SOME EFFECTS OF QUESTION WORDING AND QUESTION ADMINISTRATION ON THE PREDICTION OF VOTING BEHAVIOUR

A thesis presented in partial fulfilment
of the requirements
for the degree of
Doctor of Philosophy
in Marketing
at Massey University

Janet Anne Hoek

1996

ABSTRACT

The comparatively short history of polling has seen numerous instances of predictions which failed to match the final election outcome. Of the many possible explanations put forward to explain the divergencies that have occurred, the potential for error introduced by the undecided group has come under increasing scrutiny.

A series of surveys conducted prior to the 1993 New Zealand General Election examined the effects of three question wording or question administration variables: contextual questions; an increase in the level of confidentiality, and a probability scale, on the proportion of undecided respondents and on the accuracy of the estimates obtained. In addition, the research examined the effects of various weighting procedures. The surveys also used a probability scale to estimate respondents' likelihood of voting.

The findings suggest that, although contextual questions, the use of a secret ballot and the use of a probability scale will reduce or eliminate the undecided group, this reduction is not synonymous with an increase in the accuracy of the estimates obtained. Although proportional allocation of undecided respondents resulted in more accurate voting intention estimates than the treatments examined, some weighting factors appeared to improve the accuracy obtained by proportional allocation alone.

While replication research is required, the probability scale provided accurate estimates of turnout.

ACKNOWLEDGEMENTS

I owe a debt to Professor Philip Gendall that is larger than I care to imagine (though I am confident he has reckoned it up). Phil always had the time to listen to me discuss problems I was attempting to address; his helpful comments gave me much to reflect on, and his meticulous research skills taught me a great deal about survey research. Phil's anarchic sense of humour provided light relief in the bleaker moments of this study and taught me that life is not as serious as I sometimes imagine it to be.

My supervisor, Tony Lewis, provided me with thoughtful and constructive feedback and gave me a great deal of support and encouragement. My co-supervisor, Hugh Morton, also responded promptly to my queries and I am grateful for his expertise.

Much as I sometimes begrudged it, Don Esslemont's advice often provided me with a clearer perspective on my work. Often irascible, always provocative, Don gave his advice willingly, and his comments invariably helped me produce better work than I would otherwise have done. Bradburn & Sudman suggested a particularly sound method of pre-testing a questionnaire involved asking a group of disinterested colleagues to comment on it. They warned that this was often a humbling process and it would be remiss of me not to acknowledge the role Don and Phil played in teaching me humility.

Robert Neale read much of this thesis; his perceptive and lovingly ironic comments revitalised both my writing and my sense of humour.

The Massey University Research Fund (MURF) awarded me several generous grants which helped to fund the cost of this study; in addition, the Women's Research Award I received from MURF enabled me to devote more time to my research. Richard Long, Editor of *The Dominion*, and John Harvey, Editor of the *Manawatu Evening Standard*, also helped to fund this research by subscribing to the survey results.

As well as academic acknowledgements, I need also to thank my family for their unfailing support throughout my research. My mother has always encouraged me to strive to achieve the goals I believe are worthwhile and, as she has on so many other occasions, she offered me help whenever I was beset by the crises that seemed to occur at regular intervals during this project.

Finally, I want to plagiarise one of the finest crafters of writing I know, Robert Neale, so I can thank the most important people, Anna and Phil, who made this, as they make everything, worthwhile.

TABLE OF CONTENTS

ABSTRACT i

ACKNOWLEDGEMENTS ii

TABLE OF CONTENTS iv

LIST OF APPENDICES ix

LIST OF TABLES x

LIST OF FIGURES xi

1. INTRODUCTION 1

1.1 Background 1

1.2 Objectives of the thesis 2

1.3 Outline of Chapters 3

1.4 Summary 4

2. POLLING PROBLEMS AND DEVELOPMENTS 6

2.1 Introduction 6

2.2 The Status of Political Opinion Polls 7

2.3 Sampling 9

2.4 Identification of Likely Voters 11

2.4.1 Turnout projections 12

2.4.2 Explanation and Modelling of Turnout 18

2.4.3 Behavioural Models of Turnout 20

2.4.4 Psychological and Economic Models of Turnout 22

2.5 Undecided Respondents 26

2.5.1 Undecided Versus No Opinion 26

2.5.2 Treatment of Undecided Respondents 29

2.5.3 Methods of Reducing the Proportion of Undecided Respondents 29

2.5.4 Allocation of Undecided Respondents 36

2.6 Summary 41

3.	**QUESTION WORDING EFFECTS AND RESPONDENT BIASES**		44
	3.1	Introduction	44
	3.2	Overview of Questionnaire Design Effects	44
	3.3	Question Wording	45
	3.4	Types of Question	47
		3.4.1 Open versus Closed Questions	47
	3.5	Opinion and Behavioural Questions	51
	3.6	Question Context Effects	53
		3.6.1 Question Order Effects	54
	3.7	Errors Introduced by Respondents	58
		3.7.1 Respondent Understanding	58
		3.7.2 Respondent Knowledge	59
		3.7.3 Topic Saliency	62
		3.7.4 Threatening Questions and Social Desirability Bias	64
	3.8	Research Rationale	70
	3.9	Summary	70
4.	**SUBJECTIVE PROBABILITIES**		72
	4.1	Introduction	72
	4.2	The Intention-Behaviour Controversy	72
	4.3	Measurement of Intention	73
	4.4	Development of Subjective Probabilities	78
		4.4.1 Characteristics of Buyers and Non-Buyers	79
		4.4.2 Explanatory power of Probabilities	80
	4.5	Refinement of Subjective Probabilities	87
	4.6	Validation and Extension of the Juster Scale	92
	4.7	Problems Associated with the Estimation of Subjective Probabilities	96
	4.8	A Rationale for the Use of the Juster Scale	99
	4.9	Summary	100
5.	**SUMMARY OF THE LITERATURE**		101
	5.1	Introduction	101
	5.2	Summary of the Literature	102
	5.3	Conclusions from the Literature	105

6.	**RESEARCH METHODOLOGY**		107
6.1	Introduction		107
	6.1.1	Overview of the Primary Research	108
6.2	Sample		108
	6.2.1	Sample Selection and Composition	108
	6.2.2	Quotas	111
	6.2.3	Weighting	111
	6.2.4	Error Calculation	112
6.3	Survey Instruments		113
	6.3.1	Rapport Development	114
	6.3.2	Voting Intention	116
	6.3.3	Confidentiality	117
	6.3.4	Relative versus Absolute Judgments	118
	6.3.5	Demographic Questions	120
	6.3.6	Item Order Effects	121
	6.3.7	Piloting of the Survey Instruments	122
6.4	Survey Procedure		124
	6.4.1	Interviewer Recruitment and Training	124
	6.4.2	Interviewer Auditing	125
	6.4.3	Interview Timing	126
	6.4.4	Assessment of Accuracy	126
6.5	Potential Biases		128
	6.5.1	Respondent Recall	128
	6.5.2	Social Desirability Bias	130
	6.5.3	Bandwagon Effects	132
	6.5.4	Effects of Interviewing and Poll Publication on Subsequent Behaviour	133
	6.5.5	Non-Response Error	136
6.6	Summary		137
7	**TURNOUT PREDICTION: RESULTS AND DISCUSSION**		138
7.1	Introduction		138
7.2	Turnout Prediction		139

7.3 Relationship with the Literature 147
7.3.1 Accuracy of the Juster Scale 148
7.3.2 Effects of Social Desirability Bias 151
7.4 Summary 153

8 CONTEXT EFFECTS ON VOTING BEHAVIOUR: RESULTS AND DISCUSSION 155
8.1 Introduction 155
8.2 Context Effects on Voting Intention Estimates 156
8.3 Effects of Weighting Procedures 164
8.4 Relationship with the Literature 166
8.5 Summary 171

9 SECRET BALLOT EFFECTS: RESULTS AND DISCUSSION 172
9.1 Introduction 172
9.2 Effects of Secret Ballots on Voting Intention Estimates 172
9.3 Effects of Weighting Procedures 182
9.4 Relationship with the Literature 184
9.5 Summary 188

10 RELATIVE PROBABILITIES: RESULTS AND DISCUSSION 189
10.1 Introduction 189
10.2 Effects of a Probability Scale on Voting Intention Estimates 190
10.3 Effects of Weighting Procedures 198
10.4 Relationship with the Literature 200
10.5 Summary 205

11 CONCLUSIONS 206
11.1 Conclusions 206
11.2 Limitations of the Research 206
11.2.1 Adequacy of the Test 208
11.2.2 Sources of Error 209
11.2.3 Design of the Research 211
11.3 Future Research 211
11.4 Summary 213

LIST OF APPENDICES

APPENDIX A:	Details of the Electorates
APPENDIX B:	Interviewers' Instructions
APPENDIX C:	Survey Questionnaires
APPENDIX D:	Voting Results from the Manawatu, Miramar and Western Hutt Electorates
APPENDIX E:	Context Effects: Supplementary Tables
APPENDIX F:	Secret Ballot Effects: Supplementary Tables
APPENDIX G:	Probability Measures: Supplementary Tables
APPENDIX H:	Weighting Procedures

LIST OF TABLES

Table 1	Juster Predictions and Actual Turnout	140
Table 2	Predicted and Self-Reported Turnout Behaviour	142
Table 3	Context Effects on Turnout Prediction	144
Table 4	Context Effects on Predicted and Self-Reported Turnout Behaviour	146
Table 5	Context Effects on Voting Intention Estimates	157
Table 6	Context Effects on Mean Absolute Errors - Electorate Level	159
Table 7	Context Effects on Voting Intention: Predicted versus Self-Reported Behaviour	161
Table 8	Context Effects on Mean Absolute Errors: Predicted versus Self-Reported Behaviour	163
Table 9	Effects of Weighting Procedures on Mean Absolute Errors - Contextual Effects	165
Table 10	Effects of a Secret Ballot on Voting Intention Estimates	174
Table 11	Secret Ballot Effects on Mean Absolute Errors - Electorate Level	176
Table 12	Effects of a Secret Ballot on Voting Intention: Predicted versus Self-Reported Behaviour	179
Table 13	Secret Ballot Effects on Mean Absolute Errors: Predicted versus Self-Reported Behaviour	181
Table 14	Effects of Weighting Procedures on Mean Absolute Errors - Secret Ballot Effects	183
Table 15	Effects of a Probability Scale on Voting Intention Estimates	191
Table 16	Effects of a Probability Scale on Mean Absolute Errors - Electorate Level	193
Table 17	Effects of a Probability Scale on Voting Intention: Predicted versus Self-Reported Behaviour	195
Table 18	Effects of a Probability Scale on Mean Absolute Errors: Predicted versus Self-Reported Behaviour	196
Table 19	Effects of Weighting Procedures on Mean Absolute Errors - Probability Scale Effects	199

LIST OF FIGURES

Figure 1	Ferber & Piskie's Plan-o-meter	83
Figure 2	Byrnes' Scale	84
Figure 3	The Juster Scale	85
Figure 4	Survey Design	114
Figure 5	The Juster Scale	119
Figure 6	Proportion of Self-Reported Voters by Juster Scale Values	149
Figure 7	Proportion of Respondents Nominating Each Juster Value by Self-Reported Behaviour	150

CHAPTER ONE
INTRODUCTION

1.1 Background

The prediction of voting behaviour interests many political and social scientists and preoccupies major political parties. However, despite several decades of research, attempts to forecast election outcomes have met with variable success and pollsters, together with users of their results, have been frustrated by differences between their estimates and the subsequent election outcomes (England, 1987). Thus many pollsters now find themselves in an unenviable situation: the problems of making accurate predictions are now more profound, yet the implications of not providing precise estimates are also more serious.

Researchers have identified a number of factors that may bias poll estimates. These include the sampling frames and the procedures used to select individuals from within these, the questions asked and the response mechanisms employed; the interviewers' characteristics; the timing of polls; the identification of likely voters, and the treatment of undecided respondents. Yet despite detailed research attention into these issues, many poll results still produce inaccurate estimates of actual voting outcomes.

Although none of these problems is trivial, the problems posed by undecided respondents appear to be gaining in magnitude. The proportion of undecided respondents in polls taken at recent New Zealand General Elections has increased, a trend noted also by researchers analysing elections held in the United Kingdom and other European Community countries. Undecided respondents cannot be held solely responsible for polling inaccuracies, but several researchers have identified them as a potentially serious source of error.

While many researchers assume undecided respondents will have a vote distribution similar to that of decided respondents, and so allocate them proportionately across parties, this assumption will not be justified if those undecided respondents who turn out to vote differ in their party support preferences. Accordingly, researchers have explored mechanisms aimed at reducing the proportion of undecided respondents, in an attempt to increase polling accuracy. These measures range from variations in the questions asked and their administration, to models which allocate undecideds on the

basis of other known variables.

Yet despite some efforts by researchers, the question of how best to reduce the error introduced by the undecided group remains largely unresolved. The research which this thesis describes was designed to explore whether changes in question design and administration could reduce the proportion of undecided respondents and improve the accuracy of the estimates thus obtained.

1.2 Objectives

The general aim of this research was to examine the effects of different questions and question contexts on estimates of voting behaviour. Since question wording encompasses a broad range of possible issues, this general aim was formulated into more specific objectives, outlined below.

1. To explore whether a simple probability question could estimate turnout accurately and thus improve the estimates obtained from the voting intention question. This objective was addressed by using the results from a Juster Scale question to weight respondents' voting intentions.

2. To examine whether changes in the context preceding voting intention questions would lead to a reduction in the proportion of undecided respondents, and to an increase in the accuracy of the estimates. The development of contextual attitude and behaviour questions which preceded the voting intention question in selected questionnaire versions allowed this to be analysed.

3. To examine whether increasing the confidentiality associated with an interview, and approximating more closely the conditions surrounding voting behaviour, would reduce the proportion of undecided respondents and improve the accuracy of the estimates. In an attempt to simulate real voting conditions, some respondents were asked to complete a secret ballot independently of the interviewers and to deposit their ballot slip in a sealed ballot box.

4. To investigate whether the use of a probability scale (rather than an intention question) would improve the accuracy of the estimates obtained. To explore this objective, the Juster Scale was employed as a response device in some questionnaires.

1.3 Outline of Chapters

Chapter 2 outlines the difficulties that have beset pollsters in recent decades and the factors responsible for these problems. The chapter begins by examining the status of political polls and by reviewing the debate over whether they represent snapshots of an electorate's mood at a given point in time, or whether they constitute predictions of the election outcome. It then moves to discuss problems affecting polling accuracy, exploring specifically issues affecting sampling and turnout projections. The chapter concludes with a detailed analysis of undecided voters, the reasons underlying indecision, the effects this indecision has, and the options available to researchers who must come to terms with these effects. In essence, this chapter aims to document recurrent problems in election polling, to examine the efforts made to address these problems and the success of those efforts, and to identify residual problems requiring further attention.

Chapter 3 examines issues in question wording and context which may affect the proportion of undecided responses given in a survey. It pays specific attention to the types of questions used in surveys, to the topics (factual or opinion) these explore, and to the context in which they appear. It also examines respondent characteristics which may bias survey estimates. These characteristics include respondents' understanding of the question, their knowledge of the topic, and the effect of social desirability bias. This chapter has three aims - to identify different question design effects; to analyse specific respondent biases and the effect these may have on poll estimates; and, by isolating specific question design and respondent bias factors affecting the proportion of undecided responses given, to suggest how the research undertaken may extend the current understanding of these.

The fourth chapter explores the more specific debate over response mechanisms, and discusses dichotomous questions, questions employing intention scales, and those featuring probability scales. The development, advantages and disadvantages of

probability scales are discussed, as is the empirical evidence relating to one specific probability scale, the Juster Scale. The key aim of this chapter is to outline a rationale for the inclusion of the Juster Scale in the research methodology.

Chapter 5 summarises the key themes emerging from the literature review and restates the thesis objectives in the light of these.

Chapter 6 outlines the research methodology and how it addresses the research objectives. This chapter contains a detailed discussion of the questionnaire design, testing and evolution, of the sampling procedures employed, of the interviewer selection and training, and of the sources of error that may have affected the research.

Chapters 7 to 10 outline the research findings. Chapter 7 presents the findings relating to turnout prediction, while the remaining chapters analyse the effect the different questionnaire treatments had on the level of undecided respondents and the accuracy of the estimates.

Finally, Chapter 11 discusses the implications of the research findings and presents some conclusions about the different treatments tested and their effects on the survey estimates. Suggestions for modifications of the survey method and future research are also made.

1.4 Summary

This thesis investigated the effects various question wording or questionnaire design treatments have on the accuracy of voting intention estimates. The treatments examined include the position of the voting intention question in a questionnaire, the administration of this question, and the response mechanisms made available to respondents.

The first two treatments explored the effects contextual questions and the use of a secret ballot had on the accuracy of voting intention estimates.

The third treatment employed a purchase probability scale, the Juster Scale, to test whether a response mechanism which allows respondents to express relative

probabilities improves the accuracy of the resulting estimates.

Before examining the results of these treatments, the thesis explores the literature documenting problems which have affected political opinion polls. One problem, question wording, is then discussed in detail before an overview of the Juster Scale's development and refinement is presented.

The results of a series of surveys conducted during the 1993 New Zealand General Election which incorporated these treatments are then presented and discussed.

CHAPTER 2

POLITICAL OPINION POLLING: PROBLEMS AND DEVELOPMENTS

2.1 Introduction

In principle, political opinion polling appears a straightforward task, requiring only the administration of a couple of questions to a group of voting age people. However, in practice, even wording the questions and drawing a sample may have a marked effect on the final estimates. Perry (1979) enumerated the full range of problems: "*All one has to do is use a properly drawn sample of the electorate large enough to minimise random sampling error, get honest answers from everyone, do the questioning close enough to the time of voting to minimise changes in voting intentions, anticipate how the undecided will vote, and, finally, distinguish between voters and non-voters in the electorate*" (p. 312). Given the potential for error that exists in all of these steps, any discrepancies between predicted and actual election results seem unsurprising.

As the number of polls failing to predict election outcomes accurately has grown, so too has the range of views put forward to explain this phenomenon. Many researchers have noted methodological shortcomings which may introduce error into the survey estimates. These shortcomings fall into three general categories: sampling problems; differing likelihoods of voting, and the treatment of undecided respondents.

This chapter examines these issues and the effects they may have on political opinion poll estimates. After examining the status of political opinion polls the chapter overviews the problems that appear to have affected recent polls. These problems are then explored in greater detail by means of a brief overview of sampling and then a more detailed discussion of likely voters and their identification. The specific problems presented by undecided respondents and the effects these problems may have are then examined and further lines of enquiry, developed in detail in Chapters 3 and 4, are presented.

2.2 The Status of Political Opinion Polls

Pollsters continue to debate whether polls simply represent a "snapshot in time" of public opinion or whether they constitute formal attempts to predict the final election outcome. Recent discrepancies between polls and election results have intensified this debate and appear to have led many to seek sanctuary in the "snapshot" argument. Sieger (1990), for instance, argued that : "*Opinion research, however, is not about predicting future developments, but attempts to measure within a known degree of reliability, a snapshot of views as they are at the time of the interview*" (p. 324). Butler (1994) also claimed that, despite the media's demands, the business of pollsters is to measure the present, since the future is always essentially unpredictable.

Many other researchers have also promoted the notion that polls capture only a frame in the whole electoral montage (see Mavros, 1987; van Westendorp, 1987; Crespi, 1988; Webb, 1988, and Worcester 1992a). Although claiming that polls are accurate at the time they are taken, these researchers suggested that it is unreasonable to expect polling accuracy to extend into the future, especially when quite unforeseeable events could affect the party support distributions obtained. While some researchers have conceded that polls may document trends occurring within an electorate, the direction and strength of the trends detected have not always been reflected in the actual outcome, leading most pollsters to rely on the "snapshot" explanation.

Intuitively, this argument offers some appeal. However, it suffers from logical limitations. Irrespective of how these pollsters know their polls are accurate when taken, other facets of survey and market research address the problem of predicting future conditional behaviour and the consequences of inaccurate projections. Thus managers introducing new products or services need to anticipate demand for these in order to assess whether further development, or even launch, is warranted. Precedents from other related disciplines therefore suggest that prediction, while difficult, is nevertheless not an impossible aim.

Furthermore, closer examination of the pattern of support for the snapshot argument could lead a cynic to conclude that the number of advocates waxes and wanes in accordance with polls' accuracy. In other words, when a close relationship between polls and an election result exists, the polls were predictive. Conversely, when the

polls and final outcomes have less in common, support for the snapshot argument grows correspondingly stronger.

This trend has not gone unnoticed by polling commentators; Roper (1984) and more latterly Bartels & Broh (1989) have pointed out that pollsters who downplay the status of their results effectively fail to take responsibility for their work: "[Their] *becoming modesty notwithstanding, public pollsters' salaries are paid directly by media executives ... whose interest in election polls rests squarely on the belief that polls predict what will happen on election day*" (p. 572). Even Crespi (1988), a supporter of the snapshot argument, wrote: "*...though a pre-election poll is in itself unquestionably a measurement and not a prediction, concluding that even if a poll were conducted immediately before an election, one cannot hope to measure voter preferences accurately enough to approximate election results closely is to impugn the meaningfulness of all polls*" (pp. 4-5). Crespi's analysis of factors contributing to poll accuracy also revealed that polls which sought to predict election outcomes were significantly more accurate that those which attempted to analyse an electorate's behaviour. In concluding, Crespi noted that aiming to "analyse" an electorate *"can reduce the pressure to produce more than an approximate assessment of voter preference.... setting analysis as one's goal may be self-defeating*" (p. 173).

Despite its supporters, the snapshot defence has some unsatisfactory characteristics which undermine its logic. This is not to suggest that predicting election outcomes is a straightforward or simple task; as noted, the process may be fraught with difficulty. However, accepting the snapshot argument obviates the need to explore wider issues affecting the accuracy of polls, since there would exist no objective measure against which the polls could be assessed (see also Sigelman & Presser, 1988). Election polls which do not aim to predict a final election outcome are purely descriptive, and offer behavioural insights only when incorporated into a longer term trend analysis. The very real difficulties of making accurate predictions should act as an incentive for researchers to explore and test how methodological refinements might strengthen the alignment between polls and final results, and not as an excuse for failing to do so. The following sections examine the specific problems researchers have identified, the conclusions they have drawn, and the recommendations they have advanced.

2.3 Sampling

The accuracy of any survey may be undermined if the sample of people interviewed does not represent the population whose views or behaviours are explored or projected. Skewed samples lead directly to biased results. However, though few would dispute this statement, it is more difficult to ensure that a sample is fully representative, since not all respondents will be available, or willing, to participate in the survey.

Crespi (1988) debated the arguments for and against probability sampling and although he reported no differences in accuracy between polls based on probability or non-probability samples, he suggested potential differences were masked by variations in the polls' methodologies. According to Crespi, pollsters saw other issues, such as respondent selection, use of quotas, clustering and weighting, as more important.

Traugott (1987) raised many questions about sample selection. He argued that variations in the procedures used to select respondents from households, and in the persistence of callbacks, could affect not only a sample's demographic composition, but also the attitudinal and behavioural estimates produced (see also Mitofsky, 1981). Using polls conducted prior to the 1984 US presidential election, Traugott identified six different sampling techniques, ranging from whoever answered the phone (or door, in the case of face-to-face interviews) through to three callbacks to speak to a selected household member. To explore this further, Traugott compared the characteristics of samples obtained after one call with those secured after several callbacks. He concluded that the samples' characteristics and political preferences changed considerably, but suggested that after four callbacks the changes were not great enough to warrant further contact attempts (see also Bogart, 1985; Opatow, 1991; Gendall & Davis, 1993).

Researchers generally have recognised the potential biases caused by non-contact and have sought to reduce these by establishing quotas or by stratifying. They therefore ensure that their samples contain set numbers or proportions of people belonging to pre-specified groups. However, recent analyses of poll results have suggested that quota sampling may introduce more biases than it removes. Although it should guarantee that samples contain the "correct" proportion of people in key demographic

groups, some have argued that quota sampling results in the undersampling of certain parties (see Jowell, Hedges, Lynn, Farrant & Heath, 1993, who argued that quota sampling in the 1992 UK General Election under-sampled Conservative supporters by around 6%). Jowell *et al* noted that quota controls could only improve the accuracy of the estimates obtained if they had a direct relationship with the variable of interest. Most pollsters use age and gender quotas, since these are easier to implement and monitor than other demographic variables, yet Jowell *et al* argued that neither age nor gender were as highly correlated with voting behaviour as socio-economic variables (see also Murphy, 1978). Nor, they suggested, could researchers simply assume that people within each quota band are interchangeable. Crespi (1988), discussing Mitofsky's views, also noted that demographic quotas will not protect against sample error within demographic groups (p. 36). The implication of these arguments is clear: quota samples may do little or nothing to improve polling estimates, and may even generate errors not present in equivalent size random samples.

