing (as summarised in Chapter 8). By removing from the analysis viewers who did not repeat view simply because they were unavailable (and therefore not necessarily because they did not 'like' the programme), it will be interesting to *quantify* just how much stronger a relationship can be expected between how much different viewers say they appreciate a programme and how often they watch it.

2. The 'narrowcasting' issue was investigated in this thesis by looking at both the relative audience bias of the composition of different programmes, and at programmes that were 'most liked' by different audience segments. An extension of this is to classify individuals based on their 'most favoured programme type' or 'most favoured channel'. Are viewers who concentrate most of their viewing on a minority programme type or on a minority channel also highly 'involved' or

appreciative in comparison to average overall individual appreciation patterns for these programme types and channels?

3. A 'case study' approach could be taken to look at a few particular programmes in much greater depth and detail. For example, "Eastenders" could be analysed in detail in accordance with the principles and patterns outlined in this thesis. Given that this programme has multiple weekly showings, relationships between appreciation and repeat viewing can be examined both within specific weeks as well as across the entire five weeks of available data. Likewise, questions of viewers' consistency of scoring patterns for multiple episodes of the same series can also be addressed in greater detail.

Some currently planned research in related areas, but using new data, includes:

1. In New Zealand, Saatchi and Saatchi recently collected data and carried out research in the audience appreciation area. This data sample has been made available to the University of Auckland Marketing Department (where I am now based) and will lead to interesting opportunities for replication oriented analyses and extensions to data for a different country.
2. Also, at the University of Auckland Marketing Department, Dr. Peter Danaher has recently been investigating a concept he is referring to as Audience Attachment. This is an inferred measure of programme appreciation that is based on actual viewer behaviour rather than reported attitudes. For each programme, the idea is to track the actual percentage of a television programme (in total minutes) that each

individual viewer watches (for all individuals who view at least 20% or more of a programme's total running time). These individual percentages can then be summed and averaged for each programme to calculate a programme 'attachment' score. This should be a behavioural indication of how much viewers 'like' a programme in that viewers who are enjoying or who are highly involved in a programme should be less likely to switch away or to turn off the television. With the cooperation of AGB McNair, the company responsible for New Zealand's peplemeter ratings collection system, such attachment scores will shortly be calculated for all programmes broadcast in New Zealand. This will result in a unique database to examine and compare using similar methods to those in this thesis.

9.5 CONCLUSION AND CONTRIBUTION

This research makes a clear incremental contribution to knowledge in that it replicates and builds directly on a well-established body of prior research. Overall, the key contribution of this thesis lies in providing further insights to help in the interpretation and understanding of formal measures of television audience appreciation. Many of the findings and patterns outlined will be of interest and of potential use to media planners. If traditional television ratings measure viewing as being in the room with the set turned on, then measures of audience appreciation can at least to some extent move beyond simple presence to some indication of 'quality of exposure' to programming (and potentially to advertising).

When there were only a few television channels to choose from, small gradations in audience attitudes and behaviour were largely irrelevant for use

in media planning. As television markets continue to fragment, however, the traditional ratings systems in use today simply cannot and do not convey all the information that is becoming necessary in order to have a better understanding of how audiences react to programmes. Hence, there is an increasing interest worldwide in more qualitative measures of audience attitudes and involvement. This does not imply that audience appreciation ratings are the answer. It is hoped, however, that this thesis has contributed to the general understanding of television audience viewing behaviour and to specific knowledge on this one measure of audience involvement and its limitations. Whatever measure of programme involvement is going to gain usage in the future, there is still a need to research and develop further the links that might exist between television audience viewing behaviour, audience attitudes, and actual advertising effectiveness.

In terms of managerial and media planning implications (also discussed in more detail in the appendix), research in this area is of great current interest, and will be of increasing value as advertisers are faced with more television options and with every greater control over the environments in which their advertisements appear.

APPENDIX: PROGRAMME INVOLVEMENT AND CONTEXT EFFECTS

A.1 OVERVIEW

This appendix introduces a media planning perspective to begin considering how, and for what purposes, the type of audience appreciation data examined in this thesis might be used to help aid the effective placement of television advertising. A literature review first surveys the long history of marketing research into programme context effects (i.e. how the impact of a television advertisement might be affected by the programme in which it is inserted). As will become apparent, such research invariably involves some measure of the degree of audience involvement with a programme. The question to then be discussed is the appropriateness or usefulness of audience appreciation ratings as such a measure of audience involvement.