However, the costs of repeated callbacks in random samples, the timing difficulties, and the apparent similarity of some quota and random sample estimates have led other pollsters to argue that random sampling offers no advantage over quota samples (see Worcester 1984, and 1987; Butler, 1994). Thus, in general, researchers have tended to agree that non-sampling errors outweigh the effects of sampling errors and account for more of any difference between the polls and the election outcome (see Felson & Sudman, 1975; Day & Becker, 1984; Converse & Traugott, 1986; Buchanan, 1987, and Crewe, 1992). Nevertheless, the lack of empirical evidence to support random sampling in favour of quota sampling or vice versa should not be taken as meaning that the argument is either resolved or unimportant. On the contrary, Jowell *et al*'s comments should prompt further consideration of the quota variables currently used to determine polling samples and the weights applied retrospectively to random samples to bring them into line with known population parameters (see also Crespi, 1988).

Some researchers have expressed concerns about the widespread practice of clustering, which decreases the precision of the estimates obtained. Despite the effects of clustering, the media typically report the margins of error associated with simple random samples, thus leading some researchers to argue that the media should be

more cautious and conservative in their reporting of poll results (Roper, 1983; Pearl & Fairley, 1985; Converse & Traugott, 1986; Collins, 1988; Jowell *et al*, 1993; Butler, 1994; Hoek & Gendall, 1996). The debate over clustering has concentrated on how margins of error are calculated and reported, rather than on whether clustering might bias the estimates, and so is less directly relevant to this discussion. However, these former issues, relating to the precision of the estimates, are explored in more detail in Chapters 6 and 7.

Since sampling theory has been thoroughly debated elsewhere, and as the issues arising from this debate are not central concerns of this thesis, this section serves only to highlight specific sampling decisions which affect poll estimates. However, many related issues require more detailed analysis. For example, decisions about whom to interview, the ratio of successful interviews to non-contacts and refusals, and even the timing of polls will all affect the sample composition to varying degrees (and are discussed in Chapter 6). The following section examines the problems inherent in identifying likely voters, and, in predicting how many of those expressing each intention will turn out at a given election.

2.4 Identification of Likely Voters

Not all people who reside in an electorate may have formally registered to vote and, even among those who have, not all will actually cast a vote. Pre-election pollsters must thus ensure their estimates are based on the views of likely voters, since those people unlikely to vote, or certain not to vote, will clearly have at best only a negligible effect on the final election outcome. However, if these people constitute a sizeable proportion of a poll, and the distribution of their stated preferences differs from that of committed voters, including them in calculations of party support estimates might bias those estimates.

It is clear from the research reporting various polls that problems in differentiating clearly between likely voters and non-voters have biased some polls and have even resulted in pollsters concluding that the wrong party will win a forthcoming election see Perry, 1979, for example). Crespi (1988) acknowledged the problems poor turnout prediction could cause: *"Pollsters concur that the most accurate measurement of preference among all voting age adults can lead to an inaccurate assessment of*

candidate standing if they cannot predict who will vote .... pollsters also concur that it is one of the most difficult measurement tasks in polling and that it has not been solved by commonly used techniques" (p. 74).

Researchers have tended to approach turnout prediction in one of two ways. First, some have followed a cognitive line of enquiry, exploring the factors affecting turnout, their relationship with demographic variables, and how knowledge of the processes or issues mediating the decision to vote can be used to predict turnout. Other researchers have sought to develop more behaviourally-oriented models which include variables such as respondents' past behaviour and their behavioural intentions. The following discussion explores the problems caused by inaccurate projections of turnout and the approaches researchers have taken when addressing these.

2.4.1 Turnout Projections

Early polls did not specifically differentiate between likely voters and non-voters, a fact held largely accountable for any inaccuracies that resulted (see Perry, 1979; Buchanan, 1986, or Crespi, 1988, for a full discussion of this). Perry described the history of polling from the early straw polls through to modern methods. He argued that the sampling problems affecting early polls had been addressed and that the accuracy of the estimates had increased[1]. Other non-sampling errors, such as turnout and the problem of allocating undecided voters, were, Perry suggested, more likely to contribute to errors in poll predictions.

Perry's views were supported by subsequent election commentators who also sought to establish why poll results and election outcomes sometimes deviated so dramatically (see Abrams, 1970; Levy, 1983; Day & Becker, 1984; Roper, 1984; Marsh, 1985; Borrelli, Lockerbie & Niemi, 1987; England, 1987; Mavros, 1987; van Westendorp, 1987; Worcester, 1987; Collins, 1988; Crespi, 1988; Sieger, 1990; Jowell *et al*, 1993; Sparrow, 1993; Lau, 1994). While not all these researchers attributed polling inaccuracies primarily to a failure to differentiate accurately between likely and unlikely voters, all cited this as a contributing factor.

---

1 Perry noted changes in the sampling frame and the movement away from mail survey polls to interviews conducted as close as possible to the election date.

Abrams (1970), commenting on the failure of UK polls to predict the 1970 General Election outcome, noted a number of possible explanations, though he specifically excluded the more direct effects of sampling decisions outlined in Section 2.3. Instead, he noted a number of changes in the electoral environment[2] and suggested the combined effect of these was probably sufficient to bias poll estimates in the direction observed. Although not overtly dismissing methodological problems as potential sources of error, Abrams nevertheless firmly located the error in the external environment. However, while environmental factors can affect poll outcomes, researchers should, in principle, be able to identify and account for these factors over time. Crespi (1988) also criticised the "explanation" of large polling errors "*by reference to the unique qualities of a particular election rather than by reference to inadequacies in research designs*" (p. 171-172).

Marsh (1985) also examined problems of turnout prediction, in particular whether this was attributable to sampling deficiencies or to inadequacies in the questions used. While he found that turnout among respondents exceeded the general turnout (thus raising questions about the effects of pre-election interviews on respondents' behaviour), he noted that 9% of those claiming to be "very likely" to vote in fact did not cast a vote. He also noted that the undecided group was not less likely to vote than those respondents who articulated a clear preference prior to the election (an observation supported by Hoek & Gendall, 1993, but queried by Crespi, 1988 who suggested that many undecided respondents were also non-voters). After examining several different response groups, Marsh concluded that turnout was consistently mis-estimated and that differential turnout thus did not explain the errors in prediction that occurred. He argued instead that changes in opinion, perhaps in response to media coverage of the polls, may have accounted for the divergences he noted.

However, these conclusions were not shared by Collins (1988) who examined polls surrounding the 1987 United Kingdom General Election, again seeking to explain how and why they had diverged so greatly from the actual election outcome. While opting more firmly for the differential turnout explanation, Collins also explored other possible influences, though he discounted many of these. For example, he suggested

---

2 These factors included lowering of the voting age, the summer timing of the election and higher immigrant polling.

that late swings (endorsed by Worcester, 1986, 1992b, and proffered by Marsh, 1985, as a likely explanation of polling inaccuracies), were unlikely to account for the problem, since polls held immediately prior to the election contained the same errors as those taken earlier.

Instead, Collins suggested that the lower than desired accuracy of the polls was due to differential turnout, though he did not rule out the possibility of more fundamental changes in respondents' intentions. Collins did not set out specific procedures for dealing with this, though he commented in detail on the calculation of error margins, noting how flaws in these calculations had led to misguided assumptions about polls' outcomes. In conclusion, Collins suggested that researchers and the media should treat poll data more cautiously.

Despite Collins' recommendations, polling estimates did not improve markedly and the aftermath of the most recent United Kingdom General Election saw Sparrow (1993), Jowell *et al* (1993) and a MRS Committee of Enquiry variously examining why the polls failed to predict the 1992 United Kingdom General Election outcome. Sparrow identified three possible explanations for the poor performance of polls: turnout; people who changed their mind about who to vote for between the interview and the election, and late deciders, those who finally decided who to vote for at the end of the campaign. While arguing that a late swing accounted for most of the error, Sparrow also acknowledged the difficulty of predicting turnout and then of using this to weight the party support estimates.

Sparrow used a rather crude 4 point ordinal scale to capture voting intention. Although he did not examine the predictive validity of this scale, its construction does not compare favourably with the more rigorously tested methodologies put forward by Perry (1979) or Traugott & Tucker (1984). Sparrow's failure to consider the accuracy of the scale vitiates the remainder of his discussion, which explored differential refusal rates and logical inconsistencies in respondents' answers. Unless the scales used have construct validity, it is not clear whether the differences between predicted and actual results arise because of flaws in respondents' reasoning or deficiencies in the response devices presented to them. While Sparrow's conclusion, that differential refusals seriously biased the estimates, is intuitively logical, the range

of remedies he suggested fails to address the more fundamental problems affecting polling.

Jowell *et al* (1993) addressed the turnout prediction problem more directly, and explicitly discussed differential turnout and the various factors that could contribute to this: differential registration; differential turnout across party supporters; and the effects of overseas voters. However, they suggested there was conflicting evidence relating to differential turnout, and argued that neither overseas voters (who were simply too few in number to have any marked effect) nor differential registration were likely to have been important factors affecting the outcome. Although, like Sparrow, they concluded that problems caused by respondents who refused to divulge their views could affect poll estimates, their chief concern lay with undecided respondents, who, they argued, were primarily responsible for the errors observed.

Across the Atlantic, American pollsters were grappling with similar problems. Acknowledging criticisms over the polls' failure to predict outcomes in the 1980 elections, Kohut (1986) argued that the credibility of American polls was effectively on the line in subsequent elections, and that pollsters needed to recognise the potentially serious consequences of further inaccurate predictions. His warnings had already been noted, and researchers had, for some time, been advancing and considering the explanations developed by their British counterparts.

Like Collins (1988), Roper (1984) also noted concerns about the abuse of sampling error to make estimates appear more precise and scientific than may be the case. However, when attempting to explain the polls' failure to predict election outcomes, Roper lent more support to the "closet voter"[3] theory, although he explicitly acknowledged the problem of defining likely voters. His advice, to use respondents' past behaviour as a predictor of their future behaviour, is unfortunately limited to situations which remain constant across elections. Clearly, if a large proportion of the population are first time voters, or if previously unfamiliar candidates or parties are contesting the election, this advice becomes more difficult to implement. In addition,

---

3 This theory asserts that respondents, unwilling to give a clear mandate to any politician, refrain from making either their intention to vote or their actual preferences known to interviewers. The net effect of this on candidates in the election is unclear, but presumably this action sends a message that voters are not to be taken for granted.

although many researchers have accepted past behaviour as an important predictor of future behaviour, they have differed over how best to include this in turnout models (see Crespi, 1988 for a full discussion of this issue).

Both Marquette & Hinckley (1988) and Granberg & Holmberg (1986) suggested that respondents tend to act consistently with their *recalled* past behaviour. However, Granberg & Holmberg argued that this was not necessarily the same as respondents' *actual* past behaviour. If these assertions have empirical support, they suggest respondents' self-reported behaviour may have better predictive validity than official records of their actual behaviour. Yet while shedding some light on the question of how to assess past behaviour, this tentative finding does not resolve the wider problem of what variables could be substituted if information relating to respondents' past behaviour is not available.

Day & Becker (1984) addressed the problem of ascertaining who will vote directly and dwelt specifically on the identification of likely voters, though they also acknowledged the role of flawed sampling procedures. After considering, among other possibilities, differential turnout as an explanation of polling inaccuracies, they opted for a straight-ticket voting explanation, on the grounds that there appeared to be more evidence to support this theory than there was to support any of the other explanations they had investigated. Thus despite the preferences respondents might express in pre-election polls, their actual voting behaviour may differ, depending on the other voting decisions they make. This explanation focuses attention on question-wording and the effectiveness of questions currently used to estimate voting intentions, (explored in more detail in Chapter 3) but sheds less light on the problem of predicting turnout.

Problems associated with question wording were also explored by Borrelli, Lockerbie & Niemi (1987), who sought to understand why the Democrat-Republican partisanship gap varied from poll to poll. As well as considering question-wording, their analysis of 51 polls taken in 1980 and 1984 also concluded that the sample composition (likely voters versus the general public) and the poll timing affected the estimates. Crespi (1988) also discussed polls taken at the same time which produced varying estimates of party support. He attributed these differences to variations in the sampling

procedures and, more particularly, to the different procedures used to estimate likely voters.

Similarly, Lau (1994) concentrated on the effects sample composition may have, suggesting that the mixed predictions relating to the 1992 US Presidential election *"may have all been the result of methodological artifacts due to shifts in sampling frames ("registered voters" to "likely voters") or how "undecided"` voters are allocated among the candidates"* (p. 3). While Lau explicitly recognised the effects poor identification of likely voters could have on poll estimates, he concluded that timing factors bore the strongest relationship to poll error.

Few researchers have argued that differentiating likely from unlikely voters is unimportant. However, not all researchers have seen turnout prediction, or the identification of likely voters as pivotal factors in the accuracy of pre-election poll predictions.

One explanation of these divergent views may lie in the diversity of polls considered and the possibility that different electorate and election conditions mean the factors affecting poll estimates vary from one situation to another. However, without more detailed knowledge of these conditions, this explanation remains difficult to substantiate (though it is consistent with Converse's, 1966, normal vote concept). Notwithstanding the difficulty in defining more precisely the error associated with differential turnout, Buchanan's (1986) overview of 155 forecasts across 68 elections since 1948 concluded that difficulties in identifying likely voters had resulted in errors affecting many of the estimates he examined. The twin problems of quantifying both the level of error associated with poor turnout predictions and then its effect on party support estimates, means the potential for error to occur must remain high until more accurate prediction instruments are developed. The following section examines attempts to model and predict turnout behaviour.

2.4.2 Explanation and Modelling of Turnout

Before discussing different models and the variables which these use, it is important to examine the schools of thought regarding voting behaviour, since these schools have attached varying levels of importance to turnout prediction and have approached election prediction generally from quite different perspectives. These views typically arise from the Columbia or the Michigan schools, developed during the 1950s and 1960s. The Columbia school places a heavy emphasis on the social and political context within which voting occurs and pays particular attention to the family history of voting and to generational issues (see Campbell *et al*, 1960; Converse, 1964). Social class, according to this school, is a key predictor variable. However, although this view dominated thinking while social class was a constant phenomenon, during the 1960s and 1970s the notion of a stable electorate, fashioned chiefly by social class, began to break down. In response to this, the Michigan-based ideas of party identification gained in popularity, as models based on this concept explicitly recognised that ideological sympathies could transcend class barriers (see Downs, 1957).

Chandler (1988) noted that the Michigan school, with its heavy emphasis on partisanship, maintained strong predictive validity at least until the 1970s, when growth in the proportion of independents rendered partisanship generally less stable. Refinements to the Michigan model thus involved an acceptance that partisanship may develop though early socialisation and through more recent experiences, including evaluation of economic factors and other salient issues.

Burnham's ideas attempted to combine elements of both approaches (see Chandler, 1988). While suggesting that political attitudes did not play a primary role in electoral participation, Burnham instead saw elections as part of a historical process which was subject to periodic realignment (see Mitau, Thorson & Johnson, 1969 and Shaffer, 1974, for tests of this approach). Chandler (1988) investigated the relationship between these views of turnout and election-behaviour and concluded that while both could provide more or less accurate predictions, neither could be formulated in the precise mathematical terms that would make possible a more rigorous assessment of their respective abilities.

However, Converse's (1966) normal voting model did attempt to specify a procedure for estimating both turnout and election outcomes. Drawing on the Michigan school, Converse tried to separate out the effects of longer term and more short term issues. He proposed that elections featured not only the short-term specific characteristics which affect individual elections, but also longer term traits which affect voting behaviour in a number of elections. He defined a "normal" election as one which had no intervening short-term factors, and developed a model which specified the relationship between the long-term variables and voting behaviour. Party identification played a key role as a long-term variable which Converse argued acted as an umbrella variable representing a number of long-term factors.

To produce estimates, Converse first had to estimate "normal" turnout and "normal" voting behaviour. He achieved the latter by discounting the effect of the short-term factors and ascertaining each partisan group's turnout by regressing this against the overall election turnout. This procedure established turnout within each group, although it did not indicate overall "normal" turnout, which Converse hypothesised would be slightly below average.

Establishing partisanship involved a similar procedure to assess the level of defections associated with particular parties (regressing the proportions of Democrat defectors against the proportion of Republican defectors and identifying points with equal co-ordinates). Once both turnout and partisanship (based on party identification) were calculated, the normal vote expected to be associated with a party could also be established.

Although the schools of thought summarised briefly above have attempted to explain the distribution of party or candidate preferences, and any shifts in these, they have also influenced the choice of variables used in turnout prediction models. Debate continues about whether turnout is a function of social class, party identification or some combination of these and other variables, (see Curtice & Jowell, 1992). The remainder of this section examines various turnout models and the theoretical background from which they have developed.

While retaining links with either the Michigan or the Columbia schools, or some amalgam of these, turnout prediction has tended to evolve in a divergent manner.

Down one path have headed researchers following a more behaviourally oriented approach. Although their models may employ both demographic and ideological variables, they are dominated by measures of respondents' historical and projected behaviour. An alternative pathway is taken by researchers more concerned with the role played by social, psychological and economic variables. Often presented as an eclectic combination of factors arising from a variety of disciplines, this approach typically examines turnout within a broader context. The following discussion explores the more behaviourally oriented approach before turning to consider other perspectives.

2.4.3 Behavioural Models of Turnout

Initial attempts to predict behaviour were rather rudimentary, and essentially involved a direct question which elicited either a "yes" or "no" response or (in some instances) an undecided response. Problems associated with these dichotomous questions are outlined and considered in detail in Chapter 4; it suffices to note here that attempts to predict voting behaviour using "yes/no" questions have not met with great success (Perry, 1979). Rather it became clear that this decision was more complex and involved several different factors. Perry (1979) attempted to accommodate these factors within a turnout scale he developed. Thus he not only questioned respondents about their intention to vote and the certainty surrounding this intention, but he also explored their voting registration status, previous voting behaviour, political allegiances, the intensity of any preferences they expressed, and various attitudinal variables. His model clearly incorporates elements of the Michigan school, although previous voting behaviour and political allegiances are arguably proxies for social class, the mainstay of the Colombia model.

Perry's scale was very comprehensive and employed variables which had clear relationships with turnout, factors held responsible for the improved predictions it made possible. However, while this scale enabled him to rank individuals according to their likelihood of voting, Perry still had to predict turnout as a separate task, and for this he relied on more judgmental criteria. Validation of the scale by comparing its predictions to records of actual voting behaviour showed it provided very accurate predictions in high turnout elections. Yet where turnout dropped to around 40%, or below, the scale's accuracy deteriorated.

Traugott & Tucker's (1984) work followed on directly from Perry's findings and sought to identify variables which had a strong relationship with respondents' likelihood of voting. After some experimentation, they identified three variables: voter registration, past voting, and interest in the election, which were used to estimate respondents' likelihood of voting.

Traugott & Tucker approached the estimation task in two ways. First they constructed an index which allowed them to identify likely voters, and to estimate the effects of different levels of turnout. The second approach used logit regression to develop maximum likelihood probabilities of voting for each respondent. These probabilities then served as weights, producing a probable electorate composed of differential contributions from all respondents (p. 334). Only the second approach actually produced estimates of turnout; the first allowed different assumptions about turnout to be made, but did not provide a single turnout predictor. While the weighting procedure appeared to provide accurate estimates, the co-efficients used could not be assumed to apply across elections, thus each new data set required their re-calculation. While this model offers a solution to the problems of predicting turnout, it is complex and should not preclude the search for simpler procedures, of equal accuracy.

Petrocik (1991) extended Traugott & Tucker's model to include variables associated with non-voting, which he used to discount respondents' probability of voting. He then compared the extended and original Traugott & Tucker models and concluded that the former produced the most accurate estimates.

Buchanan (1986) was also firmly of the view that failure to differentiate between likely and unlikely voters had affected the predictive accuracy of many of the more than 150 polls taken since 1948 that he examined. To counter this bias, he adopted many of the criteria Perry (1979) described (past voting, registration, and interest in the election) either to down-weight or eliminate those respondents unlikely to vote or certain not to vote. However, he did not elucidate a model outlining either the role or the performance of these variables.

Nor did Jowell *et al* (1993), while echoing the call to do more to establish respondents' likelihood of voting, set out more specific means of achieving this.

Although they suggested four questions, encapsulating the variables Perry (1979) had included in his model, they did not report on the predictive validity of these questions. Nevertheless, their argument in favour of developing and testing "*the psephological equivalent of laws of behaviour*" (p. 256) established common ground with previous research and provided a focus for future work.

As well as the models examined here, researchers have also used a variety of screening questions exploring, for example, respondents' registration status as a proxy for turnout. However, respondents' proclivity for over-reporting their registration to vote meant the estimates were invariably inflated. Even treating those respondents unsure about their registration status as unregistered failed to compensate for the over-reporting bias.

Other researchers have used simple measures of past turnout behaviour to predict likely voters (see Crespi, 1988). However this approach has been criticised for its failure to accommodate new voters and for its assumption that respondents will behave consistently.

This overview briefly summarises some of the major work on turnout models. Perry's (1979) research represents a seminal study on which most subsequent research projects are based, even if this is not explicitly acknowledged. Perry's work, in turn, drew on aspects of both the Columbia and Michigan schools and incorporated variables relating to social class, as well as those with a more ideological orientation. However, his work has not been the only attempt to combine elements of these formerly disparate approaches. A parallel research effort has encompassed psychological and economic theories, using these to explain both turnout and party support preferences. The following section summarises findings relating to this approach.

2.4.4 Psychological and Economic Models of Turnout

Several research efforts have compared psychological and economics based models. For example, Chapman & Palda's (1983) research specifically compared the predictive accuracy of models based on social-psychological or rational choice constructs (see also Anderson, 1986). While social-psychological models posit that voters develop a psychological attachment to the political system which then prompts their

participation in it, rational choice models suggest voters evaluate the costs and benefits associated with voting before deciding whether to exercise their right to vote. Thus the former model depends on some idea of socialisation, where voting becomes a norm adhered to by responsible individuals, while the latter rests on the notion of rationality, manifested by voters acting when the expected utility of doing so is positive. Turnout, according to the latter model, is higher when the electoral contest is close (respondents' votes could "count" more) and where concurrent elections are held (since these produce economies of scale).

Basing their model on rational choice theory, Chapman & Palda developed a regression equation where they viewed the total number of votes cast as a function of the number of registered voters, of the closeness of the contest, of the level of campaign expenditure, and of income, education, and employment variables. After examining the predictive accuracy of this equation across ten Canadian provincial elections, Chapman & Palda concluded that, while their findings provided strong support for the rational choice model, they could not be construed as disproving the social-psychological model.

However, despite the correlations they noted, the construction of their model depends on retrospective information (such as campaign expenditure and the closeness of the election). Thus while the model's explanatory power may be high, its role as a predictive instrument appears limited (see also Chandler's, 1988, comments). Nevertheless, Chapman & Palda confirmed the importance of education and income in turnout decisions, and both these variables may be employed in predictive models (although the extent of the multi-collinearity Chapman & Palda noted would require further investigation).

Rather than differentiating between economic and socio-psychological models, Fife-Schaw & Breakwell (1990) suggested that rational models have a good predictive ability because they incorporate the psychological factors that militate against voting. Acknowledging the increasing dominance of the Michigan model, Fife-Schaw & Breakwell criticised recent work conducted into non-voting for relying too heavily on demographic variables, and argued that the psychological factors underlying non-voting deserved greater attention. Among such factors they included: whether respondents felt their vote would count, or be a futile gesture; respondents'

understanding of philosophical differences between parties; their view of and interest in politics, and any feelings of alienation they might experience.

Fife-Schaw & Breakwell examined teenagers' attitudes, as well as socio-economic characteristics, and concluded that attitudinal variables were better able to discriminate between potential voters and non-voters than were socio-economic variables. This finding tends to support the Michigan school, though it does not prove that social class, as proposed by the Columbia school, has no role to play.

Straits (1990) also sought to explore and understand the social context of voter turnout, noting a variety of factors that affect turnout: "*The likely mechanisms by which social environments stimulate turnout include the channelling of political information, persuasive organisation, social pressure, co-operation and friendly competition*" (p. 64). In addition to these factors, he suggested that spousal influence has profound, but typically unmeasured, effects on respondents' likelihood of voting. Indeed, Straits concluded by suggesting that the poor predictive power of models based solely on individuals' views and behaviour could be largely due to their failure to take account of respondents' social environments and peer relationships (p. 72).