A.2 THE LITERATURE ON PROGRAMME CONTEXT EFFECTS

The literature discussed in this thesis, in the introductory literature review and in each of the five analysis chapters, was restricted to prior research that directly related to the investigation into patterns of audience appreciation ratings. In now considering some wider implications of findings from this research, it is important to return to the literature.

Involvement has long been a topic of great interest in the marketing and psychology literature, yet it has proven to be a difficult construct to define and operationalise. While the literature yields a wide variety of definitions,

there is agreement that it is a complex rather than a simple factor. Years of research and debate have produced little agreement on how to define or measure the concept of programme involvement, much less on how to apply it to commercial decision making.

Involvement as a single term has often been used interchangeably to describe several quite different phenomena (Batra and Ray 1983; Bonfield 1983). In the marketing literature, interpretations have mostly seen involvement as either a characteristic of the product class (e.g. product classes are often classified as high or low involvement due to such factors as the cost and risk associated with a purchase) or as a subject-centred information processing characteristic. The subject-centred view is the one that would typically apply to involvement with television. According to Batra and Ray (1983), “message response involvement exists not as an enduring predisposition, but as an interactive outcome of many situational factors”. Viewer involvement is seen as a *situational state* that is influenced by an individual’s motivation, opportunity and ability to get involved with and to get something out of television programmes and television advertising.

Interest in the association between viewer involvement in television programmes and advertising effectiveness is certainly not new. Since television commercials can be inserted into a wide variety of programme environments, researchers have long considered the question of how the impact of a commercial may be affected by the programme in which it is inserted. Although nearly all studies have found some relationship, two opposing theories have evolved: the “positive effects hypothesis” and the “negative effects hypothesis”. The positive effects hypothesis suggests that high levels of programme appreciation, liking, or involvement will carry over into the commercial break and enhance the effectiveness of any advertising.

The negative effects hypothesis, in contrast, maintains that viewers will consider advertisements that interrupt interesting and involving programmes to be intrusive. Such commercials will therefore be avoided altogether (e.g. by leaving the room), filtered out, or perceived negatively.

(a) Negative Effects:

Schwerin (1958), a pioneer in the advertising research industry, found support for the negative effects hypothesis in an experiment which exposed subjects to one of three programme types (a quiz show, a musical, or a courtroom drama) and then assessed programme involvement at three levels (“relaxed”, “concentrated”, and “tense”). His findings led him to conclude that the same ads tended to be less effective in tense programmes, such as courtroom dramas, than in more relaxing programmes.

Steiner (1966) concluded that advertisements are considered by viewers to be “objectionable interruptions”, especially when they interrupt interesting shows. He did not explore, however, whether the fact that viewers found commercials objectionable would also render such commercials ineffective.

In a similar argument to that of Steiner, Kennedy (1971) hypothesised that greater involvement in programmes would provoke a stronger “viewer drive for closure” that would render commercials as intrusive and ineffective. Using a random sample of residents of London, Ontario, he ran a small, tightly-controlled experiment using two programme treatments which differed in their ability to generate viewer drive for closure (a suspense-thriller and a situation comedy) and three commercials embedded in natural breaks (for coffee, potato chips/crisps, and breakfast cereal). Six tests were used to measure commercial

impact: unaided recall of the brand name; aided recall of product features and benefits; aided recall of the commercial format; liking of the brand (attitude towards the brand); liking of the advertisement; and a simulated pre and post-test buying decision measure. Although Kennedy found that unaided recall of commercials in the situation comedy was higher than for commercials in the suspense thriller, these results were not conclusive. Using his other measures, some advertisements benefited from being placed in strong plot programmes, while others did not. He did conclude, however, that at the least these experimental results show strongly that programme environment can affect commercial performance.

Bryant and Comisky (1978) examined the effect on recall of positioning an advertisement in differentially cognitively involving portions of a television segment. Their reasoning and expectations built on cognitive psychologists' notions of *proactive inhibition* (which refers to the diminished recall of a message due to interference from information presented prior to the item to be remembered), and *retroactive inhibition* (the interfering effect of subsequent learning on the recall of a previously presented message).

This laboratory experiment involved placing an advertisement for Hamm's beer in different positions within an eight minute segment of the action drama series "Banacek". The four commercial placement conditions were: (1) ad placed between two moderately involving portions of the programme which occurred a few minutes prior to both the action-climax and the resolution of the suspense; (2) ad placed immediately prior to the moderately involving resolution; (3) ad placed subsequent to the climax and immediately subsequent to the resolution; and (4) ad placed between two minimally involving portions of the control programme. Subjects were 120 undergraduate students who had not watched "Banacek" on television the

previous week and who liked beer. Half the subjects received a test of advertising content recall shortly after viewing, while the remainder were tested after one month. Bryant and Comisky's findings support the negative effects hypothesis in that advertising message recall was inversely related to the cognitive involvement potential of the programme material presented immediately *before* and *after* the commercial message.