Although concentrating less on family relationships, Norrander (1991) also examined the various roles played by attitudinal, behavioural and demographic variables in turnout decisions. She examined the explanatory power of variables relating to the campaign environment itself, individuals' exposure to the campaign, respondents' voting history, their satisfaction with the incumbent president (or government), and their evaluation of the candidates contesting the election. Using presidential primary elections to test these variables, she concluded that key demographic characteristics, age and education, together with interest, combined to enable classification of core or habitual voters. However, she also noted that campaign characteristics, such as the closeness of the race, affected turnout. Since these factors varied greatly across campaigns, their specific relationship to turnout was more difficult to define consistently. Again, these findings suggest that a merging of the Colombia and Michigan schools may provide the best means of predicting turnout, a view shared by

Curtice & Jowell (1992).[4]

The latter researchers suggested that a combination of both schools, involving a representative study of voters, combined with a study of the social and political context within which they make their decisions, will provide the most accurate and informative predictions about their likely behaviour.

While they perhaps originate from different sources, it appears that both research paths are converging and that neither the Columbia nor the Michigan school dominates current turnout research. Equally apparent is the fact that the models put forward involve a number of variables, which themselves may not remain constant across elections, and which require quite detailed examination (c.f., Converse's, 1966, long and short term effects).

Overall, although researchers have explored many different avenues of enquiry, turnout prediction remains a complex task which has not yet been fully elucidated. Debate continues over whether turnout is a function of social class, of party identification or history, or of some combination of these and other variables. Irrespective of what motivates respondents to cast a vote, if respondents who have a low chance of voting also have a different party support distribution to those candidates who do plan to vote, the resulting estimates will be biased. Accurate differentiation between voters and non-voters in pre-election polls is therefore critical to the calculation of accurate party or candidate support estimates. Consequently, an instrument which provided a straightforward and accurate means of classification would solve a conundrum that has perplexed generations of researchers.

However, this is not the only issue to mystify researchers; an equally puzzling problem is that of allocating undecided respondents. Again, a variety of views exists, ranging from the dismissal of undecided respondents through to complex allocation models. The following sections examine the problems posed by undecided respondents and the methods currently employed to deal with them.

---

4 Crespi's (1988) discussion of attitudinal variables also noted that researchers typically used these measures in conjunction with other variables.

2.5 Undecided Respondents

Very few researchers investigating problems associated with election predictions have not acknowledged the difficulty of dealing with undecided respondents. Fenwick, Wiseman, Becker & Heiman (1982) stated: "*the voting behaviour of the undecided group can have a significant impact on the final results...* [and] *no matter which* [allocation] *approach is used, there is not yet an entirely satisfactory way of determining the likely voting (or non-voting) behaviour of those who claim they are undecided*" (p. 383-384). More recently, Daves & Warden (1993) expressed the same view: "*The public and pollsters judge a poll by how closely it reflects the outcome of the election. Polls with high numbers of 'undecided' or 'no opinion' responses cannot accurately reflect election day results*" (p. 2). There exists, therefore, a clear, perhaps even unanimous, view that incorrect allocation of respondents who are unwilling or unable to express their likely voting behaviour will introduce error into the sample estimates. The following sections explore this problem in greater detail before turning to examine the remedies that researchers have applied to it.

2.5.1 Undecided Versus No Opinion

Although some researchers do not differentiate between people who state they are undecided and those who claim to have no opinion on a particular matter, another group has argued that these respondents require separate treatment. As early as 1979, Perry had noted that sample members may describe themselves as undecided for a variety of reasons: "*reluctance to take a position, reluctance to refuse the interviewer and a desire to hide under the label of undecided*" (p. 317).

Researchers have long displayed an interest in the characteristics of undecided respondents and well over two decades ago Converse (1976) had explored the "no opinion" group, investigating whether respondents proffering no-opinion responses possessed characteristics that differentiated them from other sample members. Recognising that previous research had identified respondents with less education as more likely to give a no-opinion response than their more highly educated counterparts, she investigated whether certain question characteristics contributed to the proportion of no-opinion answers given. Overall, Converse found that, as the complexity of the task required of respondents increased, so too did the proportion of no-opinion responses. She also concluded that education remained the best predictor

of no-opinion responses.

Following on from this work, Francis & Busch (1975) examined non-substantive responses (NSRs) in general and, in confirming Converse's findings, also argued that NSRs were not a random phenomenon, but were related to respondents' characteristics or characteristics of the data gathering procedures. After examining data from surveys conducted around three presidential elections, they concluded that NSRs showed a consistent link with respondents' education, family income and gender. Francis & Busch also noted a relationship with respondents' level of political involvement: that is, the poorer their understanding of the role and power of government, the more likely they were to give NSR responses. Overall, Francis & Busch argued that researchers should not exclude these respondents from their analyses, nor should they simply combine them with other response categories. Unfortunately, however, they provided no concrete guidelines outlining procedures for researchers to follow when analysing and interpreting data containing NSRs.

Faulkenberry & Mason (1978) also examined further the relationship between respondents' interest in a topic and their likelihood of expressing an opinion on that topic. Differentiating between non-opinion and no-opinion response groups, they investigated attitudes toward wind energy and corroborated Francis & Busch's conclusions. Their examination of four groups (don't know, no opinion, oppose or favour) revealed that differences between the groups existed, and that undecided (or ambivalent) respondents typically had a higher level of knowledge and greater exposure to mass media than did those respondents who had no clear opinion on the topic. While these results depended on interviewers making standard and consistent classifications of respondents, and so may contain classification errors, they nevertheless added weight to the view that undecided and no-opinion respondents differ in some important ways (see also Gilljam & Granberg, 1993).

Feick (1989) sought to break down the undecided group further and argued that it comprised at least three separate groups: those who are ambivalent about the alternative response options from which they must choose (a phenomenon he described as item-specific equivocation); those who remained generally ambivalent about a topic, despite having considered it in detail (topical equivocation), and finally those who may give an undecided response because they have not considered a topic

and may not be aware of or interested in it (see also Duncan & Stenbeck, 1988).

Bishop, Oldendick & Tuchfarber (1984) had earlier investigated the first group Feick identified (those who are ambivalent in some way about the question or questions they are asked) and who Converse (1970) and Francis & Busch (1975) had earlier suggested were likely to have had less education than other respondents and to be less interested in the topic at hand. Bishop *et al* hypothesised that questions exploring respondents' interest in particular phenomena were likely to be answered according to respondents' most recent experience of or with those phenomena. As a consequence of this, they suggested responses were likely to depend heavily on the question context. To test this recency effect further they developed three questionnaires in which buffer questions were used to separate items they hypothesised would affect the responses to subsequent questions, if placed immediately prior to those questions. However, they found the intervening buffer questions did not reduce the order effects they had earlier noted, leading them to conclude that: "*context problems cannot be avoided merely by separating items that are known, or expected, to affect one another, though that may help in some instances*" (p. 518, see also Husbands, 1987).

In addition to querying the effects of buffer questions on responses, other researchers have suggested that filter questions may have some unfortunate side effects. Hippler & Schwarz (1989) argued that the use of filter questions may, instead of simplifying respondents' task, actually give respondents the impression that the task ahead of them is complex: "*Respondents who are exposed to a filter question expect more, and more difficult, follow-up questions, and doubt that they have sufficient knowledge to answer them*" (p. 86). Ironically, treatments designed to provide respondents with a context that may help consolidate their views may instead contribute to the proportion of undecided responses given.

In summary, a wider exploration of public opinion research in general suggests that undecided respondents may be expressing either ambivalence or a lack of an opinion on a topic. Although not conducted specifically in the area of voting behaviour, the research suggests that respondents expressing these views differ not only from other sample members, but also among themselves. Accordingly, researchers have recommended treating the groups as distinct, but have not yet established how best to

do this. At present, therefore, the techniques typically employed have tended to treat the two groups as one. The following section outlines some of the treatments and allocation methods used and analyses the success these have had in enhancing the accuracy of pre-election poll estimates.

2.5.2 Treatment of Undecided Respondents

Researchers undertaking pre-election polls and using traditional voting intention methodologies are almost certainly bound to encounter a group of respondents who are either unable, or unwilling, or both, to articulate clear responses to the questions posed. These researchers thus have two options: they can attempt to reduce, or even eliminate, the undecided groups, and thus dispose of the problem *a priori*, or they can devise methods of allocating the undecided groups and establish what are essentially *post hoc* solutions. This section first examines various methods used to reduce the proportion of undecided respondents before examining allocation models.

2.5.3 Methods of Reducing the Proportion of Undecided Respondents

Efforts which effectively reduce or eliminate the proportion of undecided respondents often prove easier to implement than allocation models (which depend on variables that may themselves be prone to non-response or non-substantive response error). Thus several researchers have tested treatments which result in a net reduction of the undecided group. In developing these treatments, researchers have attempted to identify underlying factors which may prompt or influence respondents' propensity to select undecided response options. Many of these factors are related to aspects of question wording. Thus Feick (1989) included misunderstanding, poor instructions, lack of attention, and ambiguous phrasing (p. 526) in his summary of factors which contribute to item ambiguity. Sigelman & Presser (1988) expressed a similar concern in more general terms, noting that: "*A basic concern of survey research is that alternative measurement procedures (e.g. question wording differences or differing survey contexts) can produce very different results*" (p. 326).

However, questionnaire construction cannot be held solely responsible for respondents' indecision; several researchers have already noted that respondents' own characteristics may also contribute to the number of undecided responses obtained. For example, Perry (1979) suggested nonresponse was also caused by evasiveness and a lack of

candour among respondents (see also Brusati, 1987). In addition to describing this rather furtive response behaviour, Perry also noted the pressures generated by social desirability bias. He suggested this pressure might lead respondents to offer a neutral response (such as no opinion or undecided) rather than engage in the less helpful behaviour of refusing to comment on a particular item or declining to participate in the interview altogether.

Perhaps noting Perry's comments, other researchers have documented the need to establish a rapport with respondents, suggesting that, if this is achieved, respondents will relax and reveal personal opinions or details that they might otherwise decline to impart (see Perry, 1979; Sopariwala & Roy, 1990). Husbands (1987) referred to this as the "warm start/cold start" controversy (p. 407), and noted that while developing a relationship with respondents may lead to greater disclosure of personal details, the addition of questions designed to create this rapport may actually bias responses to subsequent questions. In other words, while developing a rapport with respondents may improve the accuracy of the answers they give, it may also detract from this. Crespi (1988) noted this problem "*...asking questions about issues and the candidates before asking for candidate preference can shape the 'perceptual environment' and bias poll results.... It might be that voicing opinions on issues creates a preference where one did not previously exist. Or it might be that having voiced opinions on issues, respondents may then feel constrained to voice a preference between candidates as well*" (pp. 104-105).

In addition, as an interview proceeds, respondents may become more aware of divergences between their views and those they (rightly or wrongly) perceive to be held by the interviewer or their wider reference group. Noelle-Neuman (1970) suggested this realisation could lead to a "spiral of silence" whereby respondents who hold views which contravene perceived prevailing norms are less and less likely to express their views as the interview proceeds. Thus as the disparity between respondents' opinions and their perception of an "external" view increases, so too does their inclination to conceal their views and their propensity to give undecided or no opinion responses.

The above section represents a very brief overview of factors researchers have

identified as contributing to undecided responses. It is clear from this overview that many factors relate to respondent misunderstanding, social desirability bias, and other issues of question wording and interpretation. These matters are discussed in detail in Chapter 3 and are only summarised here. The remainder of this section discusses researchers' attempts to reduce the proportion of undecided responses.

Perhaps the most obvious way to deal with an undecided response is to probe the respondent's views further. Some researchers have done this by seeking to understand the motives that may prompt respondents to support particular parties (see Antaki, 1988, for example), or by assessing the level of certainty they associate with their voting intention (see Bassili, 1993). However, a more widespread practice involves repeated questioning of undecided respondents. Thus instead of simply accepting an undecided response, interviewers proceed to investigate respondents' leanings and, where these exist, they use them as surrogate preference estimates (see Traugott & Tucker, 1984; Crespi, 1988; Daves & Warden, 1993, or Green & Schickler, 1993 for a discussion of this approach). This approach can substantially reduce the proportion of undecided respondents in the sample; however, doubts remain over whether extensive probing elicits thoughtful and accurate responses. Crespi (1988) summarised this debate by questioning whether detailed probing merely elicited non-attitudes or whether it uncovered previously hidden refusals. Thus despite the widespread adoption of this technique in the United States, at least, there is no empirical evidence attesting to the accuracy of responses obtained in this way. Rather, at least some evidence to the contrary exists.

Sanchez & Morchio (1992) explored whether persistent probing improved the quality of the estimates they obtained. They reported: "*it would appear that exhaustive probing on the part of TEL* [telephone] *interviewers manages to encourage the less informed respondents to voice a substantive choice, but that the choice involved a large amount of guesswork*" (p. 463). Thus, although probing left fewer undecided respondents, Sanchez & Morchio concluded that guesswork was largely responsible. Although these findings cannot be treated as definitive, they do raise questions about the overall accuracy of responses given to questions designed to establish undecided respondents' political leanings. Clearly, the reduction of error introduced by undecided respondents remains a primary goal, but this is unlikely to be addressed

satisfactorily by the use of a technique which may concurrently introduce and remove error.

Probing respondents' answers further has not been the only approach taken to questionnaire construction that could reduce the size of the undecided group. Lampert (1978) and Lampert & Tziner (1985) suggested that polls produced high numbers of undecided respondents and inaccurate predictions because they did not use instruments that enabled respondents to integrate their attitudes with their intentions. To address this problem, Lampert developed a pollimeter, an instrument similar in appearance to a slide rule, which respondents used to indicate their support for the two candidates located at either end of the continuum. Crossing the scale mid-point was interpreted as support for the candidate whose "zone" the respondent had just entered. Likewise, the closer to the mid-point respondents placed themselves, the greater their chances of switching their support to another party or candidate.

Lampert argued that this instrument had two chief benefits: its ability to force respondents to choose between competing candidates or parties, and its ability to provide insights into those respondents placing themselves near the mid-point (who would otherwise have classified themselves as undecided). In addition, he argued the pollimeter would provide quick results which would be analogous to those based on more direct intention questions.

However, while Lampert outlined an elegant description of his instrument's benefits, he did not present empirical support attesting to the pollimeter's predictive accuracy. Nor did he address a key problem: how researchers should deal with a situation where an election is contested by more than two main candidates or parties. While this instrument may work effectively in two-party contests, many countries have a more diverse political environment, and Lampert did not discuss how the pollimeter might be adjusted to deal with this.

In a subsequent paper (Lampert & Tziner, 1985), Lampert addressed some of these issues and outlined a pilot scale test of the pollimeter. However, the findings from this paper revealed that 20% of the 186 respondents classified themselves as undecided, presumably by placing themselves on the very mid-point of the pollimeter. Furthermore, a discriminant model incorporating the attitude and intention variables

as measured by the model had variable success in predicting how the undecided respondents would behave (47% were correctly classified when attitudes were used as the main discriminating variables). While the pollimeter instrument Lampert developed does not appear to have provided respondents with an appropriate medium for combining their cognitive and conative responses, this finding does not constitute a rejection of Lampert's original hypothesis (rather it constitutes an inadequate test of it). Thus the original question Lampert raised, whether present polling research instruments can adequately accommodate and capture the complexity of voting intentions, remains a viable and unresolved question.

Panagakis (reported in Crespi, 1988) used a constant-sum approach analogous to Lampert's pollimeter and asked respondents to allocate 10 votes among all the candidates standing for a particular office. However, according to Crespi, Panagakis used this technique to measure respondents' attitudes: "*The implicit assumption of this strategy is that for accurate prediction it is crucial to measure attitudes in their complexity, something that summary, forced-choice measures of attitude cannot accomplish*" (p. 102). If this is an accurate representation of Panagakis' work, it appears to place more emphasis on cognitive variables that it does on behavioural factors. Chapter 4 examines the attitude-intention-behaviour relationship in more detail.

Several researchers have alluded to confidentiality issues, noting that respondents may feel reluctant to divulge their personal views to someone they had never previously encountered (Brusati, 1987; Perry, 1979; Butler, 1994). One method used to address this problem involves the administration of a secret ballot. Although difficult, if not impossible to administer via phone interviews, secret ballots have been used successfully in both pre-election and exit polling. Perry (1960 and 1979), who documented some of the history of secret ballots, noted that when introduced immediately following the survey introduction they were associated with a higher level of refusal. Later, however, secret ballots were incorporated as part of a standard interview, and were administered after respondents had become more accustomed to the interviewer. In this context, split-sample experiments revealed that secret ballots yielded more accurate estimates than direct questioning approaches. In addition, Perry found that secret ballots reduced the proportion of undecided respondents more than

leaning questions had done.[5] He concluded: "*It* [these results] *suggests that the relatively high undecided percentages sometimes reported in general elections are artifacts of the method of questioning and do not represent true undecided*" (1979, p. 317). Interestingly, Perry's comment suggests that the problem of confidentiality should be addressed though changes in the question administration, rather than through the construction of a series of rapport building questions.

Perry (1960 and 1979) noted that while the enhanced confidentiality contributed to more accurate estimates, the secret ballot's close approximation of actual voting procedures also enhanced the accuracy of the estimates.

In addition, Perry examined the question of data accuracy, and explicitly acknowledged that a reduction in the proportion of undecided respondents is not necessarily equivalent to an increase in the accuracy of the predictions elicited. He noted that the change in party support distributions of matched samples administered either via a direct or a secret ballot questionnaire was often greater than the reduction in size of the undecided group. However, Perry nevertheless concluded that there appeared to be no over-riding reasons why secret ballots should not produce reliable estimates. Thus he asserted: "*Experience has shown that a confidential voting choice in a survey gives every indication of producing valid responses from all but a negligible number of respondents*" (p. 319).

Fenwick *et al* (1982) also acknowledged the role of secret ballots and "leaning" questions in reducing the undecided group and suggested that the latter were typically used in phone interviews, where the characteristics of the interview medium prevented the use of secret ballots. Although not traditionally a problem encountered in Western democracies, Sopariwala & Roy (1990) noted that where respondents may fear unpleasant repercussions, or even political violence, should they express unsanctioned views to "official" interviewers, secret ballots provided a medium which reduced this fear and allowed them to express their true views in privacy.

---

5 Leaning questions reduced the initial 17% undecided to 9% while a secret ballot resulted in 6% undecided.

Butler's (1994) recommendations for improving UK polling accuracy also included a suggestion that secret ballots be investigated further to asses whether they could reduce the proportion of people giving undecided responses to the voting intention question.

Overall, there appear to be a number of procedures researchers can adopt to decrease the size of the undecided group. The methods outlined above all relate to changes in the questionnaire or the way in which interviewers administer this. However, aside from these internal research decisions, researchers have also noted that external factors (such as the timing of the poll) rather than internal factors (such as questionnaire design or administration) account for at least some of the undecided responses given.

Felson & Sudman (1975) suggested that polling accuracy in general was related to the gap between the date of the poll and the election itself, and that the smaller the time lag, the more accurate the poll was likely to be (see also Crespi, 1988). Although they did not specifically relate the improvements in accuracy they noted to a reduction in the size of the undecided group, others have argued that early polls inevitably record higher proportions of undecided respondents since some people have not yet formed or consolidated their voting intentions (see Lewis-Beck & Rice, 1982; Weimann, 1990, Worcester, 1992a, but also Campbell & Wink, 1990).

This argument inevitably raises the question of why pollsters continue to conduct surveys some months prior to an election date. The newsworthiness of poll results appears an obvious answer to this question. Although large time lags between a poll and an election date always leave open the possibility that unanticipated events will lead to changes in respondents' stated intentions, the media seem unlikely to desist from polling. Consequently, the quest for stable and accurate methodologies continues.

Overall, secret ballots appear to reduce any real or perceived threat associated with voting-intention questions; they also represent an alterative means of decreasing the proportion of undecided respondents. Practically, however, secret ballots present some logistical problems and it is not clear whether the improved accuracy they obtain offsets the difficulties incurred in their administration. However, neither secret ballots (which increase the confidentiality of respondents' answers) nor leaning questions

(which help clarify the stance taken by ambivalent respondents) completely eliminate the undecided group. Almost inevitably, therefore, researchers must decide how to treat those respondents who, for whatever reason, decline to express a firm voting intention. The following section examines some of the techniques used to deal with this latter question.

2.5.4 Allocation of Undecided Respondents

This second approach to the treatment of undecided respondents aims less to prevent, than to ameliorate the problem they present. As well as reducing the overall size of the undecided group, researchers allocate undecideds among the parties or candidates specified in the voting-intention question. The simplest method of allocation excludes undecided respondents from any subsequent analyses. In effect, this procedure allocates undecideds proportionately. Whether acknowledged explicitly or not, this practice assumes that undecided respondents have the same overall probability of exercising their vote as those who did nominate a choice, and that their pattern of voting will mirror that of respondents who stated an intention. However, neither the validity of these assumptions, nor the overall logic of this approach has been clearly established.

Converse's (1966) notion of a normal vote distribution appears to provide a rationale for allocating undecided respondents on the basis of partisan affiliations in the same proportions as equivalent party identifiers who have expressed a preference (see Traugott & Tucker, 1984; Crespi, 1988). His theory appears analogous to leaning questions, described earlier, as it uses an expression of ideological affiliation as a surrogate for specific voting intention. However, it involves a further step in logic to allocate undecided respondents in the same proportions as more committed party supporters when no clear evidence of a partisan affiliation exists. Yet this widespread practice appears to have received little attention in the literature.

The apparent lack of evidence to support the proportional allocation of undecided respondents has led researchers to devote considerable efforts to devising allocation models and it is surprising that some of these are not more widely used. For example, Fenwick *et al* (1982) described a model which uses known attitudes, candidate evaluations, and demographic variables to explain the preferences held by respondents

who did express a voting intention. Using discriminant analysis, they allocated these respondents to one of the three candidates contesting the election standing and found that the model assigned 86% of the cases correctly (i.e., to the same candidate nominated by the respondent). While acknowledging the problems implicit in using the same data set both to devise and to test the model, Fenwick *et al* nevertheless argued that this level of accuracy was well above what could be expected from chance alone. Furthermore, they claimed the model represented an improvement on the more traditional method of proportional allocation. Tests of the method on undecided respondents showed 86% were correctly allocated.[6]

Use of discriminant classification models has characterised the research undertaken in this area. Daves & Warden (1993) also developed and tested a discriminant model, together with three other allocation procedures: proportionate allocation among the candidates, equal allocation, and allocation to the major challenger. The discriminant model utilised demographic variables (gender, age, income and education) and party identification, and so was very similar to Fenwick *et al*'s model. However, despite the obvious differences between the methods and their underlying rationale, the results each achieved were not greatly different. Although allocation to the main challenger resulted in less accurate predictions, the other three methods generated a similar error level, leading Daves & Warden to conclude: "*assigning 'undecided' respondents in pre-election polls to a candidate is indeed a chancy business, no matter which method is used*" (p. 20).

Hoek & Gendall (1995) also reported using a discriminant model to classify undecided respondents. Developed and tested in the same way as Fenwick *et al* (1982) described, their research compared the accuracy of models based on demographic variables, or on opinion and behavioural variables, or on an amalgamation of these. They reported low accuracy of the demographic and opinion and behaviour models (both classified just under a third of respondents correctly), although this improved to around 50% when variables from other models were combined (c.f., Ray, 1984, who

---

6 The proportion of undecided respondents correctly allocated was ascertained though post-election interviews. Thus the information used was not actual, but self-reported behaviour. While there is no overt reason to suspect bias, it is nevertheless possible that bandwagon effects (documented by Bradburn & Sudman, 1988, among others) may have affected respondents' answers.

claimed that information on respondents' political attitudes and conservatism could explain a high proportion of the variance in their voting behaviour).

Hoek & Gendall's work was conducted during a turbulent electoral period, thus complicating the use of historical variables, and none of the discriminant models demonstrated a comparable level of accuracy to simple proportional allocation.[7] This led them to conclude that researchers should direct their energies toward reducing the overall size of the undecided group since, if this is small enough, choice of allocation procedure becomes largely academic.

Shamir (1986) identified possible reasons why discriminant models had displayed variable success. He suggested some allocation variables might be very situation-specific, and argued that even apparently objective classification measures such as the party or candidate voted for in the last election may have low validity because of false reports.[8] Furthermore, he also noted that the theoretical foundations underlying classification rules were also quite weak, while the boundaries separating support for one party from another were not always straightforward. However, while Shamir's comments recognised some very real problems, they did not resolve any of the difficulties associated with the treatment and allocation of the undecided group. Thus, unless, as Hoek & Gendall suggested, pollsters manage to reduce the proportion of undecided respondents to a negligible size; their options remain limited to a range of procedures which tend to have high error levels associated with them.