In a widely cited study, Soldow and Principe (1981) found further partial support for the negative effects hypothesis. A sample of 87 people from a medium-sized business organisation in a major metropolitan area was divided into three groups of 29; two treatment groups and a control group. Programme involvement, the treatment condition, was operationalised by asking a small number of respondents to rank order 15 programmes from most to least "suspenseful." A programme judged to be suspenseful ("Baretta") was shown to one treatment group, while another programme judged non-suspenseful ("The Brady Bunch") was shown to the other treatment group. Subjects in the control group were shown only the three commercials (for a food product, a car, and a household agent) that were embedded within the programmes for the other two groups. Four measures of commercial effects were collected immediately after viewing: unaided brand recall; unaided message recall; attitude toward the advertisement; and intent to buy. Soldow and Principe's findings were that, in general, advertising effectiveness was lower in the case of the more involving (suspenseful) programme.

To summarise, numerous studies have found support for the negative effects hypothesis. These studies, however, have almost invariably been based on off-air tests, under forced exposure (viewers have no choice of what programme to watch), in artificial environments (typically laboratory theatre-test settings),

with relatively small, non-projectable samples (often students). The Soldow and Principe study is a case in point (although they did not use students), in that findings were based on a very small sample, under a forced viewing theatre-test setting, and using a very simple unidimensional definition of involvement (“suspensefulness”). Although ‘significant’, their results seem too weak and inconclusive to draw any sort of generalisable conclusions.

(b) Positive Effects

A report by the Home Testing Institute (1968) is generally cited as early evidence in support of the positive effects hypothesis (TAA, 1984a; Lloyd and Clancy, 1991a and 1991b). This report summarised a series of advertising industry studies carried out in the both the UK and US during the 1960’s. Across these studies, the conclusion drawn was that viewers’ preferences (in terms of increased liking, interest or enjoyment) were consistently associated with higher aided and unaided advertising recall scores, and with stronger pre-post attitude shifts towards the brands being advertised.

Clancy and Kweskin (1971) conducted a comprehensive study which aimed to identify reasons for the apparent lack of reliability in the scores resulting from on-air tests of advertising recall. Such tests, best known as “day-after-recall” (DAR) tests of a commercial, are used to evaluate the ability of television commercials to generate awareness and register a brand message. The typical procedure is to place a commercial in a regularly scheduled broadcast, and then to follow this up twelve to twenty-four hours later with telephone interviews of approximately 200 viewers so as to measure their commercial and message recall. Clancy and Kweskin’s review of unpublished data on the reliability of such tests showed that if the same commercial was

tested twice (meaning that it would appear, if not in a different programme, then in a different episode) the observed recall scores were almost as different as for two different commercials.

To explore reasons for this lack of reliability, Clancy and Kweskin developed a database consisting of 25 on-air tests of 30 second commercials for a broad range of product categories, and involving more than 6,000 respondents. Five variables were then investigated: consumer attitudes towards the *programme* carrying the test commercial (rated on the Burke Marketing Research five-point scale ranging from “my favourite programme” to “poor”); whether they viewed the whole programme; whether they were users of the test brand; education; and age. Of these variables, the single best predictor of advertising recall was found to be the attitudes respondents expressed towards the programme carrier. This evidence (the more people liked the programme, the higher their recall scores) provides clear support for the positive effects hypothesis.

Krugman (1983) addressed the question of programme involvement by exploring whether advertisements shown within “interesting” programmes are less effective. His data concerned all 56 corporate television programmes that were sponsored by General Electric in the US from 1971 to 1980. Measures were collected on attitudes towards General Electric, and questions were asked about viewers’ recent exposure to the briefly described programmes. Krugman concludes by suggesting that “when an interesting show is interrupted by an interesting commercial, the momentum of aroused interest does carry over”. He qualifies this conclusion, however, by stressing that the fact that such aroused interest carries over does not necessarily imply that this makes the advertisements more effective. The answer to that question is left for future investigation.

Perhaps the most comprehensive research into the relationship between commercial effectiveness and the attitudes and behaviours of television viewers was carried out in the US by Television Audience Assessment (TAA, 1984b). In earlier research, TAA (1983a) had developed two complementary indices of programme involvement: Programme Appeal and Programme Impact. Programme Appeal was a similar measure to the UK's audience appreciation measure, providing a summary index of a programme's entertainment value based on each viewer's "personal programme rating" on a scale of 0-10. These scores were then multiplied by 10 to convert the index to a 100 point scale. Programme Impact was a measure of "demandingness", representing the average of an individual viewer's responses to a programme on two scales: "This programme touched my feelings" and "I learned something from this programme". There were four possible answers to both questions ("not at all", "only a little", "some", and "very much") and responses were merged to create a single Impact Index based on a 100 point scale.

TAA's 1984 study was designed to examine the predictive value of the Appeal and Impact measures on various measures of commercial effectiveness. The study took place in Kansas City, Missouri, and used a sample of 470 women between the ages of 18 and 49. In many ways, this was a typical experimental study, with commercials inserted in different programmes to be viewed by subjects in a laboratory viewing environment. TAA's work, however, was unique in that two different viewing environments were used and that extensive measures of advertising effectiveness were taken (brand name recall, message recall, overall evaluation of a commercial's content, credibility of the product's claim, and product purchase preferences). Both viewing environments were laboratory settings, but one involved a theatre test, with rows of seats facing a television screen, while the other was a

simulated “natural viewing environment” where subjects were placed in a modelled living-room setting where there were other distractions available. Subjects in both settings had no choice or control over the programme that was shown.

TAA’s conclusions strongly support the positive effects hypothesis. Increased appeal/impact led to greater paying of attention, while likability and credibility were shown to carry over from the programme to the commercial message. Furthermore, increased programme impact led to a greater intention to purchase for both new and established brands. Importantly, these positive effects of programme impact were much more apparent in the natural viewing situation. This suggests that perhaps many inconsistent “negative effects” studies in the past were distorted by the overly artificial settings in which respondents were tested.

To summarise both sections (a) and (b) above, the positive effects hypothesis tends to be supported by on-air tests of advertising in natural viewing environments. However, most of these on-air studies have used recall as the only measure of advertising effectiveness. Television Audience Assessment’s work therefore provides the most methodologically sound research to date, especially in its use of multiple measures of advertising effectiveness, and in its attempt to create a “simulated-natural” viewing environment. Nevertheless, the programme environment effects question has not been settled. Remaining issues of concern with respect to TAA’s study include the fact that there was no free choice of which programmes to watch (subjects were assigned to programmes) and the reality that even a simulated ‘distractions-available’ setting is no substitute for natural in-home viewing.

(c) Positive and Negative Effects - An Inverted-U Relationship?

In one recent article, there is even talk of an “inverted-U relationship” between programme involvement and advertising effectiveness (Tavassoli, Shultz and Fitzsimons 1995). Here, an experiment found that viewer involvement had a positive effect on ad memory and attitude toward the advertisement as programme involvement increases from low to moderate levels. However, further increases in programme involvement led to lowered advertising effectiveness.

(d) Recent Advertising Industry Research

Two recent UK industry reports, by Carat Research UK Ltd. and by Zenith Media Research) have directly explored the relationship between viewers’ appreciation of television programmes and the ‘effectiveness’ of advertising carried within the programmes. Both of these studies reflect the considerable level of current advertising industry interest in this area (Gullen 1993, Zenith Media Research 1993).

Zenith Media’s research on programme appreciation and advertising recall involved a quota controlled sample of 291 adults in the London ITV region. Telephone interviews were conducted on Sunday February 7, 1993 to determine whether respondents had viewed TV between 6 and 11 pm the previous evening. If so, they were then asked to rate their enjoyment of the programmes they had watched and to answer various spontaneous and prompted advertising awareness questions. Based on this data, Zenith examined the relation between programme appreciation indices and the recall

of ads in the programmes' centre breaks. The conclusion was that there was no strong correlation, either positive or negative. "AIs would be of very limited use in deciding where and when to advertise on TV."

One strength of the Zenith study, especially when compared with many of the laboratory based studies previously discussed, is that it was based on viewers' reactions to television programmes that they had chosen to watch in their natural at-home viewing environment. As with the majority of past on-air tests of advertising, however, the Zenith study suffers in that recall was the only measure of advertising effectiveness used. Furthermore, Zenith's conclusions are difficult to either believe in or to refute in that the study was not strictly comparing 'like with like'. No theoretical attempt was made to allow for advertising working in different ways in different situations. No distinction was made between the different types of advertisements being viewed (i.e. emotional or factual; a new campaign execution or an existing campaign), between the varied brands being advertised (new or established; large or small), or between users and non-users of the brands. In concluding, Zenith outline a number of possible influences that should be taken into account, yet no attempt was made to control for these influences. The personal relevance of advertising, for example, was identified as a key factor yet no measures were taken of brand usage or involvement.