Neither Fenwick *et al*'s nor Daves & Warden's models included a weighting factor to account for the likelihood that an undecided respondent will vote.[9] Implicitly, therefore, these models assume undecided respondents have the same probability of voting as respondents who did express a view. Perry (1973) examined this question and reported that his analysis of respondents' voting preferences and their overall

---

7 The average mean absolute error of the discriminant models was 10.3% compared to 2.3% for the proportional allocation method.

8 Although Granberg & Holmberg (1986), discussed in Chapter 6, argued that the accuracy of respondents' recall is less important than the ideological affiliation it conveys.

9 Hoek & Gendall's model did incorporate a variable examining respondents' overall probability of voting. However, because the number of non-voters was very low (<10), this group was not included in the final version of the model.

likelihood of voting revealed that undecided respondents were less likely to vote. However, Hoek & Gendall's (1993) work with the Juster Scale, an 11 point probability scale, suggested that undecided respondents had a 75% chance of casting a vote.[10]

As in the modelling of turnout, researchers have brought divergent approaches to bear on this issue. The preceding section has documented the more behavioural route some have followed; the next section examines how researchers have also used a wider range of variables to assist their prediction of likely voting behaviour.

Although not eschewing the use of ideological or historical variables, other researchers have incorporated economic measures into their prediction models on the grounds that these reflect the success or failure of policies implemented by the incumbent president or government. Thus Campbell & Wink (1990) argued that most forecasting models are based on the notion that people's evaluation of the president's (or government's) performance is the precursor to their voting behaviour. Thus they suggested that models incorporating candidate popularity scores, measures of economic conditions and a variable accounting for the length of term, could account for over 90% of the variance in presidential vote (p. 252). They tested their model using presidential trial-heat data and found that while the addition of economic data greatly improved the model's predictions, candidate evaluation scores were necessary to obtain maximum accuracy.

Lewis-Beck & Rice (1984) had earlier developed a forecasting model using economic as well as ideological variables. After investigating economic performance, international involvement, political experience and presidential popularity, they found that only the latter variable showed a strong predictive ability when used as the basis of bivariate forecasting models. However, combining the popularity and GNP variables resulted in a model at least as accurate as the final Gallup polls taken before that election. Nevertheless, Lewis-Beck & Rice, while concluding that the model's performance was much better than chance, conceded that changing political environments might also change the predictive power of the model's variables.

---

10 Other sub-samples had the following probabilities of voting:

Decided respondents	86%
Declared non-voters	30%

Overall, these models offer some appeal in an environment where two clear alternatives compete (such as in presidential elections). Yet they appear less suited to environments containing multiple parties, some of which are not clearly differentiated from others. Moreover, where several minor parties exist, it will be difficult, if not impossible, for respondents to assess the likely economic effects of the policies issued by those parties.

Researchers have also attempted to define the psychological processes underlying or mediating political attitude-formation, in the hope that such knowledge might improve the accuracy of prediction models (see Williams, Weber, Haaland, Mueller & Craig, 1976; Fishbein & Coombs, 1974).

Newman & Sheth's (1985) work went beyond the prediction of undecided voters' behaviour; their discriminant model sought to predict the behaviour of the whole sample and incorporated a number of cognitive factors including: policy issues; social imagery; emotional reactions to candidates; candidate image; current events; personal events, and epistemic issues. Using a mail survey questionnaire, they examined respondents' political background as well as these seven belief components. These pre-election interviews elicited a 78% response rate and were followed by post-election phone interviews designed to ascertain how respondents had voted.

Newman & Sheth constructed a number of discriminant models which they used to allocate voters to one of the two presidential primary candidates and then compared this to respondents' self-reported voting behaviour. However, the final model's accuracy of 90% is undermined by the unexplained exclusion of nearly three quarters (476) of the respondents originally interviewed. If this is put to one side, though, and relative assessments of the initial models made, the results showed that the attitudinal model correctly allocated more respondents than did a model based solely on demographic variables or on variables representing respondents' level of political involvement. Yet Newman & Sheth warned that there was no guarantee all seven dimensions would necessarily be relevant in the prediction of subsequent elections. In addition, they noted that the timing of the early qualitative work, which was required well in advance of the election in order to develop the more quantitative questionnaire, introduced a risk that the time lag may result in shifts in respondents' preferences which would ultimately affect the accuracy of results produced by the

model.

Despite the work conducted into the construction and refinement of classification models, researchers have made little definitive progress. For some time, researchers have highlighted the arbitrary choice necessitated when selecting an allocation procedure. Indeed, some have suggested that it is more productive to reduce the size of the undecided group than it is to expend effort developing an allocation model. Crespi's (1988) comments support this view: "*Pre-election polls that allocated the undecided vote themselves or reported a small undecided vote are significantly more likely to be accurate than those that reported a moderate or large percentage undecided .... it is... clear that minimising the undecided vote contributes to accuracy*" (p. 109).

While researchers have tested and developed several allocation models, those demonstrating the highest predictive ability have typically been used in a very specific context. Likewise, a number of researchers have commented on the situation-specific nature of at least some of the variables they have incorporated in their models. Thus the problems Converse (1966) dealt with when conceptualising and implementing his normal vote model still appear to beleaguer researchers today. While this does not mean researchers should desist from pursuing research into classification models, it does suggest that other lines of enquiry may offer more promise.

2.6 Summary

Although researchers have conducted work into a wide range of factors which could help explain the rather variable accuracy of opinion polls conducted around the world, two key problems have dominated much of the literature: the identification of likely voters and the classification of undecided respondents. While other issues, such as the sampling procedures used and the degree to which unavailable respondents are pursued, remain important issues, the central concerns of this thesis are with the former two problems.

A great deal of work into turnout prediction has been conducted, and researchers have developed several models which appear, in some cases at least, to show strong predictive power. However, many of these models are complex to construct and some

appear better suited to retrospective explanation than to prediction. Although complexity should not be a reason for dismissing a strong predictive tool, simpler mechanisms offer a clearer appeal and more obvious advantages to pollsters.

As well as modelling turnout, researchers have also attempted to explain how and why inaccurate estimates may have arisen and, while a number have related this problem back to sampling inadequacies, many have noted that issues related to question wording may also have affected the estimates.

In addition, researchers have sought to identify the various roles played by environmental factors, such as social class or historical developments, and by turnout. Overall, while methods of identifying likely voters and of predicting turnout do exist, these have not resolved all the problems facing researchers and turnout prediction remains a difficult and only partially elucidated task.

Similarly, the allocation of undecided respondents continues to intrigue researchers and confound pollsters. While researchers have identified the existence of several different groups within the general undecided category, this finding has not yet been formally translated into procedures for allocating these people.

However, work exploring the factors underlying indecision has identified results similar to those put forward by turnout researchers: question wording, respondent understanding and social constraints all seem to play a role in influencing the proportion of undecided respondents. While some procedures for dealing with these factors exist, their adoption, utilisation and validation by researchers has been piecemeal.

As a consequence, many pollsters still need to allocate undecided respondents. Yet of the available allocation procedures, most do not have a record of strong predictive accuracy.

Clearly several factors influence respondents' decision to vote and their choice of which party or candidate to support, and researchers face many difficulties if they attempt to chart and measure each stage of these inter-linked decision processes. Yet researchers can manipulate variables within their control and then ascertain the effect

of these changes on respondents' ability to predict their behaviour accurately.

Many researchers working to produce turnout or allocation models have alluded to question wording and the related issue of respondents' understanding. Since researchers can do a great deal to change and clarify the questions which they put to respondents and the ways in which interviewers, working on their behalf, administer these questions, the research reported here seeks to extend the current understanding of question wording and administration. The following chapters therefore explore first the general problems of question design before moving to document the rationale underlying an alternative approach to the estimation of turnout and the prediction of election outcomes in general.

CHAPTER 3

QUESTION DESIGN EFFECTS AND RESPONDENT BIASES

3.1 Introduction

Problems of question wording and questionnaire design have troubled and intrigued researchers for many years. Few academics or practitioners would dispute that the words used to frame a question, the context in which it appears, the way in which interviewers deliver it, and the intangible social connections which develop between respondents and interviewers may affect the responses elicited. Kalton & Schuman (1982) summarised the difficulties: *"The survey literature abounds with examples demonstrating that survey responses may be sensitive to the precise wording, format and placement of the questions asked.* " (p. 42). However, the magnitude of these effects, and the topics and respondents most susceptible to them, remain difficult to quantify.

In addition to addressing question design and context effects, and issues relating to the perceived saliency of the research topic, researchers must also consider how best to minimise the biases respondents themselves might introduce. Issues such as the extent to which respondents attempt to perform particular roles, and the relationship between these roles and the accuracy of respondents' answers, all require detailed attention.

This chapter examines these problems and their effects on the information given by respondents. It begins with a general overview of questionnaire design effects and then considers the more specific issues of question order and context. Next the chapter explores respondent biases, particularly the interlinked problems of respondent ignorance and the perceived social desirability of some responses. Finally, possible solutions to these problems are presented and a rationale for their exploration and inclusion in the research methodology is outlined.

3.2 Overview of Questionnaire Design Effects

Questionnaires consist of several layers, yet researchers can rarely deal with these discretely. Instead, their decisions about the words to use, the question structure these words support, and the types of questions employed must all, paradoxically, be made both consecutively and simultaneously. Labaw (1980) argued that questionnaire

design is analogous to painting: *"While a painting consists of such components as lines, lights and darks, forms, composition, colours, and actual application of the paint, the artist cannot use a step-by-step sequence, handling one component entirely, followed by the next. All the components must be integrated and handled simultaneously"* (pp.12-13).

The following sections explore a number of questionnaire design decisions, and the approaches researchers may take to them, in more detail.

3.3 Question Wording

Rugg & Cantril's (1942) widely cited illustration of the problems caused by apparently simple changes in wording alerted researchers to the fact that even what appear to be logical substitutions for words in questions may generate markedly different response distributions.[11] However, Adams (1956) queried this finding. He substituted the phrase "better than" for "worse than" and found little difference in the resulting response distributions. He therefore concluded that respondents appeared equally able to express strongly favourable and strongly unfavourable opinions.

Clearly some words and phrases have more straightforward synonyms and antonyms than others. Since Rugg & Cantril's work, many researchers have identified particular problem words which generate ambiguous connotations and which may confuse respondents (see Payne, 1950; Belson, 1981; Converse & Presser, 1986, for thorough overviews of problematic words and expressions).

Other work to explore the ambiguity of certain words investigated whether some question formats produced more stable responses than others. Gendall & Hoek (1990)

---

11 Rugg & Cantril substituted the words "allow" and :"forbid" in the following questions:

"Do you think that the United States should allow public speeches against democracy?"

and

"Do you think the United States should forbid public speeches against democracy?"

The first question elicited support for "free speech" from 21% of the respondents questioned. However, the second question, which was expected to elicit around 21% of negative responses (making it equivalent to the first question) was denied by 39%.

devised a series of mirror image attitude statements[12] and examined the distribution of responses to these. Their results confirmed the "allow/forbid" findings established earlier (Rugg & Cantril, 1942; Schuman & Presser, 1981; Hippler & Schwarz, 1986) and found a number of other significant differences between the distributions to ostensibly reversed statements. Gendall & Hoek (1990) confirmed that these differences were not solely due to the use of antonyms in the matched statements (which might be expected to generate additional ambiguity). Many differences occurred even when the only change was the addition of a negative (not) to the statement, thus suggesting that respondents did not interpret the statements as logical opposites. Apart from this general conclusion, Gendall & Hoek found it difficult to identify a pattern that could be used to explain their results and predict others.

Gendall, Assendelft & Hoek (1991) explored these issues further and tested whether presenting forced choice questions as mirror images would produce stable responses. While some questions had very consistent response distributions across versions, others showed less stability. Again, the lack of a clear pattern in these results led Gendall *et al* to reiterate their earlier view that even very subtle changes in question wording may produce quite different responses (see also Stubbs & Hutton, 1976, for a discussion of a similar study).

Given the dynamic and metaphoric nature of language, problems caused by the substitution of one word for another are unlikely to disappear. Rather, as new words become introduced, and as semantic shift alters the meanings associated with existing words, these problems appear likely to increase. Belson (1981) recommended comprehensive pre-testing to identify, and then eliminate, words which respondents interpret to mean something other than the researcher intended. Researchers using the double-back technique he devised begin by administering a question to a respondent. Once the respondent has answered the question, researchers then explore the meaning

---

12 Mirror image attitude statements are those which are logical opposites. For example, some of the statements tested by Gendall & Hoek included:

"The law should allow/forbid public speeches which promote racism."

"The police should be/should not be armed".

that respondent associated with key words or phrases in the question. The ensuing discussion identifies the question the respondent believed he or she answered and highlights discrepancies between the information researchers sought to collect and that which the respondent gave.

While pre-testing methods remain open to debate (a topic covered in more detail in Chapter 6), Belson's general advice has also been advocated by many major methodological researchers (see Schuman & Presser, 1977; Sudman & Bradburn, 1982; and Converse & Presser, 1986, for example). The necessity of pre-testing is clearly a widely accepted survey research principle, even if individual practices vary.

Just as specific words can generate confusion, so too can the questions themselves. While the individual words and phrases that make up the question may be clearly interpreted by respondents, it is nevertheless possible that the question itself is biased. For example, it may not include all the possible response options, or the question may make some responses appear more acceptable than others. If this is the case, the answers elicited may be neither reliable nor valid. Section 3.4 below addresses more fully problems associated with question design effects.

3.4 Types of Questions

Researchers must make many question design decisions as they craft a questionnaire. One of the most fundamental of these involves deciding whether to use open or closed questions. Use of closed questions then necessitates decisions about the response categories made available to respondents, the scales employed, and whether an explicit undecided option is provided.

3.4.1 Open versus Closed Questions

Research investigating the relative merits of open and closed questions has been well documented in the literature (see Payne, 1946; Bradburn & Sudman, 1979; Labaw, 1980; Sudman & Bradburn, 1982, Kalton & Schuman, 1982; Converse & Presser, 1986; Converse & Traugott, 1986). There is a general consensus that although open-ended questions elicit a wider range of responses, they do not allow researchers to gauge accurately the proportion of the population who hold the views or responses elicited. Closed questions, on the other hand, will produce more accurate estimates, since each respondent must select from the same response options. However, closed

questions may not include all the alternatives that respondents would have selected, had they been made available in the original list. In an attempt to deal with this problem, some researchers have advocated the use of combined open and closed questions. Questions designed in this way present respondents with a list, but also allow them to nominate ideas or categories not contained in the list. However, Schuman (1986) found few respondents proffered a response outside those explicitly made available by an interviewer; thus he argued the additional information gathered by an "other" category is negligible (see also Gendall & Hoek, 1990).

Perhaps because of the time, difficulty, and cost of coding open-ended questions, as well as for methodological reasons, most experienced researchers use open-ended questions sparingly. Instead they employ closed questions, using either specifically designed response categories or scales widely employed in survey research (e.g., a Likert scale, or a semantic differential scale). However, closed questions do not present an unambiguous solution to the problems posed by open questions. Researchers have debated whether response categories should be balanced, to avoid predisposing responses in either direction, and whether response lists should contain an explicit "don't know" option, or whether respondents should have to volunteer this response.

Kalton, Collins & Brook (1978) tested the specificity of response options provided to respondents and found that the effects of offering no specific alternative, a token alternative only, or a substantive alternative, varied across the topics examined. Noelle-Neumann (1970) had earlier argued for the explicit inclusion of alternatives, suggesting that failure to provide these would mean some respondents did not make the trade-off necessary before they could provide an informed and accurate response.[13]

---

13 Noelle-Neumann asked a split sample of housewives one of two questions:

"Would you like to have a job, if this were possible?"
or:
"Would you prefer to have a job, or do you prefer to do just your housework?"

Of those answering the first question, 17% said they would prefer to have a job, 38% said they would like part-time work, 19% said they would prefer to do just [sic] their housework and 26% were undecided. The responses to the second question were 10%, 14%, 68% and 8%, respectively. Clearly, these questions varied in other respects aside from the explicit inclusion of an undecided alternative (one used the verb "like" while the other employed "prefer", and only the first question contained the phrase "if this were possible"). However, it would be extremely unlikely if the inclusion of an alternative had not contributed to the large variations in responses.

Although the research into this topic is limited, it is possible that the effects depend to some extent on the saliency of the topic explored. That is, where respondents have some familiarity with and knowledge of the options, the inclusion of specific alternatives is less necessary. However, where this is not the case, their omission may lead to substantial biases, as Kalton *et al* (1978) suggested: *"There is a danger that respondents replying to an opinion question will choose one of the alternatives offered even when they hold no view on the issue under study."* (p. 153).

Assuming questions do include specific alternatives, researchers must then decide whether, and how, to deal with the problem of respondents volunteering opinions when in fact they hold no view on the topic. Perhaps the most obvious response to this question has been the inclusion of middle categories, which effectively allow respondents to declare their indecision, or to nominate a neutral position. Schuman & Presser (1977) explored this issue and compared the distribution of responses where middle alternatives were and were not offered. They found that more people selected a middle position or an undecided option when it was included. However, they also noted that including these response alternatives did not affect the ratio of people selecting the polar options provided.

Since researchers often have to comment on the proportion of people selecting each response category, the decision whether to include an undecided option has important ramifications. Making this option explicit will result in a higher proportion of undecided respondents and, where the size of this group exceeds the proportion who expressed a definite view, will almost invariably question the stability of those results. However, requiring respondents to select a specific option may also produce unstable results, since a number of those classified as holding a particular view may not do so. This dilemma has led many researchers to recommend that less emphasis be placed on specific results and more attention paid to responses across a variety of questions (see Schuman, 1986). As Kalton & Schuman (1982) proposed: *"By tying an important concept to at least a few items that differ among themselves in form, wording and context, the investigator is unlikely to be trapped into mistaking a response artifact for a substantive finding."* (p. 55).

Schuman & Presser's (1977) results suggested that the inclusion of a middle

alternative or an explicit undecided option did not affect the response behaviour of people holding a definite opinion. Nevertheless, researchers have devoted considerable effort to developing techniques that will reduce the proportion of people likely to proffer an undecided response. Some have devised opinion screening questions which explore whether respondents hold a view on a particular topic. Only if the response to the filter question is affirmative, do interviewers proceed to administer the actual opinion question (Kalton et al, 1978; Bishop, Oldendick, & Tuchfarber, 1980a; Bishop, Oldendick, Tuchfarber & Bennett, 1980b). However, the potential danger this presents is that respondents, learning that a declaration of no opinion leads to the avoidance of questions, may claim to have no opinion simply to expedite the completion of the interview. Considerable evidence suggesting that respondents learn how to skip sections of questionnaires employing routine formats already exists (see Noelle- Neumann, 1970; Labaw, 1980; Sudman & Bradburn, 1982). Thus regular usage of filter questions may increase the risk of screening out people wishing to bring an interview to a rapid close as well as those who are genuinely undecided.

Where respondents select their response from a list of options presented on a showcard, read by an interviewer, or contained within a self-completion questionnaire, there is a tendency for them to choose items positioned at the top or bottom of the list. Researchers have suggested this occurs because the items' position makes them easier to recall or nominate, rather than because respondents' views align with those options (see Payne, 1951; Brook & Upton, 1974; Schuman & Presser, 1977; Kalton *et al*, 1978; Belson, 1981; Kalton & Schuman, 1982; Sudman & Bradburn, 1982). This phenomenon, known as item order effect, has led researchers employing closed lists to rotate the order in which they present the response options. Where rotated lists obtain similar response distributions, researchers can simply combine the two sets of responses. However, where the distributions differ, Noelle-Neumann (1970), among others, has argued that researchers should not amalgamate the two sets of data without first considering the extent to which the differences may arise from variations in the sub-samples, or from instability in the list itself.

Smith's (1987) review of surveys spanning fifty years identified a number of questions which had changed little during that period, despite changes in the survey topics. Just

as questions survive through time, so too do some of the decisions researchers must confront. As well as considering the structural aspects of question design and wording, researchers need also to consider whether they question respondents about their opinions or their behaviour. While it is often desirable to include both types of questions in a questionnaire, they vary in important aspects. Sudman & Bradburn (1974), among others, differentiated between behavioural questions, the responses to which can, in principle, be verified by reference to some external reality, and opinion or attitudinal questions, which cannot be verified in the same way. Since many polls attempt to predict respondents' future voting behaviour,[14] but also usually include a number of opinion questions, the following section examines the problems associated with both types of questions.

3.5. Opinion and Behavioural Questions

The first problem researchers face is that of defining different types of questions. Kalton & Schuman (1982) highlighted the difficulty of defining precise question classifications and drew a similar distinction to that proposed by Sudman & Bradburn, (1974), who attempted to differentiate between factual and opinion questions. They suggested that questions attempting to establish respondents' knowledge, reasons, preferences and perceptions combined aspects of their behaviour and opinions on particular topics, and they argued that clear boundaries between question types did not always exist. Like Sudman & Bradburn (1974), Kalton & Schuman argued that, rather than attempting to define the characteristics associated with different types of questions, researchers should instead consider issues of accuracy in greater detail. While noting that researchers could ascertain the accuracy of strictly factual or behavioural questions by comparing the estimated and actual outcome of an event, such as an election, they also recognised that validating non-factual questions presented researchers with a number of problems, not least of which was the lack of an event against which to compare the results.

The theories on which factual questions are based and test are themselves tested by the relationship between the predicted and actual outcomes. However, theories tested

---

[14] Chapter 2 contains a discussion of the debate over whether polls predict elections or describe electoral opinion at a specific point in time.

by non-factual questions are evaluated according to the strength of the relationship between the estimates obtained and those predicted by the theory. This lack of an objective comparison makes the evaluation of non-factual questions a more difficult and arbitrary process. The non-appearance of an expected relationship may therefore result from faulty logic or some other inadequacy in the theory as much as it may from the construction of the questions used to evaluate it. For this reason, Kalton & Schuman (1982) advocated using split-ballot tests to allow an examination of the consistency of responses to non-factual questions across versions. Although numerous researchers had employed split ballot tests prior to this, Kalton & Schuman's work reinforced the need to consider response accuracy explicitly within the question design and selection process. They also noted the desirability of external validation where this is possible, and suggested alternatives where it is not.

Although the debate over whether behavioral or attitudinal information provides a better basis from which to predict future behaviour continues (and is discussed in Chapters 2 and 4), a growing number of researchers have acknowledged that, regardless of other insights they might offer, respondents' opinions may bear little or no relationship to their current or subsequent behaviour. Bradburn & Sudman (1988) commented to this effect: *"Liking something, believing positive things about it, and acting positively toward it are not exactly the same things"* (p. 145).

Using the same line of reasoning, Labaw (1980) had earlier argued that researchers should include a battery of behavioural questions in their instruments on the grounds that respondents' past behaviour would best predict their future behaviour. Although stating that respondents often neither knew nor understood the reasons why they performed particular actions, she nevertheless advocated including attitudinal and opinion questions in order to provide a context for the behavioural questions. Others, while recognising the desirability of establishing a rapport with respondents, have cautioned that contextual questions may introduce errors (see Crespi, 1988).

Overall, researchers have long accepted that different questions elicit different types of information, vary in their predictive ability, and differ according to the ease with which they can be validated. However, many recommend employing a combination of factual and non-factual, or behavioural and opinion, questions as a means of

increasing the saliency of key topics, and thus enhancing the accuracy of the estimates obtained. More recent research attention has therefore moved away from the specific wording effects explored in the earlier literature and now focuses more on general effects, such as how questions interact with each other, and how changes in the context within which a question appears affects its response distribution. Implicit in this shift in attention is a recognition that if different words, different response options, and different types of question affect the responses obtained, so too will variations in question sequence. The following section analyses these issues and the implications they have for question designers.

3.6 Question Context Effects

Even the most rudimentary questionnaires constructed by novices are more than a random selection of questions, in an arbitrary sequence. Researchers recognise the importance of creating questionnaires that perform the same fundamental task as roadmaps. Both contain signposts, provide opportunities for respondents to rest and find their bearings, and outline clear routes along which respondents travel to arrive at, on the one hand, their destination and, on the other, conclusions and considered views on the survey topic.

The context in which questions appear, and the sequence they follow, thus play important roles in maintaining respondents' co-operation and maximising their involvement in the interview process. Sudman & Bradburn (1982) recommended that questionnaires begin with easy, salient and non-threatening questions. Such introductions, they suggested, contribute to a funnelling process which Labaw (1980) also described. Respondents begin by answering more general questions, then as the interview proceeds, the context and questions become more specific, and their content narrows until they address quite precisely the objectives of the research (Sudman & Bradburn, 1982, p. 219).