Carat Research refers to their March 1993 advertising recall study as "the most comprehensive study of unforced viewing ever carried out in UK". This study took place in the Granada ITV region. A gross sample of 1,000 adults was broken into 5 daypart samples of 200. Respondents were given programme diaries, in which to record their viewing, along with sealed questionnaires on advertising recall. They were asked to open and fill in these questionnaires on a specific day at a specific time. In this way, with

measurements taken 1 to 2 hours after programme/advertising transmission, Carat was able to compile a database on viewing and advertising recall by:

- Time of day and channel
- The level of attention being paid while viewing
- The positioning of advertisements (centre vs. end break)
- Previous weight of advertising
- Programme appreciation
- Reasons for watching
- Recall of advertising (prompted by brand name)

Carat's conclusion from this research was that programme environment does have an effect on advertising, and that programme appreciation or enjoyment is a key measure of such programme involvement. "Advertising is 70% more effective when it is seen by people enjoying the programme. It is 50% more effective when seen by primary viewers (those who especially chose to watch) and 70% more effective amongst those paying attention."

Like Zenith's study, Carat's research benefits from its use of natural at-home viewing. Also, in asking questions about the level of attention being paid, and about the viewer's motivation for viewing, Carat's study provides more measures of viewer involvement than simply depending solely on audience appreciation scores. As with the Zenith study, however, prompted recall was Carat's only measure of advertising effectiveness. Also, the study lacked any theoretical framework for how and where advertising recall might be likely to vary regardless of programme appreciation influences. Importantly again, Carat also identified "the relevance of the ad to the target group" as a key influence, yet no measure or indication of this factor was taken.

A.3 AUDIENCE APPRECIATION AND MEDIA PLANNING

The often contradictory findings in the above literature review suggest two main points. Firstly, there is little agreement on how to measure the concept of programme involvement. Measures explored have included the level of a programme's "relaxation or tenseness", "cognitive involvement" or "suspensefulness", and various measures of appreciation, liking, enjoyment, interest, attention, impact, and reasons for watching. Secondly, while everyone is concerned with advertising effectiveness, there is not even any consensus on how this should be measured (e.g. advertising recall or recognition, brand name recall, purchase intentions, etc.).

A further complication on the relationship between programme involvement and advertising effectiveness is possible interaction effects. Even if one were to conclude that involving programmes (however measured) are 'better' carriers for advertising, placing an advertisement on this basis assumes that all advertising will perform similarly in similar environments. Might certain types of advertising for certain types of programmes perform better in low involving situations? Shouldn't media planners also consider many other inherent qualities of the *particular* programme and the *particular* advertisement? Might a particular advertisement, for whatever reason, seem incongruous and offensive or ineffective in a particular programme environment?

Despite all of this complexity, it appears that there are certainly implications from this thesis research in relation to media planning issues concerned with evaluating the efficiency of different programmes as vehicles for commercial messages. Many of the programme context effect studies summarised earlier, especially the 'positive effects' investigations based on on-air tests of

advertising effectiveness, do suggest and make use of programme appreciation (liking, interest, enjoyment, etc.) as an appropriate measure of programme involvement. Even without definitive answers on the issue of advertising effectiveness, advertising practitioners already hold beliefs and make decisions on which measures of advertising effectiveness they will use and trust. Similarly, based on ‘gut feeling’ or on a considered review of past literature, advertisers will have come to reasoned conclusions on whether or not programme context does matter and consequently on what measures of programme involvement they will use and trust.

What advertisers do need, therefore, is a greater understanding of how to interpret and make use of the programme involvement measures they might consider. This is where this thesis has direct managerial implications. As illustrated by the two recent UK industry studies outlined above, and by recent research by Saatchi and Saatchi in New Zealand (as mentioned in “Directions for Future Research” in Chapter 9), advertisers are already collecting and making use of similar types of appreciation data. Audience appreciation data is not cheap to collect, especially on an ongoing week-by-week basis, as a large panel of respondents must be maintained in order to achieve effective sample sizes. Even on a syndicated industry-wide collection basis, these costs and the lack of extreme variation in this data calls into question the need for collecting it on a continuous basis.

For most media planning purposes, sufficient and appropriate audience appreciation data could probably be collected through occasional surveys rather than through expensive ongoing consumer panel methods. The greatest likelihood for the future, therefore, is that advertising agencies or media buying firms will collect their own data on a more ad-hoc, perhaps seasonal basis. (Companies could do this either individually, as with the above UK and

New Zealand studies, or in cooperation.) Patterns described in this thesis will greatly help in the interpretation of such future studies where the data may be more limited in scope and extent.

The main patterns found in this thesis that should be of direct current interest to media planners include:

1. With the exception of a limited Double Jeopardy pattern, programme appreciation does not measure the same thing as audience size (i.e. it does not measure the same thing as traditional ratings systems in that two programmes with audiences of equal size and composition can differ substantially in their programme appreciation scores). This Double Jeopardy effect nevertheless needs to be recognised when interpreting and comparing scores for programmes with audiences of quite different sizes.
2. While programme appreciation scores do not vary a great deal, they do show systematic patterns by programme type and channel. Appreciation scores for a programme will mostly remain quite stable from episode to episode, but this suggests they reflect the audience's inherent qualities of the overall series. At an individual response level, viewers' appreciation for different programmes does show much more variation. Individuals can be segmented along demographic and viewing behaviour lines according to their appreciation scoring patterns. Again, a knowledge of such patterns is important in understanding and interpreting individual and programme level appreciation scoring patterns.

3. While not powerfully associated with ratings of audience size, appreciation scores are associated with other direct measures of viewing behaviour (e.g. frequency of repeat viewing). This relationship between the appreciation scores individuals give to a regular series and their frequency of viewing, suggests that appreciation can provide an indication of 'loyalty' to the series.
4. Limited findings in this thesis about 'narrowcasting' are important in that they may challenge both some of the 'hype' about the future of television and the 'conventional wisdom' about how audiences behave (e.g. highly targeted programmes 'should' be particularly liked by those targeted and therefore 'highly involved' viewers who are watching). If, even when looking at extremely biased or targeted audience programmes, such commonly expected 'narrowcasting' patterns do not appear, then this is indeed a case where "null results are not dull results".

In the current television advertising environment, institutional factors (such as the fact that advertising is often bought on a run-of-station basis instead of in terms of particular programmes) may limit the immediate applicability of these findings. With increasing flexibility and choice, due to television market fragmentation and a lessening of the dominance by major broadcasters, advertisers who desire it will have increasing control over the particular programme placement of their advertising. Yes there are relatively small deviations in audience appreciation data, but the television advertising industry already exploits seemingly small deviations in demographic audience composition data to provide savings in buying airtime for targeted advertising. Media planners are, and will continue to be, interested in any supplementary

measures that can usefully complement and begin to move beyond traditional audience ratings data.

REFERENCES

- Ang, I. (1991). Desperately Seeking the Audience. London, Routledge.
- Armstrong, J. S. (1991). "Prediction of Consumer Behaviour by Experts and Novices." Journal of Consumer Research 18 (September): 251-256.
- Aske Research Ltd. (1973a). Audience Reaction to Different Episodes and to Similar Programmes: Exploratory Analyses of Audience Appreciation and Behaviour. London, Report Prepared for the Independent Broadcasting Authority.
- Aske Research Ltd. (1973b). Repeat-Viewing and the Audience Appreciation Index. London, Report Prepared for the Independent Broadcasting Authority.
- Aske Research Ltd. (1975). Some Correlations of the Appreciation Index. London, Report Prepared for the Independent Broadcasting Authority.
- Aske Research Ltd. (1978). Systematic Variations in the AI. London, Report Prepared for the Independent Broadcasting Authority.
- Aske Research Ltd. (1980). Attitudes to Episodes and Programmes. London, Report Prepared for the Independent Broadcasting Authority.
- Aske Research Ltd. (1981a). Interesting or Enjoyable?. London, Report Prepared for the Independent Broadcasting Authority.
- Aske Research Ltd. (1981b). Some Insights from Canada. London, Report Prepared for the Independent Broadcasting Authority.
- Barwise, T. P. (1985). Mass Attitudes and Routine Choice Behaviour. Doctoral Dissertation, London Business School, University of London.
- Barwise, T. P. (1995). "Good Empirical Generalisations." Marketing Science 14 (3, Part 2 of 2): 29-35.
- Barwise, T. P. and A. S. C. Ehrenberg (1982). The Liking and Viewing of Regular TV Programs, London Business School, Report prepared for the Markle Foundation, New York.
- Barwise, T. P. and A. S. C. Ehrenberg (1987). "The Liking and Viewing of Regular TV Series." Journal of Consumer Research 14 (1): 63-70.