However, other researchers have found that the placement of questions, especially general and specific questions, can have a significant influence on the responses to subsequent questions (see Duncan & Schuman, 1980 or Schuman, Kalton & Ludwig, 1983 for a discussion and illustration of this). Thus as well as providing a sequence which simplifies the demands made on respondents, researchers may also have to consider how this sequence affects the responses they obtain. In other words, the

provision of a funnelling sequence of questions may also have caused question order effects and shifts in the response distributions (see Becker, 1954, for a general discussion of item order and Schwarz, 1991; Gendall, Carmichael & Hoek, under review, for discussion of a conversational-logic model proposed to explain this).

3.6.1 Question Order Effects

Sigelman (1981) provided a comprehensive definition of question order effects:

> *"A question order effect is said to operate when responses are influenced by the question placement within a survey. Most research on order effects pertains to the situation in which people's responses to one item in a survey are consciously or unconsciously influenced by their responses to a previous item; had the order of the items been reversed, or had the first one not been asked, their responses to the later question would have been different"* (p. 199).

Schuman (1992) suggested context is even broader and includes: "*more than just the influence of one setting on another. It also includes the effects of the interviewer, the interview setting, and indeed the historical setting*" (p. 13). Paradoxically, while the context in which a question appears defines its meaning, defining the context itself proves more difficult, since the number of possible influences on respondents appears vast.

In addition to increasing the saliency of a research topic, utilising a specific question sequence may also elicit responses that differ markedly from those elicited when a different context is employed. These differences may occur at many different levels, perhaps the least less complex of which is analogous to item order effects. This differs from more complex context effects, since the response pattern varies according to whether the question appears at the beginning or end of the questionnaire, rather than because of the specific questions preceding it.

Kraut, Wolfson & Rothenberg (1975) conducted research which established the existence of such an effect. Their survey used a split sample of two groups, members of which received one of two questionnaires which differed according to where a block of 46 Likert items appeared. They found that when the questions appeared later

in the questionnaire, respondents were more inclined to use modal categories, and the overall mean scores were lower than when the items were positioned earlier in the questionnaire. While the length of their questionnaire (168 items) suggests respondent fatigue may have contributed to these differences, the research nevertheless points out the risks of assuming responses will remain consistent when any aspect of a questionnaire is changed.

Schuman & Presser (1981) have conducted several studies into one of the main context effects identified to date. Their essentially chance discovery of what has come to be known as the general-specific, or part-whole, effect occurred when they noticed a major difference in the response distributions of a general question used in two studies exploring support for abortion in different circumstances (Schuman, Presser & Ludwig, 1981). Despite the researchers' attempts to balance the two samples' demographic characteristics, the differences in the response distributions persisted. Replication of the two studies revealed the same phenomenon: where a general question about abortion preceded specific questions about the circumstances under which abortion should be permissible, it elicited significantly higher levels of support than when it followed the specific questions.

Schuman *et al* (1981) described this result as a "part-whole" effect. By this they meant that the general abortion statement effectively contained the specific statements, as agreement with the general statement normally implied agreement with the specific statements, though the reverse was not necessarily true (p. 220). They suggested that where the specific item was asked first, respondents interpreted the general statement as excluding the specific situations explored in the preceding statements. That is, the context for the general statement became much narrower when it followed the specific statements. Bradburn & Mason (1964) had earlier identified and discussed a related phenomenon, redundancy. Where an overlap in responses to different questions existed, they found some respondents did not report relevant information because, having stated it in response to an earlier questions, they believed it was no longer relevant.

Since Schuman *et al*'s (1981) work, many researchers have recommended placing general questions before specific questions (see Kalton & Schuman, 1982; Converse

& Presser, 1986) although not all have accepted the theoretical basis of "part-whole" effects. McFarland (1981) explicitly tested the assumption that specific questions create a saliency likely to affect the responses to general questions. He asked respondents specific questions on religion and politics, and on the economy and energy crises, and followed these with general questions. He found significantly more interest in politics and religion when the general questions followed specific questions on the same topic but, overall, concluded that question order had little effect on the strength of the relationship between general and specific questions.

Like Schuman *et al* (1982), McFarland also noted the difficulty of determining which context provided the "truest" responses. Despite his initial questioning of the premise underlying "part-whole" effects, McFarland concluded that: *"...the common recommendation that general questions should precede questions on more specific content appears justified"* (p. 213).

He also suggested that questions varied in their susceptibility to order effects and that the more specific the question, and the more concrete the response required, the less prone to order effects the question would be (p. 214; see also Bradburn & Sudman, 1988).

However, not all context effects can be explained in terms of "part-whole" effects. Crespi & Morris (1984) reported investigating two apparently inexplicable sets of results which gave quite different response distributions, despite their use of the same question, the same time frame and the same sample location. After examining the composition of the samples and weighting to ensure these were identical, they found the differences remained. Further investigation revealed that the order of the questions varied: one version sought respondents' preferences for the state governor before eliciting preferences for senator, whereas the other questionnaire reversed the order in which the preferences were sought. Although the interactions and cross-pressures generated by this difference were too difficult to identify exactly, it was obvious that differences in the prior question had established a fundamentally different response context for the subsequent question.

Researchers examining context effects across a number of surveys have also noted the

confounding effect of variations in question wording. In other words, researchers examining several questionnaires may have difficulty in separating question wording from question context effects (see Kalton & Schuman, 1982). Although researchers have attempted to ascertain whether shifts in opinion have occurred by controlling the question context for these variables (see Borrelli, Lockerbie & Niemi, 1987; Lockerbie & Borrelli, 1990), their findings are quite general. Thus while their conclusions may aid interpretation of survey results, they are of less value in guiding the construction of a questionnaire.

Most survey researchers attempting to explore question context effects resort to using split samples. However, although any behavioural questions included allow validation against external parameters, thus offering a means of determining the context most conducive to accuracy, non-behavioural questions treated in this way prove more difficult to analyse. Over two decades ago, Noelle-Neumann (1970) argued that researchers should expose more directly the instability of the images they present in precise and quantified terms. Schuman has also argued on a number of occasions that researchers are behoven to detail and comment on any discrepancies they detect. However, many researchers apparently continue to average the estimates in the belief that this cancels out any effects that have occurred. The logic of this view is obviously flawed and Noelle-Neumann has suggested researchers will come under increasing pressure to acknowledge the limitations of the estimates they produce (see also Roper, 1983).

Other researchers (see Perreault, 1975) have advocated harnessing computer technology to produce, for each individual respondent, attractively formatted questionnaires which vary in the presentation of key question sequences. However, while researchers routinely rotate the presentation order of list items, few, if any, have adopted the suggestion to rotate groups of questions within questionnaires. This reluctance may stem from subsequent work suggesting the almost infinite permutations generated when the number of question wording and context items manipulated increases. Bradburn & Mason, (1964) earlier noted how, once researchers have ordered questionnaires to ensure they create a rapport with respondents and follow a logical structure, the length and complexity of interview schedules make opportunities for variations in question wording difficult to organise.

In summary, the preceding sections highlight the difficulty of constructing questions which elicit the desired information in a manner which is both reliable and accurate. Not only may the words themselves create ambiguity, but the question construction, the type of question employed and the context in which it appears may alter the response distribution. However, if the researcher's task appears fraught with difficulty at this point, the possibility of designing accurate, reliable, and meaningful questions appears even more remote when the errors respondents may introduce are considered.

3.7 Errors Introduced by Respondents

It is sometimes possible to gain the impression that, if only issues of question wording and construction were resolved, the difficulties afflicting survey researchers would disappear. Indeed they might, but once the questionnaires appeared in the hands of an interviewer, or in front of a respondent, new and equally perplexing problems would arise. To begin, respondents may have little knowledge of or interest in the survey topic, or they may wish to conform to particular roles they believe will enhance their standing with the interviewer or researcher, or they may, in some cases, simply struggle to understand what is asked of them. The remainder of this section examines these problems, and their effect on question responses, in greater depth.

3.7.1 Respondent Understanding

Implicit in asking people questions is an assumption that they will understand and be able, and willing, to answer those questions. Bradburn & Sudman (1988) suggested that because respondents participate voluntarily in surveys, their willingness to answer questions truthfully should not be in doubt (p. 189). However, although respondents may embark on an interview with noble intentions, their ability to fulfil those intentions may not equal their desire to do so. Collins (1982), commenting on Kalton & Schuman's paper, suggested a number of errors that respondents may introduce. He noted that to answer a question, respondents must first interpret the information they believe it seeks to elicit from them. Clearly where respondents' interpretation differs from researchers' intentions, error occurs. Second, he pointed out that even where respondents' interpretation concurs with the researcher's intention, they may find it hard to recall the information sought and some may find it hard to establish accurate time periods in both the past and the future (see Weir, 1975; Himmelweit, Jaegar & Stockdale, 1978).

The first of these problems has been discussed in terms of semantic issues in section 3.3, but variations in the vernacular employed may be only a manifestation of more fundamental differences between researchers and respondents. Labaw (1980) encapsulated the social and cultural differences that may also affect respondents' ability to interpret questions as they were intended:

> *"A major problem in the design of questionnaires and delineation of content areas occurs because questionnaires are constructed primarily by white, upper-class college graduates for administration to the general public or to important subgroups of the general public completely unlike the researchers themselves..."* (p. 39).

Her point is made also by Terris (1949); Converse & Presser (1986); Schuman (1986); and Fowler, (1994), all of whom stressed the importance of simple language and the use of respondents' own idiom, even when this appears to flout conventional rules of grammar and style. Kalton & Schuman (1982) went even further, arguing that respondents must comprehend the question itself and the concepts and frames of reference implied by it. To give a correct answer to factual questions, respondents must have the necessary information accessible to them and, if the question refers to past events, respondents must recall the information accurately.

While problems in interpretation and understanding can be treated in the same way as ambiguity - by pre-testing - the results of comprehensive studies in this area serve only to reinforce the scale of the problem. Belsen's (1986) work using his "double-back" technique has suggested well over half the respondents thus questioned had misinterpreted the question. Nor are his findings isolated experiences. Gendall (1994) has also reported obtaining a wide range of interpretations of apparently straightforward questions. Thus while pre-testing helps to identify the presence and size of respondent misunderstanding effects, there are few rules enabling researchers to identify the potential for these effects to occur. Even where respondents correctly understand the question put to them, their interest in the topic, and its general relevance to them, may lead to the emergence of a new set of biases.

3.7.2 Respondent Knowledge

If respondents lack knowledge about the research topic, or believe it is of little relevance to them, they may give less thoughtful and considered answers. Bogart (1967) suggested that the proportion of truly reflective answers given in surveys was so low that researchers should instead measure: *"the degrees of apathy, indecision or conflict on the part of the great majority, with the opinionated as the residual left-over"* (p. 337).

Yet until researchers adopt Bogart's recommendation they will still have to deal with at least two problems which emerge when a topic lacks relevance to respondents, or when they are largely ignorant of it. First, respondents may express an opinion, even when they do not hold one, possibly because they believe the interviewer expects it of them, or because they worry that not to do so would make them appear uninformed, and less like a "good citizen". Second, they may lose interest in the survey and, at best, declare themselves undecided and, at worst, terminate the interview.

The issue of social desirability bias, and its role in contributing to this behaviour, is dealt with in detail below, (the problem of response alternatives, and whether to include a specific undecided category, was examined in section 3.4.1). However, the relationship between topic saliency and respondent knowledge falls somewhere between these latter two concerns and is discussed in detail below.

Many researchers have examined the question of respondent knowledge in terms of the level of undecided responses received. However, even undecided responses may mean different things. Kalton & Schuman (1982) suggested that undecided responses differ across factual and non-factual questions. That is, a "don't know" response to a factual question indicates that a respondent has not been able to proffer a required piece of information. In other words, an answer to the question exists, but, for whatever reason, the respondent could not articulate it. With opinion questions, however, an answer may not even exist, and "don't know" responses consequently reflect a different situation.

Ferber (1956) had earlier explored the issue of respondent ignorance in detail and

concluded that at least three specific categories of respondent existed. First, he described informed respondents, who both offered an opinion and subsequently demonstrated some knowledge of the topic. Second, he commented on undecided respondents, who declined to offer an opinion and third, he identified misinformed respondents, who offered an opinion on the topic, but whose actual knowledge of the topic was flawed.

Ferber explored the views of 407 people who had expressed opinions on particular topics, in order to gauge their actual knowledge of the survey topic. He found fewer than half the sample had a sound understanding of any of the four topics examined. Perhaps surprisingly, no real differences in the willingness of misinformed or informed respondents to offer opinions existed, although sharp differences in the views they expressed were evident. Ferber suggested that the high frequency of neutral opinions found in many polls may reflect a high level of ignorance of the topic and a tendency on the part of uninformed respondents to take a neutral position (p. 596). To help deal with this, he recommended researchers continue to solicit respondents' opinions, but that they supplement these questions with others designed to test the level of knowledge underlying respondents' opinions. However, he did not suggest how responses from misinformed respondents should be treated.

Respondent ignorance is not the only reason why respondents might give neutral or undecided responses. In some cases, they may simply have difficulty in recalling their experiences, or in defining the time period referred to in the question. Several researchers have explored this issue, suggesting the use of bounded recall to reduce telescoping (see Sudman & Bradburn, 1974; Kalton & Schuman, 1982) and recommending that researchers devise a more comprehensive context for the questions of interest. Other researchers have asked whether these procedures are sufficient and whether differing treatments should be applied to responses given by different types of respondent. Bogart (1967), cited earlier, had suggested that all responses received to a survey may not be equally valid but stopped short of suggesting that those deemed less valid be discounted, or even excluded, from the analyses.

Until a widely accepted procedure for differentiating between responses on the basis of their quality exists, researchers typically choose at least one of the following options. The first, noted earlier, involves acknowledging and discussing more fully

any limitations affecting the data set and perhaps investigating the response distributions of different sub-groups. The second, which has received widespread attention, involves addressing the issue of saliency in an attempt to reduce the problem.

3.7.3 Topic Saliency

As noted in section 3.4, several researchers have commented on how providing a context for the questions of interest may help to increase the saliency of the topic, thereby making information relating to it more accessible. The remainder of this section examines some of the suggestions put forward, and their implications, in detail.

Many researchers (see Kalton *et al*, 1978; Sudman & Bradburn, 1982; Converse & Presser, 1986; Bishop, 1990; Martin & Harlow, 1992) have noted that people may be more knowledgeable about and more interested in a topic when they have had an opportunity to think a little about it, and to reflect on any experiences they may have had that relate to that topic. Sharp & Frankel's (1987) "respondent burden" index also explicitly recognised the importance of the survey topic and respondents' view of its importance.

However, not all introductory questions have the same effect on respondents' willingness to answer questions. Martin & Harlow (1992) examined the use of filter questions and their effects on respondents' subsequent answers. Their results suggested that respondents who could not answer a filter question, had a lower level of interest in the survey topic than they had indicated prior to being asked the filter question. While Martin & Harlow's interest lay in discussing the cognitive processes underlying this pattern, it seems logical that topic saliency will increase when respondents feel they are able to contribute in a meaningful way to the discussion. In turn, this implies that filter questions should provide opportunities for respondents to display their knowledge rather than to reveal their ignorance (see also Bishop, Oldendick & Tuchfarber, 1980a). However, researchers have not yet explored whether the use of these questions increases social desirability bias.

Assuming respondents do not find the tenor of introductory questions demeaning,

providing a context for questions central to the research objectives should improve the accessibility of information required to answer those questions. Sudman & Bradburn (1982) summarised the notion of a research context:

> *"... the order of questions provides a context within which questions are answered. Questions that are quite closely related, particularly those that are related to the same attitude object, may increase the saliency of particular aspects of the opinion or provide a further definition of a particular aspect"* (p. 142).

Blakenship (1943, cited Kalton *et al, 1978*) noted higher levels of undecided respondents when the key questions appeared at the beginning of the questionnaire, and Sigelman (1981), who also tested this notion, arrived at a similar conclusion. Sigelman used two questionnaires to explore whether respondents' willingness to offer an opinion varied according to when they were asked to provide it. The first questionnaire version asked respondents their view on the President's performance after only three essentially demographic questions, while the other version preceded these questions with a range of related opinion questions. Sigelman found no differences in the approval ratings themselves, but significant differences in the level of opinionation. More educated respondents showed no change in the level of undecided responses given, but less educated respondents were much more willing to offer an opinion if administered the "contextual" version.

Strack (1992) explained this effect by suggesting that providing a context was analogous to priming respondents. In other words, having to activate information to answer an earlier question makes that information more accessible to subsequent questions, and so increases the proportion of respondents offering an opinion. Schwarz (1990) also noted that the response alternatives provided to respondents may help them access information which they then use to provide more detailed responses to subsequent questions.

However, while measures exist to assist respondents to recall specific information, they may also declare themselves undecided because they are genuinely unable to choose between two or more conflicting views. Some respondents also offer undecided response because a question may threaten them in some way and, rather

than confronting that threat, they choose to evade it. Little attention appears to have been given to the notion of multiple opinions, although Bogart (1967) suggested the model that people hold a single opinion had serious flaws:

> *"This model has the virtue of great simplicity, but it makes no sense, because conflicting and contradictory opinions may be held simultaneously and because they constantly jostle each other for dominance"* (p. 344).

Actually proving that people hold incongruent opinions may be difficult, given that respondents often make efforts to appear consistent. Accordingly, it is perhaps not surprising that research attention has focused more on the question of social desirability bias, which is easier to detect and document using empirical measures.

3.7.4 Threatening Questions and Social Desirability Bias

Coombs & Coombs (1976) suggested that an undecided response: *"... may reflect something other than just a difficult discrimination or an indeterminate category on an ordinal response scale"* (p. 497). They argued "don't know" responses may be better treated as meaning "don't want to know" for a host of reasons. Among other ideas, they surmised the question itself may be too sensitive, making respondents either reluctant to reveal their own views, or worried that, in making a response, they may also reveal their own ignorance.

Non-response to threatening questions and social desirability bias thus appear distinct, but closely related phenomena. Sudman & Bradburn (1974) outlined the relationship between the two problems:

> *"Some questions call for the respondents to provide information on topics that have highly desirable answers.... if a respondent has... engaged in a socially undesirable behaviour, he may face a conflict between a desire to conform to the definition of good respondent behaviour, which says that one should tell the truth, and a desire to appear to the interviewer to be in the socially desirable category. It is frequently assumed that most respondents resolve this conflict in favour of biasing their answer in the direction of social desirability"* (pp. 9-10).

Unfortunately, unless researchers have some external validation for each respondent, they have no way of establishing in which direction a respondent ultimately resolves the conflict. Given that Bradburn & Sudman (1979) noted discrepancies between respondents' views on how threatening they found particular topics and how threatening they believed others would find the same topics, the potential bias generated by efforts to appear knowledgeable or socially well rounded appears to be great.

Parry & Crossley (1950) had reached similar conclusions nearly three decades earlier. More concerned with measuring accuracy, they attempted to assess the aggregate and individual accuracy of specific behaviours, which could be externally validated, and which all carried a certain degree of prestige. Over the six behaviours investigated, only a third of those questioned provided entirely correct answers. Parry & Crossley found that inaccuracy typically followed social pressures, and was especially pronounced in over-reporting of voting (see also Cahalan, 1968; Katosh & Traugott, 1981).

These examples deal with respondents' attempts to conform to socially desirable patterns, but as Bradburn, Sudman, Blair & Stocking (1978) pointed out, respondents may also wish to present themselves as the sort of people who would not participate in or condone socially undesirable behaviours. Thus social desirability biases work in two directions: the over-reporting of commendable behaviour and the under-reporting of less widely sanctioned behaviour.

Given that researchers have long agreed that these effects exist, they have, in recent years, devoted less attention to documenting the bias and more to attempting to minimise the size of any bias affecting their results. Establishing a strong rapport with respondents represents an obvious means of reducing respondents' perception that interviewers "judge" the answers they provide. Smith (1992) suggested that a rapport leads not only to a more trusting and open exchange of information, but also to more accurate estimates as respondents become more accustomed to the scales used and the tasks required of them (see also Sudman & Bradburn, 1974; Labaw, 1980; Sigelman, 1981). However, he also warned that a rapport could also introduce bias. Smith suggested that where questions about a subject stimulated respondents to report more behaviour, or interest, related to the topic, they may exaggerate their experiences in

order to fulfil more completely the role they see themselves playing.

According to Bradburn & Mason (1964), effects arising from an almost excessive rapport can be explained in two ways. First, having created what they believe is a good impression, respondents answer subsequent questions so they are consistent with their earlier judgments. Alternatively, they suggested that as a questionnaire proceeds, respondents may become fatigued, leading them to give more perfunctory answers (see also Herzog & Bachman, 1981, who discuss evidence of straightline responding). Regardless of which explanation is actually true (or whether truth in such issues can be ascertained), creating a rapport has the potential to introduce as well as reduce bias.

Other researchers have suggested using longer questions, especially when eliciting behavioural information. The rationale for this is that when respondents receive more information about the task required of them, the question itself may contain memory cues which effectively assist their recall (Laurent, 1972; Sudman & Bradburn, 1982). However, while this technique may increase the proportion of people reporting socially undesirable behaviours accurately, it may also increase the numbers reporting more wholesome behaviours.

Sudman & Bradburn (1982) suggested overcoming this latter problem by including in the introduction to the question ideas that might increase the acceptability of giving a less socially desirable response. For example, they preceded questions seeking details of respondents' past voting behaviour with a statement which acknowledged that unexpected events may sometimes prevent people from voting. Kalton & Schuman (1982) also argued that by effectively desensitising the desirability of voting, respondents might feel more able to volunteer details of their actual behaviour.

Presser (1990) also considered means of reducing respondents' impulse to over-report certain behaviours. He tested whether it was possible to mute the unfavourable connotations associated with non-voting by giving respondents other opportunities to display their good citizenship. Using a split ballot sample he asked half his respondents whether they knew and could name their nearest polling station before asking whether they voted; the other half answered a question exploring their voting behaviour only. Presser surmised that people who could not name their poll location would be less likely to claim that they voted. His results supported this, although a

quarter of those unable to name the poll location still claimed to have voted. In addition, the results were somewhat confounded by the fact that many non-voters did know the poll location, leading Presser to conclude that misreporters may not have been put in a position of admitting something inconsistent with voting (as he had originally hoped).

In a second test of contextual effects, Presser provided explicit opportunities for half his respondents to indicate the regularity with which they had voted over several years before asking them specifically about their voting behaviour in the election just held. However, his exploration of whether respondents who had an opportunity to demonstrate their good citizenship would be less likely to misreport their most recent behaviour, led him to conclude that the results did not support his hypothesis. Although difficult to establish precisely why the surmised results did not appear, Presser suggested respondents, having established an impression of "good" behaviour, may be unwilling to sully this by admitting to non-voting. This fits well with Bradburn & Mason's (1964) notion of consistency, but is difficult to verify empirically.

Other tests of how context might diminish the perceived social desirability bias of certain responses include the use of conditioning attitude statements and embedded questions (see Sudman & Bradburn, 1982). Gendall, Hoek & Blakeley (1992) explicitly tested the role of attitude statements in reducing the undesirability of admitting to drink-driving offences. Four questionnaires, employing positive statements downplaying the offence, negative statements highlighting the offence, a list of traffic offences in which the behaviour was embedded, and a direct question were employed. The version which preceded the behavioural question with positive attitude statements obtained the highest admission rates, and the direct question the lowest. However, varying response rates across versions raised the possibility that some versions encouraged response behaviour better than others, and made the results more difficult to interpret.

Sudman & Bradburn (1982) cautiously recommended the use of embedded questions, but noted that they added to the length of the questionnaire and thus increased the costs of the research. Researchers, they suggested, should attempt to balance the increases in accuracy likely to result against the costs incurred.

Since social desirability bias arises from a desire in respondents to portray themselves in a certain manner, the problem should diminish if the key audience, the interviewer, becomes less prominent in the research process. Studies using anonymous response mechanisms thus appear likely to reduce social desirability bias. Researchers have explored whether a less personal survey medium (phone versus face-to-face) reduces over-reporting, but found no significant improvement (Presser, 1990).