- Barwise, T. P. and A. S. C. Ehrenberg (1988). Television and Its Audience. London and Beverly Hills, Sage.
- Barwise, T. P., A. S. C. Ehrenberg, and G.J. Goodhardt (1979). "Audience Appreciation and Audience Size." Journal of the Market Research Society 21 (4): 269-289.
- Bass, F. M. (1993). "The Future of Research in Marketing: Marketing Science." Journal of Marketing Research 20 (February): 1-6.
- Batra, R. and M. L. Ray (1983). "Operationalising Involvement as Depth and Quality of Cognitive Response." Advances in Consumer Research 10 : 309-313.
- Beale, C. (1993). Advertisers to Act on Quality. Marketing: 19.
- Beville, H. M., Jr. (1985). Audience Ratings. Hillsdale, New Jersey, Lawrence Erlbaum.
- Bonfield, E. H. (1983). "Involvement, Attitude and Behaviour: On the Nature of the Relationships." Advances in Consumer Research 10 : 314-318.
- Bryant, J. and P. W. Comisky (1978). "The Effect of Positioning a Message within Differentially Cognitively Involving Portions of a Television Segment on Recall of the Message." Human Communication Research 5 (1): 63-75.
- Buzzard, K. (1992). Electronic Media Ratings, Turning Audiences into Dollars and Sense. Stoneham, MA, Butterworth-Heinemann.
- Carrie, D. (1993). Finding the Core of the Youth Market. Marketing: 26.
- Chaturvedi, A and P. E. Green (1995). "SPSS for Windows, CHAID 6.0." Journal of Marketing Research. 32 (2): 245-254.
- Clancy, K. J. and D. M. Kweskin (1971). "TV Commercial Recall Correlates." Journal of Advertising Research 11 (2): 18-20.
- Collins, M. (1992). "The Data Reduction Approach to Survey Analysis." Journal of the Market Research Society 34 (2): 149-162.
- Dawkins, D. and J. Samuels (1994). "'Barbara' and Audience Appreciation." Admap (February): 71-83.
- Ehrenberg, A. S. C. (1982). "How Good is Best?" Journal of the Royal Statistical Society 145 (3): 364-366.
- Ehrenberg, A. S. C. (1986). Advertisers or Viewers Paying? London, Admap.

- Ehrenberg, A. S. C. (1991). A Primer in Data Reduction. London and New York, John Wiley & Sons.
- Ehrenberg, A. S. C. (1992). "The Problem of Numeracy." Admap (February): 37-40.
- Ehrenberg, A. S. C. (1993). "Even the Social Sciences Have Laws." Nature 356 (September 30): 285.
- Ehrenberg, A. S. C. (1994). Theory or Well-Based Results: Which Comes First? Research Traditions in Marketing. G. L. Laurent and B. Pras. Boston, Kluwer.
- Ehrenberg, A. S. C. (1995). "Empirical Generalisations, Theory and Method." Marketing Science 14 (3, Part 2 of 2): 20-28.
- Ehrenberg, A. S. C., G. J. Goodhardt, and T. P. Barwise (1990). "Double Jeopardy Revisited." Journal of Marketing 54 (3): 82-91.
- Ehrenberg, A. S. C. and J. Wakshlag (1987). "Repeat-Viewing with People Meters." Journal of Advertising Research 27 (1): 9-13.
- Goodhardt, G. J., A. S. C. Ehrenberg, and M.A. Collins (1975). The Television Audience: Patterns of Viewing. (Updated edition 1987) Aldershot, England and Brookfield, VT, Gower.
- Gullen, P. (1993). Measuring the Quality of Television Viewing and its Link with Advertising Effectiveness. London, Carat Research UK Ltd.
- Gunter, B. and M. Wober (1992). The Reactive Viewer. London, John Libbey & Company.
- Hoffman, D. L. (1984). A Multivariate Analysis of Audience Attitudes and Behaviours During Television Viewing. Doctoral Dissertation, University of North Carolina and Chapel Hill.
- Home Testing Institute Inc. (1963). Must Reading About HTI and TVQ.
- Hubbard, R. and J. S. Armstrong (1994). "Replications and extensions in marketing: Rarely published but quite contrary." International Journal of Research in Marketing 11 : 233-248.
- Johnson, H. (1992). "Audience Appreciation: a New Future." Admap (October): 40-44.
- Kasari, H. J. (1985). Patterns of Television Viewing in Finland. Helsinki, Doctoral Dissertation, University of Helsinki.