Others have examined randomised response models, where respondents answer either a question prone to social desirability bias, or a quite innocuous one, while the interviewer remains ignorant of the actual questions answered. Kalton & Schuman (1982) stated randomised response techniques reduced over-reporting, but led to a large increase in sampling error (see also Presser, 1990). Sudman & Bradburn (1982) went further, arguing that the technique led to even higher levels of over-reporting of socially desirable behaviours. Presumably this increase in over-reporting occurred as respondents did not have to directly misreport their behaviour and thus avoided any cognitive dissonance this might have produced. Nor did they have to face the real or imagined gaze of a possibly sceptical interviewer. An additional disadvantage noted by Kalton & Schuman (1982) was that the results cannot be compared to the responses given to other variables. That is, since researchers have no way of knowing which questions each specific respondent answered, they cannot compare the response distributions for the sensitive question administered in this way to, for example, demographic variables. In effect, this means they have no means of establishing which respondents, if any, the treatment affected most.

Singer explored other confidentiality mechanisms in detail (Singer 1978a; 1978b; Singer, Hippler & Schwarz, 1992; Singer, von Thurn & Miller, 1995). Her work has concentrated on the confidentiality assurances provided to respondents, in particular, the level of detail associated with these, and their effects on response rates. Among other conclusions, she noted that excessively long confidentiality assurances may arouse respondents' suspicions (Singer *et al,* 1992). More recently, Singer *et al* (1995) reported that although confidentiality assurances do improve response where the data sought are sensitive, their effect on less sensitive data is negligible.[15]

---

15 The same effect was found whether "response" was measured as item non-response, response rate, or response quality.

Perhaps the greatest level of confidentiality is offered by a secret ballot or anonymity. Pearlin (1961) noted that anonymity was believed to "*preclude the possibility that as a result of his opinions a respondent will be the object of derision, distrust, or even revenge*" (p. 641). However, he found that respondents' desire for anonymity depended more on their personality than on the actual survey topic. Wildman (1972) queried whether anonymity had any effect and concluded that anonymity failed to produce effects that could not reasonably be ascribed to chance (see also Goldfield, Turner, Cowan & Scott, 1978, for a discussion of confidentiality and privacy).

Instead of employing complete anonymity, therefore, many researchers have increased the level of confidentiality associated with survey responses. Sudman & Bradburn (1982) explored the contexts in which researchers used secret ballots and found these were typically used to increase the reporting of socially undesirable behaviours. They argued that secret ballots could be more effectively used to reduce social desirability bias, a view which seems at odds with their conclusion about randomised response models, which they suggested increased over-reporting. One possible explanation of these apparently discrepant views might lie in the relationship between a secret ballot interview and the process of casting a vote. Since the behaviours performed in the interview and in the polling booth are so similar, it is possible that the nature of the task itself conditions respondents to provide responses which more accurately predict their likely behaviour.

More recently, Sparrow (1992) has reported that a group administered a secret ballot in a split-sample experiment showed a markedly different voting intention distribution compared to the group which responded to a direct question. However, he could find no clear relationship between party support and over- (or under-) reporting and rather weakly concluded that the relationship was dependent on the parties' images at the time of interviewing (see also Benson, 1941).

Overall, while researchers have a number of mechanisms available to them which might reduce respondent biases, these do not provide unequivocal solutions and, in some instances, have the potential to introduce at least as much bias as they remove. The next section summarises the sometimes contradictory evidence attesting to certain question design or respondent bias effects.

3.8 Research Rationale

Eliciting details of respondents' voting behaviour appears likely to encroach on some respondents' feelings of privacy. The same questions may prompt other respondents to give answers designed to satisfy perceived social norms. Given the wide ranging errors and biases voting intention questions may introduce, these issues appear to merit more detailed research attention.

The specific ideas tested in this research are twofold. First, whether the provision of a context for the questions eliciting likely voting behaviour reduces the proportion of undecided respondents, thereby increasing the accuracy of the estimates. Second, whether reducing social desirability pressures and privacy concerns through the introduction of confidential (not anonymous) response mechanisms results in a decrease in the size of the undecided group and an increase in the accuracy of the estimates.

The literature detailing funnelling sequences which lead respondents through general and then specific questions is well documented. Less well documented is the effect these sequences have on two key aspects of respondents' behaviour: their willingness to offer a response to the questions of interest, and the accuracy of the responses offered. Similarly, the increase in confidentiality provided by a secret ballot may well decrease the proportion of undecided responses, and the pressure to conform to certain real or imaginary role expectations, but unless it increases the accuracy of the overall estimates, it holds no real value for researchers.

This research examined both topics in an attempt to further the knowledge of question context effects and social desirability bias. However, a further attempt to deal explicitly with many of the questions and problems discussed in this chapter involved the development, refinement, and testing of a new method of predicting likely voting behaviour. Chapter 4 introduces and describes this second key aspect of the research.

3.9 Summary

Researchers have acknowledged that question design effects have the potential to change the response distributions they obtain. However, while few would dispute this, most have difficulty in predicting when these effects will occur, the direction in which error will occur, and the topics and questions they will affect most.

Similarly, a high level of agreement about the existence of various errors introduced by respondents also prevails, but researchers still remain uncertain about how best to predict the levels of bias likely to occur in a given survey.

Overall, while the effects themselves have been painstakingly documented, the means of reducing or addressing these remain open to debate and researchers have yet to develop rules of design that accommodate and neutralise the errors discussed in this chapter. However, these problems do not render the task of question wording and questionnaire design impossibly difficult. Instead, they suggest that researchers should continue to place increasing emphasis on behavioural questions rather than those that elicit opinions, since the problems discussed affect the latter questions more severely than they do the former. While general theories of question wording remain tantalisingly elusive, researchers have at least some clear avenues down which future research efforts might proceed.

The final section outlined the specific issues explored further in this thesis and set out an explanation and justification for their inclusion.

CHAPTER 4

SUBJECTIVE PROBABILITIES

4.1 Introduction

This chapter begins by briefly summarising the debate over which models should be used as the basis of predictions about consumers' behaviour. In turn, this provides a context which helps identify the differences between questions which elicit statements of intention and those which elicit probabilities about a respondent performing a particular behaviour, or set of behaviours. The chapter then analyses research which has explored these differences and studies outlining instruments designed to elicit subjective probabilities. Overall, the studies in this area conclude that intention and probability questions differ in their predictive ability. In particular, one scale, the Juster Scale, which appears to have provided accurate estimates of a variety of different behaviours, is described. Results from studies using the Juster Scale, as well as the advantages and limitations associated with this scale are then discussed. The chapter concludes by presenting a rationale for utilising the Juster Scale to predict voting behaviour.

4.2 The Intention - Behaviour Controversy

For many years academics from a variety of disciplines have debated the best methods of eliciting information from respondents about their future behaviour. Criteria for assessing superiority have included the internal logic of the models, their reliability, their validity (or predictive ability), and their parsimony. At least two quite different theories exist.

The first of these appears to stem directly from work into consumer behaviour which has dominated much of the marketing and survey research literature for the last five decades. While many models exist within this school, they are based on the common belief that respondents' behaviour can be predicted by their behavioural intentions, attitudes and beliefs, all of which form a causal sequence. Warshaw (1980) summarised the view: "*The specific nature of this relationship evolves from the Fishbein extended model (e.g., Ajzen and Fishbein, 1973), which assumes behavioural intention (BI) approximates behaviour (B), where BI is the subjective probability of*

performing an overt behaviour. That is, with little variation, people do what they say they are going to do" (p. 26).

Although Warshaw has equated behavioural intention with subjective probabilities, Fishbein's work, and subsequent work based on his findings, has tended to use scales which measure intentions rather than probabilities. This difference between intentions and probabilities is critical, and section 4.3 elucidates it in more detail.

In the context of survey research practice, and polling research more specifically, intention models have had a strong influence on the questions employed and the data sought. Typically, researchers seek information about respondents' intentions, measured dichotomously (will or will not perform an action; see Klein & Lansing, 1955; Tobin, 1959; Heald, 1970), or on a 5, 7 or even 9-point scale (indicating gradations of intention), which they assume will predict respondents' subsequent behaviour.

However, a variety of empirical tests attempting to confirm the existence of the link between intentions and actual behaviour have, at best, met with variable success. Over the last two decades, marketing and research literature has seen an increase in the number of articles documenting the weakness of the intention-behaviour link and positing an alternative approach (see Foxall, 1982; Driver & Foxall, 1986). This approach has sought to promote the use of more direct behavioural measures; these include explicit subjective probabilities, which, proponents have claimed, provide more accurate estimates of respondents' future behaviour. Instead of assessing respondents' intentions, researchers employing this approach measure the likelihood of their subjects performing specific behaviours within defined periods. The ensuing sections of this chapter discuss the problems associated with intention measurements, the development of probability instruments, the empirical evidence asserting their superiority, and the questions which still require attention.

4.3 Measurement of Intention

Regardless of how they elicited respondents' intentions, several researchers have noted that many people who declare they have no intention of performing a behaviour in fact go on to do so (Juster, 1960; Juster, 1966; Clawson, 1971; Pickering &

Isherwood, 1974; Warshaw, 1980; Gan, Esslemont & Gendall, 1985).[16] Clawson (1971) suggested that even if only a small proportion of non-intenders do act, they typically constitute such a large proportion of those who perform the behaviour that the error generated is very large, and the projected estimates are quite inaccurate.

Researchers have suggested a variety of reasons to explain how and why these discrepant results occur. In particular, the intervention of unanticipated events has received detailed attention and, to counter any influences these might have, Fishbein & Ajzen (1975) advocated measuring behavioural intention as close as possible to the performance of the behaviour in question. Minimising the timelag between the intention and behaviour effectively reduces the potential for unexpected occurrences to affect the estimates (see also Clawson, 1971; Pickering & Isherwood, 1974; Warshaw; 1980). However, Manski (1990) suggested that the unavailability of key information at the time of the prediction was likely to remain a major problem: *"Divergencies* [between intentions and behaviour] *will occur even if responses to intentions questions are the best predictions possible given the available information. The lesson is that researchers should not expect too much from intentions data."* (p. 940).

Gross & Niman (1975) expanded the range of factors contributing to erroneous predictions by listing personal, situational and methodological issues which may interfere with the chain of causation from attitudes, through intentions to behaviour. Seeking to identify the prime movers among these, Pickering (1975) explored why respondents' intentions differed from their behaviour. Based around purchase (or non-purchase) of consumer durables, his research enabled him to document a range of unforseen events such as the need to replace the item, special promotions, or personal factors, such as improved finances or family pressure, which prompted unintended purchases. He suggested failure to make an anticipated purchase resulted mainly from changes in the respondent's financial position. While based only on a small sample, these responses provided some support for Pickering's hypothesis (discussed in section 4.3) that consumers' willingness and ability to purchase merited closer scrutiny (see also Isherwood & Pickering, 1975). Overall, these studies confirmed the difficulty of making even short term forecasts where personal or situational factors may intervene.

---

16 Likewise, many people who state an intention to perform a behaviour often do not.

Other researchers concentrated less on external influences and instead explored the methodological problems affecting intention scales. In particular, they have queried the accuracy and uniformity of information collected from respondents. The problems presented by dichotomous scales were quickly identified, in particular, the tendency of these scales to oversimplify the choice posed and greatly constrain respondents' ability to report their intentions (see Ferber & Piskie, 1965).

In addition, the problem of whether respondents attach the same meaning to the various points on an ordinal scale has been well documented (see Ferber & Piskie, 1965; Worcester & Burns, 1975; Ptacek & Ross, 1979). As these scales are currently used, researchers have no way of comparing or standardising respondents' preference intensities, although they implicitly assume that respondents attach the same denotation to ordinal scale values. Researchers using these scales must also devise some way of calculating the actual purchase rate from the proportion of people in each intention category. Given the difficulty of making the scale points more concrete and universally understood, this apparently arbitrary demarcation process appears unlikely to lead to accurate estimates.

Beniger (1979), among others, has questioned the assumption that scale points denote the same meaning to different people, calling it the "horizontal fallacy" of ordinal scales. He argued that attitudes: *"are preferences in comparison to other alternatives and, as such, are not absolute and should not be subject to interval measurement."* (p. 270). Inaccuracies in prediction are thus attributable to flaws in the scales themselves, as well as to the way in which researchers use the data they educe.

While Beniger's solution to this problem was to use a rank-distance approach, studies comparing attitude and intention scales have differed over the level of inter-scale reliability that exists. Kassarjian & Nakanishi's (1967) comparison of open and limited choice scales revealed high inter-method correlations, which they suggested may be evidence "*of convergent validity for each technique*" (p. 153). However, Axelrod (1968) reached a quite different conclusion. His test of the sensitivity, stability and predictive power of sixteen different scales used to predict brand purchase behaviour suggested not all possessed a high level of sensitivity, though those that did, displayed a high degree of stability. Since the predictive power of the

scales varied, Axelrod recommended using a probability scale administered via a constant sum instrument as the best means of differentiating between those who will, and those who will not, perform a particular behaviour. Axelrod identified the key advantage of constant sum methods: *"...they produce a continuum of purchase probabilities along which consumers are distributed. This continuum provides a base for analytic work that is not provided by a go/no go type of equation."* (p. 8).

Rather than investigating the relationship between attitudes, intentions and behaviour, Haley & Case (1979) examined the convergent validity of a variety of scales. Testing 13 different scales, they sought to assess whether scales measured the same constructs, and whether some were better discriminators than others. Their comprehensive analysis identified several issues that have contributed to the refinement and development of many scales. In particular, they noted respondents' tendency to select points associated with a label, especially if that label was verbal (as opposed to numeric). Despite this, when comparing the average ratings across scales, they found a great deal of overall agreement. That is, the scales appeared to possess convergent validity.

Subsequent analysis to investigate whether the scales measured the same constructs suggested they had two key dimensions: an evaluative dimension and a salience dimension. Here the scales did vary, and Haley & Case concluded scales loading heavily on the saliency factor were better discriminators and showed a greater level of consistency with behaviour. Differing from some of Axelrod's recommendations, Haley & Case advocated use of awareness, purchase intention and paired comparison scales to predict brand choice behaviour. However, Haley & Case also explicitly acknowledged problems in the presentation of some scales that may have contributed to those scales' poorer performance: *"This tendency* [to opt for points accompanied by a verbal description] *apparently resulted in under-use of the numerical points of our purchase probability scale and may have reduced its ability to discriminate."* (p.31).

This problem suggests some scales were advantaged, and raises the possibility that the results presented might have been quite different had a more consistent presentation format been adopted.

Aside from problems in how scales measured future behaviour, researchers identified other factors that could contribute to inconsistencies in intention-behaviour measures. Warshaw (1980) suggested a variety of possibilities, including the existence of peripheral behaviours which had poorly developed intentions that were not always accessible, or, if available, able to be assessed accurately. This lack of centrality, he argued, meant the posited relationships were obscure and so could not reasonably be expected to be elicited accurately.

Leading on from this, Warshaw suggested that purchase behaviour could be more stochastic (i.e., probabilistic) than deterministic (i.e., a function of intentions) (p. 27). Assessing the probability or likelihood that respondents will undertake certain actions thus allowed an assessment of factors other than intention and, in bringing a wider context into focus, could, he argued, lead to more accurate estimates. However, Miniard, Obermiller & Page (1983) criticised Warshaw for confounding contextual and format issues. After separating and testing these variables, they concluded that contextual specificity did not enhance an intentions-behaviour relationship.

Warshaw's suggestion, and Miniard's *et al*'s reaction to this, marks the bifurcated direction research has taken. On the one hand, an active group of researchers has continued developing deterministic or intentions models. Their work, following the direction established by Fishbein and Ajzen, has sought both to address and overcome the problems noted earlier. These efforts have met with some success (see Sheppard, Hartwick & Warshaw, 1988, for a recent review of this work) but the conditions required for this success are typically so tightly controlled that they raise serious doubts about the possibility of replicating the results outside an experimental context. Driver & Foxall (1986) encapsulated the difficulty: *"...high levels of correlation can be obtained only under the tightest of circumstances, notably when the elapsed time between the expression of an intention and the occurrence of an opportunity to act in accordance with it is minimal."* (p. 154).

While these criticisms discuss the problem of transferring the research from the laboratory to the social or commercial world, others have raised more fundamental questions about the relationship of attitudes and intentions to subsequent behaviour. Working in the area of new product development, Tauber (1981) explicitly questioned

the relationship that had underpinned concept testing: *"The implicit assumption of concept testing and product testing is that attitudes and intentions expressed during a pre-market situation will relate to later adoption behaviour. This assumption is fallacious."* (p.176).

In an explicit test of Fishbein's model, Fredericks & Dossett (1983) suggested respondents' previous behaviour was a more accurate indicator of their future actions: *"In essence, once prior behaviour is included in the model, the effects of intention become nonsignificant."* (1983, p.509). However, Fredericks & Dossett's work investigated behaviour relating to fast-moving consumer goods, and their respondents would have purchased items on a regular basis. Other purchase behaviours, such as those relating to services, or other behaviours altogether, may not have such a history, thus making the role of past behaviour more difficult to establish and less easily incorporated into predictive models.

One response to these criticisms has been to replace attitude and intention variables with estimates of future behaviour, and the second research path has proceeded in this direction. Nearly twenty years earlier Juster, among others, had considered similar issues to those Warshaw (1980) summarised, and which Tauber (1981) and Fredericks & Dossett (1983) explored in another context. His work, as well as that of those who followed on from him, also charts the movement away from attitudes and intentions to probabilities.

4.4 Development of Subjective Probabilities

Researchers began addressing the problems associated with intention scales in at least two general ways. One school, following an economic model, investigated a wide range of consumers' socio-economic characteristics and attempted to use differences in these to predict whether respondents would purchase particular products. Thus respondents' purchase expectations, or buying intentions, were not explicitly sought and researchers initially placed greater emphasis on demographic and socio-economic characteristics (see Juster & Wachtel's 1971 description of a stock adjustment model, for example).

The second school concentrated specifically on distinguishing between intentional and

probability data, and on exploring whether the latter offered any advantages over the former. Neither approach specifically excluded the other, indeed both contained at least some elements of the other, nevertheless, their objectives and experiments differed, although the conclusions they arrived at shared some common characteristics. The following sections examine these two lines of research before exploring the development of subjective probabilities in more detail.

4.4.1 Characteristics of Buyers and Non-Buyers

Klein & Lansing (1955) were among the first to use the average characteristics shared by people as the basis of predictions about the proportion of buyers in a population. In an initial interview they collected data on variables ranging from respondents' liquid assets to their buying intentions. After twelve months had elapsed, all those who had not moved address during the intervening period were re-interviewed so Klein & Lansing could assess the correlation between the original variables and respondents' subsequent behaviour. Despite some evidence of a relationship between respondents' purchase plans and their actions, Klein & Lansing concluded that these did not discriminate adequately between buyers and non-buyers. Although this inadequacy, may result in part from Klein & Lansing's decision to exclude those who had shifted residence (and whose behaviour had already differed in an important respect from those subsequently re-interviewed), they could not assess the extent to which these respondents might have affected their estimates.

The questions they posed:

> *"Do you expect to buy a car this year, in 1952?"*

and

> *"Do you expect to buy any large items such as furniture, a refrigerator, radio, television set, household appliances, and so on - during this year, 1952?"* (p. 110).

also present some problems. These questions imply a simple dichotomous choice (with the possibility of a don't know response), and researchers have documented the poor discriminating power of this type of question (see Ferber & Piskie, 1965). Consequently, the question used also appears likely to have contributed to these inaccuracies because of its explicit assumption that intentions take one of two forms: wholly positive or wholly negative.

Identifying buyers and non-buyers, and isolating factors which affected respondents' willingness and ability to buy also constituted key objectives of Isherwood & Pickering's (1975) and Pickering & Isherwood's (1975) work. They hypothesised that the purchase of cars and household durables depended on consumers' attitudes and expectations, as well as on economic variables. Thus while their research continued to explore the role of socio-economic characteristics, it also encompassed factors more closely related to probabilities. Overall, however, they sought to distinguish between buyers and non-buyers, and concluded that respondents' general economic confidence, and confidence about their own ability to bear the costs of purchasing given items, were more important differentiating factors than their actual intentions. Furthermore, Pickering & Isherwood argued that measures of consumer confidence and respondents' willingness to make purchases had an important predictive value, and, in the discriminant models they developed, these variables made a contribution independently of socio-economic information.

4.4.2 Explanatory Power of Probabilities

This emphasis on variables other than intentions or respondent characteristics had also interested Juster, as he struggled to explain why intention questions often failed to obtain accurate predictions of respondents' behaviour. His work showed clearly that most respondents reporting buying intentions did not fulfil these in the time period nominated; moreover, most purchases were made by householders who had reported no intention of buying (Juster, 1960; 1966; Theil & Kosobud, 1968). Juster pointed out that, if departures from intentions were random, the aggregate estimates would remain accurate. Since this was not the case, he attempted to ascertain the direction and extent of departures from stated intentions so these could be modelled, and the predictions adjusted accordingly.

To do this, Juster developed two hypotheses about the way in which intentions, purchase behaviour, and attitudes were related. The first of these, the "additivity" hypothesis, suggested that intentions were essentially attitudes with a time dimension, thus both attitude and intention measures could help gauge respondents' state of mind when they were interviewed. The second hypothesis, the "contingent action" hypothesis, differentiated sharply between intentions and attitudes. Unlike attitudes, intentions were posited as forward looking variables which expressed, with varying degrees of certainty, subjective judgments about the future. Juster (1960) summarised

this interpretation of intention, suggesting it constituted: *"judgments about what an individual or a household expect to do, provided that expectations about what others will do are correct."* (pp. 611-612).

In testing these hypotheses, Juster found no conclusive evidence either to support or dismiss the additivity hypothesis, but found strong evidence in favour of the contingent action hypothesis. When exploring this further by regressing respondents' income expectations against their purchases, (for respondents with buying intentions of 0 or 1), Juster found that some respondents' income expectation was more closely related to their "real" level of buying intention, when their stated intention was zero (1960, p. 615). This led him to hypothesise that zero might represent a number of intentions, ranging from definitely won't buy through to various levels of uncertainty, and provided him with a further impetus to develop a scale which could record levels of uncertainty more precisely.

Juster's work continued to develop in this area and over the next five years he differentiated between intentions and probabilities, and amassed the evidence that lead him to conclude that probability instruments offered researchers a superior method of predicting future conditional behaviour (Juster, 1966). His work also drew attention to the large proportion of non-intenders (people stating they did not intend to buy) who registered greater than zero probabilities of performing the specified behaviour. While these findings suggested that intentions and probabilities represent different information, more solid evidence for this view came after he regressed both intentions and purchase probabilities against respondents' behaviour. Although intentions had a significant relationship with behaviour in the early equations, this was before the purchase probability variable was entered. Once this appeared in the equation, it showed a highly significant relationship with behaviour (both before and after the intention scores were entered), whereas the relationship between intention variables and purchase ceased to be significant. Overall, Juster concluded that these results supported his contention that respondents' purchase probabilities bore a closer relationship to their behaviour than did intentions data.

While Juster concluded that these results lent further support to the claim that purchase probabilities possessed superior predictive ability, he did not offer them as

a panacea to the difficulties that had beset earlier researchers who based their predictions on intentions data. Instead, he explicitly acknowledged that mean probability scores tended to underestimate actual purchases rates, but suggested that, notwithstanding this downward bias, they still provided better predictions than did either intentions or weighted averages of intentions.

Seeking to explain why intentions should prove such poor predictors of behaviour, Juster surmised that respondents were being asked to assign a probability to at least two possible outcomes when most had no specifically formed intentions. According to this view a "yes" meant that, after some consideration of the factors affecting the purchase decision in question, a "yes" answer was more accurate than a "no" response. Similarly, Juster suggested: *"...a good many respondents with purchase probabilities higher than zero will classify themselves as non-intenders."* (p. 7).

That is, he argued respondents must first consider the probability that they will behave in a particular way before they can assess whether the probability meets the threshold required for a "yes" response, or falls short of it, necessitating a "no" response. Not only do these comments suggest that a finer instrument to accommodate these levels of uncertainty is required, but they also suggest that intentions are essentially disguised probability statements. If this is so, Juster argued, researchers should collect probabilities directly, instead of implicitly requiring respondents to undertake their own imprecise scaling in order to fit their probability of performing a behaviour to a particular intentions instrument.

Juster suggested using an eleven point scale to collect probability information as this enabled him to incorporate the notion of chances, or odds, out of ten, with which he argued respondents would be familiar. Although an eleven point scale may appear analogous to the longer intentions scales sometimes employed, the latter remain ordinal scales whereas Juster's scale, in introducing the idea of odds, became an instrument capable of producing metric level data. Researchers, therefore, no longer had to decide at which point to differentiate likely buyers from likely non-buyers, but could simply use the mean response to estimate the proportion of the sample who would purchase the item in question, or perform the behaviour explored.

In addition, Juster believed that the use of a more detailed scale which incorporated a greater number of points would enable respondents to record their buying plans more accurately. He surmised that a more finely graduated scale would reduce the proportion of people who stated either that they had no intentions, or that they did not know their intention of acting. Instead of attempting to identify those points on an intention scale that denoted likely buyers, and basing predictions on this sub-group only, Juster used the mean purchase probability to predict the proportion of buyers. He hypothesised that this would enable a more accurate interpretation of non-intenders' likely behaviour, a finer differentiation between buyers and non-buyers, and an overall improvement in the purchase estimates. In developing his scale, he was guided by research conducted elsewhere, which also sought to identify a more accurate predictor than intentions.

Working independently of Juster, Ferber & Piskie (1965) were also employing a form of subjective probabilities. They used an 11-point "Plan-o-meter" to ascertain whether purchase probabilities could elicit more accurate estimates for purchases of durables over a six month period. The Plan-o-meter had only three labels at 0, 5, and 10 (see Figure 1).

Figure 1 **Ferber & Piskie's Plan-o-meter**

10	Certain
9	
8	
7	
6	
5	Fifty-fifty
4	
3	
2	
1	
0	No plans at all

However, the questions put to respondents sought their buying plans, rather than their likely behaviour or purchase probabilities. Thus despite the move to a probability type scale, the labels used and the questions posed were more akin to an intentions survey. In addition, Ferber & Piskie dichotomised the data, according to the level of probability, in some of their analyses, thus effectively defeating the purpose of

collecting a continuous variable. Though they found the dichotomised intentions variable made a significant contribution to explaining the variance in actual purchase rates, the raw plan-o-meter probabilities did not increase the overall goodness of fit. Yet since these served as the basis of the dichotomised intentions, Ferber & Piskie concluded that subjective probabilities could not be dismissed as having no value, though they could not specify more precisely what this value might be. In addition, the tri-modal distribution seems attributable to the labelling used, and foreshadows the comment made by Haley & Case (1970), among others, about the attractiveness of verbal labels.

Byrnes' (1964) research represented a more detailed use of subjective purchase probabilities. Again, independently of Juster, he developed an 11-point scale where all points were associated with an adjectival description, and which he used to predict purchases of household durables (see Figure 2).

Figure 2 Byrnes' Scale

10	Absolutely certain to buy	10
9	Almost certain to buy	9
8	Much better than even chance	8
7	Somewhat better than even chance	7
6	Slightly better than even chance	6
5	About even chance (50-50)	5
4	Slightly less than even chance	4
3	Somewhat less than even chance	3
2	Much less than even chance	2
1	Almost no chance	1
0	No chance	0

Although the mid-point was labelled "about even chance, 50-50", Byrnes' study also produce a tri-modal distribution. When commenting on these findings, Juster (1966) considered whether the peak of responses occurring at the mid-point of the scale was due to the label "50-50 chance" associated with that point. His initial hypothesis suggested that many respondents were uncertain about their probability of performing a behaviour, and selected the midpoint for two reasons. First, it was the only point with a quantitative label and, second, this label captured the idea that their probability of acting was greater than zero, but less than one.

To address this possibility, Juster modified the scale so all points had both qualitative and quantitative descriptions (see Figure 3).

Figure 3 The Juster Scale

10	Certain, practically certain	(99 in 100)
9	Almost sure	(9 in 10)
8	Very probable	(8 in 10)
7	Probable	(7 in 10)
6	Good possibility	(6 in 10)
5	Fairly good possibility	(5 in 10)
4	Fair possibility	(4 in 10)
3	Some possibility	(3 in 10)
2	Slight possibility	(2 in 10)
1	Very slight possibility	(1 in 10)
0	No chance, almost no chance	(1 in 100)

He then compared this scale to an intention scale in the Quarterly Survey of Intentions Experiment. In addition to collecting information about the household's projected behaviour over the next 6, 12 and 24 months, Juster asked a number of attitudinal questions and administered test questions using his scale to ensure respondents had some familiarity with the instrument before he explored the key behaviours of interest. In this way he effectively joined the models based on socio-economic and general financial data with those concentrating more directly on individuals' own predictions. Juster argued that contextual questions were important to help respondents consider the factors that might influence the behaviour of interest. He collected details about actual purchase behaviour after 6 months and found the distribution of probabilities was shaped like an inverse "J"; there were peaks at 0 and 1, but no evidence of a peak at 0.5 (as Byrnes, 1964 had found).

Juster's analysis confirmed his hypothesis that the scale would enable a more accurate assessment of non-intenders' behaviour. He noted: *"A very substantial number of respondents reporting no intentions to buy reported a non-zero probability of purchase, as did most of those reporting "don't know"." (1966, p. 21).*

More importantly, the non-intenders were divided out into more homogenous subgroups when classified by probabilities, and the proportion making purchases was lower among non-intenders with a zero purchase probability (5%) than among those with non-zero probabilities (11%). Juster developed regression equations to explore

the relationship between purchases and intentions or probabilities and found that probability variables explained up to twice as much variance in purchase behaviour than did the intention variables (Juster, 1966).

Juster suggested that probabilities displayed a superior explanatory ability because of their stronger relationship with underlying factors, such as the household's or individual's financial situation, which clearly helped determine the actual purchase rate. However, while clearly asserting the superiority of probability variables, he explicitly recognised difficulties in the analysis he had undertaken which limited his ability to generalise his results[17].

Juster also suggested that his scale had a propensity to underestimate actual purchases, though it is possible that two factors compounded any tendency present. First, Juster included some unmatched data in his study (he collected predictions from 451 people, but purchase behaviour from only 395) and it is possible that this introduced error which might explain the underestimation. (However, it is equally possible that the inclusion of these cases may have lead to error in the opposite direction, i.e., overestimation.) Second, Juster used 0.025 and 0.975 as probabilities for the 0 and 10 values. Because of the high proportion of "0" responses, it is possible that this produced an upward error, though insufficient to counter the overall trend to underestimation. Nor did Juster provide an explicit rationale for having selected these two probability values.

In addition, Juster noted that respondents assigning a zero probability of buying an item within six months, but a non-zero probability of buying that item within twelve months, actually had a "true" six month purchase probability greater than zero. Consequently, the mean probabilities of purchase within six months underestimated the actual purchase rate. A weighted average of the six and twelve month probabilities proved more accurate that the six month estimates alone.

Overall, Juster's work made several important contributions to the debate over how best to predict respondents' behaviour. First, he developed and provided some support

---

17 These related to the number of cells containing too few cases to allow meaningful comparisons.

for the contingent action hypothesis which asserted that intentions and attitudes differed, and that attitudes could not be relied on as a basis for predicting either intentions or behaviour. Second, he theorised that intentions were disguised probability statements, the predictive ability of which would improve if the underlying probabilities themselves were collected. Both his 1960 and 1966 studies supported the distinction between intentions and probabilities, and provided mounting evidence attesting to the superior predictive ability offered by probability scales.

4.5 Refinement of Subjective Probabilities

At the same time as Juster was developing the scale which later came to bear his name, other researchers were also examining alternative methods of assessing consumers' propensity to buy a particular item or to behave in a specific manner. Although not explicitly acknowledged by Juster, the notion of eliciting subjective probabilities from consumers was not new, but had been dismissed as too difficult and too demanding a task. Requiring such a procedure of respondents would, critics argued, almost inevitably lead to meaningless and inaccurate data (see Hogarth, 1975, whose criticisms are examined in detail in section 4.7). However, other researchers tested these views empirically; thus allowing the accuracy of various instruments to be experimentally established.

Work conducted after Juster's key publications joined together these research threads, (of economic or probability based predictions), and the refinement and further testing of probability scales gathered momentum. Shortly after Juster's (1966) work, Stapel (1968) conducted a survey asking respondents their "percent chance" of purchasing a new or used car. His scale contained similar descriptions to those used by Juster, and he found similar results insofar as purchase rates rose as probability levels rose. However, inadequate re-interviewing, especially of the people who stated a zero probability, or who stated no probability at all, makes it difficult to discuss the relationship between Stapel's and Juster's results in more detail.

Clancy & Garsen (1970) used an 11-point "chances of buying" scale to explore consumers' reactions to a new product. Although using a scale analogous to the Juster Scale, Clancy & Garsen's chief interest lay in whether monadic or comparative reference scales provided more accurate predictions of subsequent behaviour. Although they recommended researchers use a comparative scale, they did not

comment specifically on its predictive power.

Gruber (1970) explicitly set his work in the context of Juster's (1966) findings[18], and suggested probability scales might have particular relevance where the proportion of non-intenders was large: *"One might expect that the larger the percentage of non-intenders, the greater the potential utility from the more discriminating and predictively more accurate purchase probability scale."* (p.23).

He went on to explore further the relationship between intentions and probabilities by seeking respondents' buying intentions and purchase probabilities after exposure to colour photos of a food product prototype. When comparing the two scales, Gruber found a high positive correlation, with little variation at either extreme. However, the middle categories of the intention scale, especially the point labelled "probably would buy" were associated with a range of probability values.

Gruber aimed to calibrate intention and probability scales and so had less interest in their respective predictive abilities. In doing this, he did not detect the same magnitude of difference between the scales as that observed by Juster, (1966); Byrnes, (1964); or Ferber & Piskie (1965); he attributed this largely to his choice of a non-durable stimulus, and to the different behaviours traditionally associated with consumer goods. However, since he compared neither set of data to respondents' actual purchases, Gruber's conclusions remain rather limited.

Although Gruber's work did not add substantially to the empirical evidence supporting the use of probability scales, other work, related more specifically to the context in which Juster conducted his work, did. Gabor & Granger (1972) used the Juster Scale to collect information about respondents' purchases of cars and consumer durables; respondents were re-interviewed 12 months later about their actual purchase behaviour. Like Byrnes (1964), they also noted a very high proportion of zero probabilities, yet discovered that nearly two thirds of all purchases were made by

---

18 He used the Juster Scale with the following question:

"Taking everything into account, what do you think would be the chances that you would buy this product?" (p. 24).

about 5% of this group.

They also observed rises in purchase rates as probability levels rose, but, like Juster, found the predictions greatly underestimated actual purchases (181 compared with 306). However, a number of external factors may have intervened and affected these estimates. While it is possible to suggest some of these, such as the introduction of smokeless zones, may have affected purchases of heating products, it is more difficult to quantify the extent to which these were responsible for the errors noted. Gabor & Granger hypothesised that the type of purchase, whether a new acquisition, or a replacement item, could contribute to discrepancies between people's probabilities and their behaviour. That is, replacements are often unplanned purchases (where an appliance suddenly breaks down, for example), and 80% of purchases made by the zero probability group were for replacements.

Pickering & Isherwood (1974) also reported research concluding that non-intenders accounted for a large proportion of total purchases, and a large proportion of the variation in purchase rates over time. Arguing that intentions data did not fully explain which consumers would purchase particular goods, Pickering & Isherwood suggested that probabilities could not only do this, but could also help predict purchase levels.

Their work investigated both the accuracy of probabilities and the context within which the behaviours in question were set. Hypothesising that non-intenders, or those with zero probabilities, went on to purchase because they had experienced difficulty confining their predictions to a particular time frame, Pickering & Isherwood conducted a three stage study to test their view. This consisted of initial interviews, during which they sought purchase probabilities, plus two follow-up interviews four and fourteen months later during which they elicited respondents' actual purchase behaviour. Based on Juster's instrument, their scale also used 11 points, although only two of these, 0 and 10, were accompanied by a verbal description. However, the card also contained an explanation that read: *"A score of 8 would mean that you were 80% certain, a score of 1 that you were 10% certain and so on."* (p. 208).

Pickering & Isherwood found a similar probability distribution to that noted by Juster

(an inverse J shape), but noted that many of those who gave zero probabilities to the likelihood that they would buy a range of household durables did purchase those items. However, despite this inconsistency, they detected a clear trend: as the probabilities respondents gave increased so too did the proportion of purchases made by members of the probability groups.

They attributed at least some of the inconsistency between what the people assigning zero probabilities to particular behaviours said, and did, to the time periods used (3, 6, and 12 months), and suggested that probability statements were reliable predictors, but were less useful as predictors of behaviour during a short period (a problem Juster had already noted, and to which Neter & Waksberg, 1964, had also referred). In addition, they referred again to the problematic influence of unanticipated events which appeared to have intervened between the interview and the time at which the purchase was expected (or not expected) to have occurred. They also raised again the point noted by Gabor & Granger (1972), that the type of item (high or low ownership) combined with the length of planning period may affect the accuracy of the estimates and the direction of any errors. Pickering (1975; discussed earlier) interviewed a number of respondents who had given predictions at odds with their actual behaviour and found unexpected changes to individuals' financial situation explained many of these discrepancies.

Isherwood & Pickering (1975) also noted the importance of assessing respondents' willingness to purchase an item together with their ability to do so. Their comparison of respondents' attitudes towards and probabilities of a variety of purchase behaviours with their actual behaviour revealed that purchase probabilities were strong discriminating variables, especially when combined with measures of consumers' socio-economic status and details of their financial background.

Overall, Pickering & Isherwood's (1974) use of a scale with only labelled two points did not appear to affect their estimates, and they confirmed many of Juster's earlier conclusions. However, in seeking to extend these conclusions, they again raised questions about the time frames within which respondents situated their future behaviour and pointed to the problems caused by intervening factors. Unable to resolve these problems, Pickering & Isherwood called for further work to explore them in more detail.

Morrison (1979) took up the challenge to define more precisely the relationship between intentions, probabilities and behaviour. He aimed to address the problems of underestimation and overestimation by decomposing the relationship into a three-step model. At the first step he collected stated intentions and transformed these into estimates of "true" intention through various regression procedures. These estimated "true" intentions then underwent a process which transformed them to unadjusted purchase probabilities. This process occurred via an Exogenous Events Model which attempted to allow for changes in circumstances which could affect intentions. Finally, the unadjusted purchase probabilities were transformed into estimated purchase probabilities via a probability adjustment model. This included an adjustment to accommodate any bias caused by the average purchase intention not equating to the proportion who actually did purchase. Morrison did not explicitly test his model by collecting before and after data, but instead reanalysed Juster's (1966) data. While providing a framework for more detailed comparisons of intention and probability data, Morrison did little more than highlight issues (such as the length and description of the scale, and the time frames nominated in questions) that Juster and others had already suggested required further research. Overall, Morrison concluded that, despite the elegance of his model, employing this sort of procedure was less desirable than improving the measurement of variables used to predict behaviour.

This stage of the development of probability models has seen a clear movement away from the traditional "stock adjustment model" (Juster & Wachtel, 1971, p.564), with its emphasis on individuals' personal financial situations and their view of general economic conditions. Replacing this was a model which began to incorporate measures of consumers' anticipated behaviour, which in turn led to a model where the anticipation, or probability, variable formed the basis of subsequent predictions. All empirical comparisons of intentions and probability data suggested probabilities possessed a predictive accuracy at least as good as that displayed by intentions data, and, in some cases, accounted for more than twice the variance in purchase behaviour than did intentions.

Yet despite the clear progress made, problems had emerged. The accuracy of predictions made within short time periods remained questionable, with a number of researchers recording high proportions of people who originally stated zero probabilities of purchasing as subsequent purchasers. Researchers noted some

evidence to suggest that questions eliciting purchase probabilities, and the context in which these questions were placed, prompted a more thoughtful consideration of events or circumstances likely to impinge on a purchase decision. Despite this, completely unanticipated events, continued to introduce error to the estimates. However, these errors did not appear systematically, with some researchers reporting underestimates (see Juster 1960 & 1966; Heald, 1970; Gabor & Granger, 1972; Pickering & Isherwood, 1974 and some overestimates (see Byrnes, 1964; Juster, 1966; Clawson, 1971; Pickering & Isherwood, 1974). Despite the identification of errors within some of these individual studies (see the discussion in Juster, 1966, and Gabor & Granger, 1972), it remained difficult to identify a clear pattern that could improve the accuracy of future studies.

The main explanation of discrepancies between predicted and actual behaviour has been unanticipated events, but since these are by definition, difficult to predict, most researchers have shown less interest in analysing them retrospectively than in attempting to minimise their future effects. Thus research from here has concentrated on expanding the range of uses to which the Juster scale has been put, and validating the structure of the scale itself.

4.6 Validation and Extension of the Juster Scale

Previous users of probability scales had sought predictions about respondents' likely purchases of cars, or household durables and, with the exception of Gruber (1970), none had produced estimates about purchases of fast moving consumer goods, or services. Gan *et al* (1985) compared the accuracy of three month purchase predictions for a range of household durables, services and fast moving consumer goods using both traditional intention questions and the Juster Scale. Their results confirmed those obtained earlier by Juster and others, and once again found that non-intenders accounted for a large proportion of purchases. Overall, respondents who declared they had no buying intention made more than twice as many purchases than did those who assigned a zero probability to the behaviour (25% compared to 12%). Like Juster earlier, Gan *et al* also found that while buying intentions data did not differentiate clearly between buyers and non-buyers, the mean purchase probabilities of buyers were consistently higher than those of non-buyers.

However, the ultimate test of the Juster Scale lies in its ability to predict actual behaviour. Gan *et al's* results revealed only four of the eleven predictions fell within a 20% error margin. Purchases of nearly all durables (except car purchases) were overestimated by between 45% and 245%. However, estimates of service use and of fast moving consumer goods behaviour all came within a 40% error margin. Interestingly, all but the estimate with the smallest error (4.7%) overestimated the actual purchase rate. Again, this raises the possibility, discussed by Pickering & Isherwood (1974), that the scale is less accurate when used to project short term behaviours; however, it also questions Juster's (1966) finding that the scale consistently underestimates actual purchases.

Brennan & Esslemont (1994) explored this question, among other issues, in their test of the scale's ability to predict demand for different brands, within two categories of fast moving consumer goods: soups and yoghurts. They asked respondents their likelihood of purchasing each of three brands of soup and four brands of yoghurt within a four week period, after which they re-interviewed those respondents who had previously consented to a further interview. The prediction errors in this study were much smaller; all were under 15% and most were under 10%, and all overestimated actual purchases. Thus while these results questioned the problems Pickering & Isherwood (1974) had associated with shorter time frames, they supported a view emerging in more recent studies that the scale is in some way biased towards overestimation, (see also Byrnes, 1964; Clawson, 1971; Gan *et al*, 1985; Day, Gan, Esslemont & Gendall, 1991). Brennan & Esslemont went on to consider how this error margin could be reduced and surmised that calculating the predicted shares of each brand within a given category would lead to lower overall estimates. After performing this calculation for all brands, they found that they all now had an error of less than 5%. Furthermore, the error now appeared more randomly distributed and the tendency for errors to occur solely in the direction of overestimation had virtually disappeared.

At the same time as these research efforts were underway, other researchers were exploring characteristics of the scale and its administration. Day *et al* (1991) compared two versions of the Juster Scale: the original version, which contained both adjectival descriptions and numeric probabilities, and a second version, containing

only the numeric probabilities. In addition, they tested the scale in a self-completion context; hitherto it had been administered solely via face-to-face interviews. Purchase probabilities for a range of household durables, services and fast moving consumer goods were elicited from a panel of respondents who completed a second questionnaire about their behaviour three months later.

Overall, the standard Juster Scale produced more accurate estimates, had a more even distribution, and showed a lower item non-response level. While Day found the highest level of error associated with estimates of durable purchases, both the services and fast moving consumer goods predictions also had large errors associated with them. However, it was clear that a correlation between the predicted purchase rate and the error existed. As Clawson (1971) noted, the higher the level of prediction, the lower the error associated with that estimate. However, Day's *et al*'s findings lent support to the suggestions of Gabor & Granger (1972) and Pickering & Isherwood (1974) that ownership levels and the purchase planning period influenced the level of accuracy associated with the predictions. Overall, both versions of the scale produced accurate estimates, though Day *et al* recommended using the final version which Juster developed for the reasons outlined earlier.

As Day *et al* (1991) noted, marketers are often concerned to identify both the average purchase probability and the number of items each household will purchase. Hamilton-Gibbs, Esslemont & McGuiness (1992) reported a study which addressed this latter question and which compared two means of administering Juster-Scale questions. Administration of the first method simply involved repeating the traditional Juster question and substituting new quantities until respondents gave a zero probability, while the second method utilised a constant sum approach. Here respondents received 10 counters (each representing a 0.1 probability) and used a specially designed grid to distribute these counters to show the probability they would make various purchases of an item over the next four weeks. The seven products tested were all among the most frequently purchased grocery items in New Zealand. Respondents were then re-interviewed exactly four weeks after the initial interview to ascertain the actual number of items purchased. Overall, predictions elicited via the constant sum had a lower error across all but two of the products, and all errors were less than 10%, while errors associated with the direct question ranged from 3% to

35%. Interestingly, the constant sum errors were all overestimates of actual purchase while the repeated questioning method errors were virtually all underestimates. There was no obvious reason why this pattern emerged, but the strong performance of the constant sum approach to collecting purchase probabilities concurs with Axelrod's (1968) conclusions.

As well as experimenting with different versions of the Juster Scale, Brennan also explored the method through which it could be administered (Brennan, Hini & Esslemont, 1994). Day *et al* (1991) had earlier demonstrated that the scale could be employed in self-completion questionnaires, where it would be contained in the questionnaire itself, rather than displayed on a showcard. However, telephone interviewing offers little opportunity to introduce visual stimuli, and so presents a greater challenge to researchers wishing to employ the Juster Scale.

Brennan *et al* addressed this challenge in two ways. First, they mailed a copy of the Juster Scale to a group of randomly selected respondents together with a covering letter advising them that they would be telephoned and asked to participate in a survey, and asking them to keep the showcard near their phone. The second group received no prior notification but were contacted by telephone and read the following statement:

> *"We would like to know what the prospects are of you buying certain products during the next four weeks. I would like you to answer on a scale of "zero" to "10". If you are certain, or practically certain, that you will purchase a product then you would choose the answer "10". If you think there is no chance, or almost no chance of purchasing, the best answer would be "zero". If you are uncertain about the products, choose another answer as close to "0" or "10" as you think it should be. You can think of the numbers as chances out of "10". For example, 3 would mean 3 chances in 10, or there was some possibility. An answer of 6 would mean 6 chances in 10, or there was a good possibility of you purchasing the product."*

While the samples employed in the test were small, the verbal probability scale was consistently more accurate than the mailed version (it produced more accurate

estimates for eight of the nine products tested). This result queries Gendall, Esslemont & Day's (1991) conclusion that the Juster Scale's predictive ability derives from the combination of adjectival and numeric labels associated with each point[19]. Instead, it suggests these descriptions can be condensed and represented verbally without incurring a loss in the scale's predictive accuracy (see also Pickering & Isherwood, 1974). While this finding requires replication, with larger samples and in a variety of contexts, before it can be accepted, it suggests potential exists for the scale to be more widely employed. Given that telephone is now the most widely used survey research medium, these results point to an important route down which future research may proceed.

Overall, these studies document a strong body of evidence supporting if not the adoption, at least the further exploration, of probability scales, in particular the Juster Scale. However, while this chapter has thus far documented the theories underlying the development and use of subjective probability scales, critics, notwithstanding the accumulating body of research, have questioned respondents' ability to develop and articulate subjective probabilities. The next section explores these arguments, and their relevance to the Juster Scale, in greater detail.

4.7 Problems Associated with the Estimation of Subjective Probabilities

Tversky & Kahneman (1973 & 1974) conducted several studies designed to explore how people assess the probability of uncertain events occurring. They hypothesised that people relied on heuristic principles which reduce the complex task of assessing probabilities to simpler judgmental operations (Tversky & Kahneman, 1974, p. 1124). They discussed three heuristics commonly used. First, representativeness, where individuals make assessments of the similarities between objects by considering, for example, "the probability that object A will belong to Class B" (1974, p. 1124). Second, adjustment and anchoring, where respondents adjust an initial starting value to provide a final value reflecting their probability. Finally, availability, where respondents' judgments are aided by the ease with which they can bring instances or associations to mind (Tversky & Kahneman, 1973, p. 208).

---

19 Juster (1966) also raised the question of whether the adjectival descriptions helped or hindered the survey (or had no effect on it at all) and noted that preliminary research into this question tended to support the view that the adjectival descriptions detracted from the scale's accuracy.

However, as Tversky & Kahneman stressed, each heuristic may be accompanied by a number of biases which affect the estimates. Without discussing each of these in detail, Tversky & Kahneman enumerated many potential problems, all of which could seriously affect probability estimates. Their key points are summarised by Roshwalb (1975) who listed three general conditions, or problems, symptomatic of respondents' difficulty in coping with subjective probabilities. First, he argued that people overestimated the occurrence of events with a low probability of happening, and underestimated those with a high probability of occurring, a finding alluded to by Juster (1966) and documented earlier by both Gabor & Granger (1972) and Pickering & Isherwood (1974). Second, he posited the existence of a gambler's fallacy which suggests that an event which has not occurred for some time is, as a consequence, more likely to occur in the near future. For example, if a tossed coin gave six heads in a row, those exhibiting "gambler's fallacy" would believe the chances of it landing tails in the next throw were greater than 50-50. Finally, he suggested that people will overestimate events favourable to them and underestimate those that are unfavourable (see also Kahneman & Tversky, 1975).