- Kass, G. V. (1980). "An Exploratory Technique for Investigating Large Quantities of Categorical Data." Applied Statistics 29 : 119-127.
- Kennedy, J. R. (1971). "How Program Environment Affects TV Commercials." Journal of Advertising Research 11 (1): 33-38.
- Kleinbaum, D. G., L. L. Kupper, and K. E. Muller (1988). Applied Regression Analysis and Other Multivariable Methods. Boston, PWS-Kent.
- Kleinman, P. (1994). "Diaries make a comeback." Admap (February): 9.
- Krugman, H. E. (1965). "The Impact of Television Advertising: Learning without Involvement." Public Opinion Quarterly 29 : 349-356.
- Krugman, H. E. (1967). "The Measurement of Advertising Involvement." Public Opinion Quarterly 30 : 583-596.
- Krugman, H. E. (1983). "Television Program Interest and Commercial Interruption." Journal of Advertising Research 23 (1): 21-23.
- Lindsay, R. M. and A. S. C. Ehrenberg (1993). "The Design of Replicated Studies." American Statistician 47 (3): 217-227.
- Lloyd, D. W. and K. J. Clancy (1991a). "Television Program Involvement and Advertising Response: Some Unsettling Implications for Copy Research." Journal of Consumer Marketing 8 (4): 61-74.
- Lloyd, D. W. and K. J. Clancy (1991b). "CPMs Versus CPmIs: Implications for Media Planning." Journal of Advertising Research 31 (4): 34-44.
- Magidson, J. (1993). The CHAID Approach to Segmentation Modelling. Handbook of Marketing Research. R. Bagozzi. Cambridge, Mass., Blackwell.
- Magidson, J. (1995). SPSS CHAID Release 6.0. Chicago, SPSS Inc.
- Marquis, S. (1993). Environment vs. Weight. Marketing: 18.
- McQuail, D. (1983). Mass Communication Theory. London, Sage.
- Menneer, P. (1987a). "Audience Appreciation - A Different Story from Audience Numbers." Journal of the Market Research Society 29 (3): 241-264.
- Menneer, P. (1987b). "Audience Appreciation - A Different Story from Audience Numbers." Admap (June): 36-42.

- Monroe, K. (1992). "On Replications in Consumer Research." Journal of Consumer Research 19 (Editorials in June and September issues).
- O'Brien, T. and P.E. Durfee (1994). "Classification Tree Software". Marketing Research: a Magazine of Management & Applications, 6 (3): 36-39.
- Perry, J. (1995). "Measuring Change. How TV audience measurement might develop in a fragmentary and interactive future." Admap (February): 14-16.
- Rust, R. (1986). Advertising Media Models: A Practical Guide. Lexington, MA, Lexington Books.
- Schwerin, H. A. (1958). "Do Today's Programs Provide the Wrong Commercial Climate?" Television Magazine 15 (8): 44-47, 90-91.
- Soldow, G. F. and V. Principe (1981). "Response to Commercials As a Function Of Program Context." Journal Of Advertising Research 21 (2): 59-65.
- Steiner, G. A. (1966). "The People Look at Commercials: A Study of Audience Behaviour." Journal of Business of the University of Chicago 39 : 19-26.
- Stewart, W. G. (1995). TV Should Appreciate its Audiences. Broadcast: 14.
- TAA (1983a). The Audience Rates Television. Boston, MA, Television Audience Assessment, Inc.
- TAA (1983b). The Multichannel Environment. Boston, MA, Television Audience Assessment, Inc.
- TAA (1984a). Commercial Effectiveness and Viewers' Involvement with Television Programmes: A Literature Review. Boston, MA, Television Audience Assessment, Inc.
- TAA (1984b). Program Impact and Program Appeal: Qualitative Ratings and Commercial Effectiveness. Boston, MA, Television Audience Assessment, Inc.
- Tavassoli, N. T., C. J. Shultz II, and G. J. Fitzsimons (1995). "Program Involvement: Are Moderate Levels Best for Ad Memory and Attitude Toward the AD?" Journal of Advertising Research (September/October): 61-72.
- Thomas, J. (1992). What the Viewer Wants. The *Broadcast Conference*, London.
- Windle, R. and L. Landy (1996). Measuring Audience Reaction in the UK. ARF/ESOMAR Worldwide Electronic and Broadcast Audience Research Symposium, San Francisco, ARF and ESOMAR.

Wober, J. M. (1976). Week 37-75: Some Aspects of Week's Television Among the Audience in the Midlands. IBA Social Report, May. London, The Independent Broadcasting Authority.

Wober, J. M. (1993). "Television Ad Effectiveness and Programme Ratings." Admap (June): 41-44.

Zenith Media Research Ltd. (1993). Programme Appreciation and Advertising Recall. London.