These latter two problems relate more to the prediction of which external events will come to pass (how likely A is to occur?) than they do to respondents' prediction of their own behaviour (how likely am I to perform behaviour A?). In a voting context, it would be the difference between these questions:

> *How likely is it that Party A will win the government after the next election?*

as opposed to:

> *How likely are you to vote for Party A?*

While this difference does not mean that these two biases described by Tversky & Kahneman, and summarised by Roshwalb, do not affect individuals' estimates of their own behaviour, the examples on which they base their conclusions differ in kind from the task individuals perform when estimating their own behavioural probabilities. However, the first bias, that of over- and under-estimation remains closely associated with individuals' predictions of their behaviour, and has already been raised by a number of Juster Scale researchers (see also Ravinder, 1992). While Brennan &

Esslemont (1994) presented a means of reducing this error,[20] their suggestion requires further testing before it can be held out as a solution to this problem.

Hogarth (1975) raised a more fundamental problem when he queried respondents' ability to project their own behaviour. Drawing on Tversky & Kahneman's work, he concluded: *"...man, as a selective, step-wise information processing system with limited capacity, is ill-suited to the task of assessing probability distributions." (p. 276).*

Yet despite his obvious lack of optimism, Hogarth did not dismiss this task as impossible, but went on to suggest how the method of questioning could be structured to improve the chances of eliciting accurate probabilities. While the majority of Hogarth's comments related to predictions made by experts, rather than members of the general public, his comments about question context have a wider application, and Juster (1966) had already commented on this issue[21].

However, he concluded that thinking in terms of probabilities represents a new way of thinking for most people, thereby explicitly contradicting the rationale that Juster established as the basis of his scale. Further support for Hogarth's queries came from Seymour, Brennan & Esslemont (1994) who reported that respondents administered probability questions via a constant-sum method showed a better understanding of the task than did respondents who answered multiple probability questions. However, they noted that: "*Very few respondents realised that the probability of buying any items should relate inversely to the probability that they would buy zero items. On the occasions that the two probabilities did correspond correctly, it was often due to chance rather than* [to] *a complete understanding of the concepts*" (p. 29). Despite this, both methods of collecting probabilities provided some very accurate estimates,

---

20 Their method involved predicting the number of purchases within a product category and using this as a basis for calculating the number of buyers of each brand.

21 Juster (1966, p.38) noted: "*There is much to be said, in my view, for the proposition that probability judgments would be sharpened by making the household explicitly aware of all the considerations that ought to have some relevance to purchase prospects. Thus a survey which, prior to asking about probabilities, contains questions on the household's income, income prospects, asset holdings, stocks of durables, repair experiences on durable stock, actual and prospective labour market participation, etc., may obtain more accurate judgments than a survey which does not.*"

with mean prediction errors ranging from 11% to 16%.

Whether thinking in terms of probabilities is too difficult for respondents, or whether as Juster surmised, intentions are disguised probability statements, can only be determined by repeated testing of the instrument's predictive power (see Ehrenberg & Bound, 1993). Thus while the concerns Hogarth presents cannot be dismissed, the real test of how seriously they affect an instrument's predictive ability can only be ascertained through rigorous and repeated use of that instrument in a variety of circumstances.

4.8 A Rationale for the Use of the Juster Scale

Research to date has highlighted a number of deficiencies in the models which posit behavioral intention as the precursor of behaviour. Although many such models still enjoy widespread use and popularity, a growing number of researchers have suggested their limitations are endemic, and not easily remedied. Consequently, a body of research has documented the growth and refinement of subjective probability scales, and has presented evidence which suggests the Juster Scale, in particular, has a superior predictive ability. While the Juster Scale is also prone to error (Juster reported an underestimation bias, though some subsequent studies have reported an overestimation bias), explicit comparisons of it with intention scales have concluded that the Juster Scale is more highly correlated with the behaviour it sought to predict. In other words, imperfect as it is, it appears to offer an opportunity to obtain more accurate estimates than attitude or intention-based scales have hitherto elicited.

Virtually all tests of the Juster Scale have sought information about consumers' future purchase behaviour and, while this has ranged across various product categories, few researchers have used the Juster Scale to predict other types of behaviour. Logically, there seems no reason to suppose that a scale used to predict conditional behaviour in one context could not also predict behaviour in another context.

Given that the questions traditionally used to elicit voting intention are analogous to dichotomous intention questions[22], the discrepancies between predicted and actual behaviour are perhaps not surprising (see Chapters 2 and 6 for a detailed discussion of this problem). Since the problems experienced in eliciting voting intentions are akin to those experienced by researchers eliciting buying intentions, it is possible that the substitution of subjective probabilities for voting intentions could produce more accurate estimates. Accordingly, this research was designed to explore the hypothesis that estimates of respondents' probability of voting for different political parties would provide more accurate predictions than their voting intentions.

4.9 Summary

This chapter has summarised the movement away from the collection of intentions data to the elicitation of subjective probabilities, the problems which provided the impetus for this shift, and some of the consequences that have arisen. In particular, the development and refinement of probability scales, and the process by which the Juster Scale arrived at its present form were explored.

Empirical tests of the Juster Scale, including its application to different types of purchase behaviour, the use of different time frames, and the medium though which it was administered, have been discussed. Criticisms of the use of subjective probabilities have also been examined, and the extent to which these apply to the current work considered.

Finally, the chapter concluded by outlining a rationale supporting the use of the Juster Scale to elicit likely voting behaviour.

---

22 That is, the traditional question:
If an election had been held yesterday, which party would you have voted for?
is analogous to:
If an election had been held yesterday, would you have voted for Party A...Party N?

CHAPTER 5

CONCLUSIONS FROM THE LITERATURE

5.1 Introduction

Chapter one outlined the broad research objective, to examine the effects of different questions and question contexts on estimates of voting behaviour. The chapter then specified more precise objectives, namely:

1. To explore whether a probability question based on the Juster Scale could estimate turnout accurately and be used to weight the estimates obtained from the voting intention questions.

2. To examine whether changes in the context surrounding voting intention questions would lead to a reduction in the proportion of undecided respondents and an increase in the accuracy of the estimates.

3. To examine whether increasing the confidentiality associated with the interview, and approximating more closely the conditions surrounding voting behaviour, would reduce the proportion of undecided respondents and increase the accuracy of the estimates.

4. To investigate whether the use of a probability scale (the Juster Scale), rather than an intention question, would improve the accuracy of the estimates obtained.

The chapters thus far have documented the past research undertaken to address these issues or related questions, the following chapters present the findings of empirical research undertaken to address these objectives. The remaining sections in this chapter summarise the key findings from the literature as they relate to these objectives.

5.2 Summary of the Literature

Chapter 2 elaborated on the problems affecting polling research and discussed the various explanations put forward to explain inaccuracies in a number of pre-election polls. These wide ranging explanations included the sampling procedures adopted, the use of random or quota samples, the number of callbacks made, and the use of clustered samples. However, while poor sampling practices will inevitably result in biased samples, and hence less accurate results, a variety of other factors also appear to have affected the accuracy of polling estimates.

These factors include the problem of identifying, *a priori*, those respondents who will go on to cast a vote. Numerous researchers have cited inadequate identification of likely voters as an explanation of polling inadequacies. However, while the evidence supporting the existence of this problem is well documented, knowledge of its scope, and means of addressing it, have not received such detailed attention. While some voting intention, or turnout, scales, do exist, these tend to suffer from one of two problems. First, some are very crude, taking the form of simple ordinal scales whose predictive accuracy is neither established in this context, nor well supported by research undertaken in other contexts. Second, the alternatives to this are often complex indices, requiring the collection and integration of several variables. Although the latter have a stronger predictive accuracy, they remain demanding to implement and so do not appear to have been widely adopted by pollsters.

Pollsters must also deal with an equally perplexing problem: how best to allocate respondents who are unwilling or unable to nominate a party or candidate for whom they plan to vote. Researchers have devoted considerable effort to exploring the factors underlying indecision, including examination of the questionnaire format, as well as the specific wording of the voting intention question itself. Consequently, some researchers have attempted to develop a rapport with respondents before examining more sensitive topics, while others have interrogated respondents in the hope of identifying the direction in which they "lean".

Others researchers have developed instruments which attempt to capture respondents' equivocation, while yet another group has sought to increase the level of confidentiality associated with respondents' answers. However, these efforts have

often left unanswered problems of bias, and their implementation sometimes creates difficulties in administering the questionnaire. As a consequence, researchers have also examined how to model the behaviour of undecided respondents. However, although some encouraging results have emerged, the success of these models remains variable, leading some to conclude that research efforts might be better directed at reducing the scale of the problem.

An underlying theme of this chapter has been the problem of eliciting respondents' voting intentions. While the disappointing results of pre-election polls can be attributed to many factors, an increasing number of researchers has identified question wording as playing a key role in the success, or otherwise, of poll predictions.

Chapter three explored the problems associated with question wording in more detail and also examined issues related to questionnaire design. This chapter noted the large differences in response distributions that can result from apparently minor changes in wording. In addition, the chapter examined fundamental decisions researchers must make about the questions they employ, and the implications these decisions have for the questionnaire, and the research design, implementation and interpretation generally.

In discussing the various effects the type of questions employed may have, the chapter noted that, while behavioural or factual questions are not without problems, they appear to have fewer difficulties associated with them than do opinion, or non-factual, questions. However, opinion questions appear to play an important role in creating a context within which researchers may place behavioural questions. This practice explicitly recognises the problem, identified in Chapter 2, that fewer respondents may give undecided responses if they have already thought about how their opinions and past behaviours relate to the survey topic. Not unexpectedly, the practice of providing a context also incurs some problems, in particular, it raises the question of whether the context helps consolidate respondents' views on their likely behaviour, or whether it introduces error into the resulting estimates.

In addition to discussing errors inherent in question wording, the chapter also considered biases respondents can introduce. First, the chapter examines respondents' understanding of survey questions, and the related issues of their knowledge of the research topic and its relevance to them. Further exploration of context effects

suggested the provision of questions related to the survey topic may make the actual behavioural questions more salient and may also increase the accuracy of respondents' answers.

Questions investigating voting behaviour appear especially prone to social desirability bias, and the chapter also examined procedures, such as secret ballots, which may ameliorate the effects of this problem. Overall, the chapter concluded that a context, or a funnel of questions leading respondents through to the behavioural questions may reduce the proportion of undecided responses, though the effect of contextual questions on the accuracy of the estimates remains open to some debate. In addition, the chapter noted evidence suggesting that the provision of more confidential response mechanisms appears likely to achieve further reductions in undecided responses.

However, while these measures could no doubt reduce the proportion of people declaring themselves undecided, they do not resolve a residual problem. That is, many respondents may give undecided responses not because they have no opinion, or because they feel too timid to voice their opinion, but because they have considered the various options in detail, yet find themselves unable to make an unequivocal choice between these options. Thus their indecision results from the absolute nature of the question put to them.

Since voting questions typically take the form of intention questions which explicitly require absolute judgments it seems unlikely that measures designed to build a rapport with respondents, or to increase the confidentiality of their answers, will completely eliminate the undecided group. Chapter 4 documented the development and refinement of intention scales and discussed the problems associated with these. In particular, the chapter examined the poor predictive performance of questions employing intention scales and outlined factors that may have contributed to this error. Problems inherent in intention scales, particularly respondents' interpretation of the various scale points, were also discussed.

Chapter 4 advanced an alternative methodology, based on the collection of probabilities, as a possible replacement for intention-based questions. Detailed discussion of one particular scale, the Juster Scale, suggested that this scale has superior predictive accuracy when compared to intention scales. Furthermore,

predictions obtained using the Juster Scale have correlated highly with respondents' subsequent behaviour. Accordingly, the chapter concluded by proposing to use the Juster Scale to collect estimates of respondents' probability of voting for the various parties represented in an electorate.

In summary, these chapters documented the range of problems affecting polling accuracy, and the role question wording plays in creating these problems. Additional factors that researchers can control, such as the level of confidentiality associated with the survey, and the choice between relative and absolute response devices, have also been discussed, and proposals for employing these as a means of reducing the proportion of undecided respondents reported in pre-election polls have been put forward.

The remaining section in this chapter draws together the conclusions reached from the preceding chapters.

5.3 Conclusions from the Literature

While at least some of the conclusions put forward here remain the subject of continuing debate, much of this debate centres on the scope and role of the problems, rather than on their existence *per se*.

On the basis of the literature examined, it is clear that respondents' indecision can greatly affect polling estimates in at least two ways. First, respondents may be uncertain as to whether they will vote and, second, when interviewed they may not yet have decided for whom they will vote. In addition, respondents may feign indecision if they fear their answers are somehow inappropriate, or if the questions themselves appear too personal. Researchers can address these problems by using a number of measures.

First, they can develop a series of contextual, or rapport-building, questions which lead respondents gradually to the question of real interest, which may, in turn, address more sensitive or personal issues. Second, they can increase the level of confidentiality traditionally associated with the interview in the hope that respondents will express their views more freely. Third, they may experiment with different types

of questions and response devices in an attempt to achieve a reduction in the proportion of undecided respondents and a simultaneous increase in the accuracy of the survey estimates. Finally, they can use the information about respondents' likelihood of voting to weight the behavioural estimates they give (regardless of whether these are intention or probability based estimates).

These measures form the basis of the research objectives outlined in Chapter 1, and summarised above in section 5.1. The following chapters first describe the manner in which these objectives were addressed, before analysing the results of these research efforts.

CHAPTER 6
RESEARCH METHODOLOGY

6.1 Introduction

The preceding chapters have drawn together the problems and issues relating to turnout prediction and to the allocation of undecided respondents, and have described the research efforts undertaken to address these problems. This chapter has a narrower focus and describes the approach taken in the research presented in the following chapters, drawing on the literature only to illustrate or clarify specific issues, or to provide a rationale for certain decisions.

After a brief overview of the primary research, the chapter examines in greater detail its component parts: the survey sample, its characteristics and selection, the weighting procedures adopted, and how sampling procedures may have affected calculation of the error margins associated with different estimates.

The chapter then documents the questionnaire construction process, discussing the different treatments tested, and the rationale for these. It also outlines the design and piloting of the different questionnaire versions.

The next section discusses the administration of the different questionnaire versions and provides full details of the interviewer training and auditing, and the steps taken to measure the accuracy of respondents' predicted behaviour.

Since few, if any, survey research decisions can be made with total objectivity, (and, even assuming this were possible, respondents are unlikely to respond in a purely objective manner), certain errors will affect all research to varying degrees. Section 6.5 examines possible sources of error, the effect these errors might have, and possible means of countering their effects.

The chapter concludes by summarising the key decisions taken and the logic underlying these.

6.1.1 Overview of the Primary Research

During the period 27 August to October 26, 1993, a total of 4040 face-to-face interviews were conducted with residents of the Manawatu, Miramar and Western Hutt electorates. Two sets of interviews, one prior to and one during the campaign period, were undertaken, and the overall response rate was approximately 59.6%[23].

Following the General Election, a sample of respondents was re-interviewed by phone and details of their self-reported voting behaviour were obtained. This phase achieved an 89.4% response rate. Although it would have been preferable to compare respondents' predictions to actual voting records, the New Zealand electoral and privacy legislation prevented the release of such data, even in aggregate form.

The following sections outline the more specific research decisions taken in relation to these two phases of the research, and their implementation.

6.2 Sample

This section examines the decisions taken with respect to the survey samples. It includes discussion of the sample selection, the identification of individuals with whom interviews were conducted, and subsequent treatments of the data.

6.2.1 Sample Selection and Composition

Although purely random sampling from an up-to-date electoral roll would have increased the precision of the estimates obtained, previous experience suggested that it could also lead to several problems. Hoek & Gendall, 1993 used the most recent electoral roll in their study but found that a greater than expected proportion of the individuals identified had shifted abode, and could therefore not be interviewed. More

---

23 However, since some interviewers (3) failed to follow the instructions they received and did not record full details of the contacts they made (or those they were unable to make), this figure was calculated on the figures provided by the (96) interviewers who completed these records as requested.

The response rate was calculated as follows:

$$\frac{\text{Successful Interviews}}{\text{(Total contacts - ineligible contacts)}}$$

seriously, the electoral roll used as a sampling frame in 1990 was out of date within weeks of its issue, since well over 10% of the eligible electorate chose to register after the preliminary roll was issued. Moreover, the administration of face-to-face interviews drawn in this way is difficult compared to cluster sampling, where time and travel costs are minimised. Finally, since the final roll is typically issued only days before the actual election, using this final version was not a viable alternative.

For these reasons, random sampling from the electoral roll was eschewed in favour of selecting (from the most recent roll) random starting points around which to conduct a cluster of interviews. Although clustering often decreases the precision of survey estimates, cluster sampling can nevertheless bring cost savings and timing advantages.[24] Since both cost and timing were important concerns, the practical advantages offered by clustering appeared to outweigh its disadvantages.

Starting addresses, selected at random from the electoral roll, were plotted on a map of the electorate to ensure that no suburbs were either unrepresented, or under-represented[25]. The starting addresses were then grouped into sets of six for each interviewer. Interviewers approached each dwelling noted on their starting address list and stood with their backs to that dwelling. They then attempted to conduct interviews at the six adjacent houses to the immediate right of the starting address, using the right hand rule to guide their random walk. (Appendix A contains full details of the electorates; details of the instructions and training given to interviewers are contained in Appendix B.) Where dwellings were unoccupied, or obviously non-residential, interviewers substituted these with the next occupied dwelling until they had completed six interviews in that cluster.[26]

---

24 Clustering need not necessarily decrease the precision of the estimates since the sample may be increased to compensate for the increased design effect; any compensating increases in sample size do not normally incur greater costs than would be incurred by simple random sampling.

25 The only electorate not full covered in the interviewing was Western Hutt, a sprawling urban electorate covering a large geographic area. In this electorate, the interviewing concentrated on the Hutt valley area while the northern region of Whitby was excluded from the sample to ensure the research stayed within its budgetary constraints.

26 The right hand rule states that interviewers should turn right at each intersection they encounter until the number of interviews conducted in the cluster is completed.

At each identified dwelling, interviewers used the "next birthday" method to select an individual, 18 years of age or older, to interview.[27] If the individual meeting this criterion was not present, the interviewers attempted to ascertain when he or she would be available for an interview. Interviewers made two additional attempts to interview that individual before abandoning the household and substituting another dwelling. If the individual meeting this criterion was not going to be available within the period during which the interviews were conducted (for example, some potential respondents were students studying at another university, and some individuals were overseas at the time the interviewing was conducted), the interviewers then attempted to secure the co-operation of the individual whose birthday came immediately after that of the unavailable person.

In some cases, dwellings were clearly occupied, but no residents were at home when the interviewer called. When this occurred, interviewers noted the time at which they first called, made an additional call that same day and, if no contact had been established, made a final call the following day. If after three attempts at contact the interviewer was still unable to arrange or conduct and interviews, the dwelling was replaced in the manner described above.

If residents declined to answer the door, despite persistent efforts to attract their attention, interviewers were instructed to treat the response as a refusal, and to substitute another dwelling as described above.

Where hospitals or other care facilities fell within a cluster, interviewers substituted the institution on the grounds that it was not ethical to approach people who were unwell or recuperating from illness. Any respondent, in fact, who was clearly incapacitated in some way was treated as ineligible to complete the survey and replaced.

---

27 The "next birthday" procedure, as its name suggests, involves identifying and interviewing the individual in the household whose birthday falls next. Salmon & Nichols (1983) compared four methods of selecting telephone survey respondents within a household and concluded that the next-birthday method is an efficient procedure for selecting a sample representative of all household members. Other studies, however, have criticized this (and other) methods for under-representing younger people and over-representing older people (see Hess, 1994). In this study, any bias created by this selection method is less important, since individual level data as well as aggregate data are used.

Interviewers working in some urban areas encountered a number of households whose residents had a poor ability to communicate in English. If it was clear that these individuals were not permanent residents of New Zealand, the interview was abandoned, since non-residents are not entitled to vote. However, when the respondent was a permanent resident, the interviewers attempted to complete the interview, abandoning it only if they felt continuation would seriously compromise the data quality. In other words, where interviewers were unable to obtain responses to the specified questions without paraphrasing them to assist respondents' understanding, the interview was terminated or, if it was not possible to do this without appearing rude, a note was made on the questionnaire which was then removed from further processing.

Otherwise, interviewers excluded a dwelling or its occupants from a survey only when they felt that entering would have compromised their personal safety.

6.2.2 Quotas

Within each cluster, interviewers were required to interview three men and three women and to ensure that, overall, equal numbers of men and women answered each questionnaire version. Although the research evidence has not identified a clear relationship between turnout or partisan affiliation and gender, roughly equal proportions of men and women do vote and it seemed logical that the samples reflect this distribution (see Section 6.2.3 for further discussion of this issue).

Quotas were also established to ensure that the sub-sample sizes relating to each questionnaire version were approximately equal. Interviewers thus began each cluster by asking to speak to the person within that household whose birthday fell next and changed this request to specify a particular gender only when they had completed the quota for one gender within that cluster.

6.2.3 Weighting

The results reported in Chapters 7 to 10 were calculated in three ways: estimates based on the raw unweighted data are then weighted to correspond to the age-gender distribution of the electorate; the same data are weighted by respondents' probability of voting; finally, the raw data were weighted so respondents' reported voting behaviour in the 1990 General Election matched the actual results for each electorate.

Some researchers have questioned the relationship between age or gender and voting behaviour (see Murphy, 1975; Husbands, 1987). However, many research companies in New Zealand, at least, weight polling data (and, indeed, data relating to a wide range of survey topics) to ensure the age-gender distribution of the sample matches that of the population. Thus while there was no compelling evidence to suggest that this procedure would improve the accuracy of the estimates, it nevertheless seemed both interesting and opportune to compare the results obtained from the raw and weighted data sets.[28]

However, while the literature relating demographic characteristics to partisanship reaches rather ambivalent conclusions, research exploring past behaviour as a predictor of future behaviour has produced more definite findings (see Pearl & Fairley, 1985). In particular, although debate over whether measures of actual or recalled behaviour should be used, researchers appear to accept that measures of respondents' past behaviour will bear a strong relationship to their future behaviour (see Murphy, 1975; Walter, 1987; Butler, 1994). The data were therefore weighted so that the distribution of respondents' reported voting behaviour in the 1990 General Election matched the actual voting distribution for each electorate. Since some respondents (i.e., those voting for the first time) had no prior voting history, their responses remained unweighted.

6.2.4 Error Calculation

Several researchers have criticised pollsters for failing to calculate the true sampling errors associated with their estimates (see Pearl & Fairley, 1985; Mullet & Karson, 1985; Jowell *et al*, 1993). Instead of including the effects of the clustering in error calculations and accepting that the design factor was now greater than one, many pollsters have instead calculated errors as though they had used a simple random sample[29]. Furthermore, some researchers noted that pollsters or the media regularly reported responses from sub-samples within the survey without noting that the error margins increase as the relevant sample size decreases (see Hoek & Gendall, 1996).

---

28 Voss, Gelman & King's (1995) overview of the polling practices employed by major US pollsters also includes an interesting discussion of weighting.

29 The design factor shows the extent to which a standard error associated with a particular estimate should be adjusted to be comparable with the error that would be associated with a simple random sample of the same size.

Perry (1979) raised a number of questions relating to both bias and precision and suggested that, if pollsters accounted for the increased design effect brought about by clustering, the standard errors associated with poll estimates would increase by a factor of between 30% and 40%. Likewise, Buchanan (1986) commented: "*On the basis of the long-term performance of the polling profession as a whole, press accounts of predictions should say '5 percent margin of error' where they now say '3 percent' and make comparable adjustments to all such estimates based on sample size*" (p. 225). Jowell *et al* (1993) and Lau (1994) found few pollsters had adopted these suggestions and called again for pollsters to stop treating highly clustered samples as though they were simple random samples. However, they noted that pollsters should not bear the full blame for the problems brought about by miscalculation of errors, and both articles berated the media for failing to pay sufficient attention to the error margins.

As demands for explanations of polls' inaccuracies have become more strident, so more recent researchers have devoted greater attention to how they and others calculate error margins, and whether the procedures commonly adopted adequately reflect the precision of the estimates. Collins (1988) examined the problems associated with error margins in detail and, like others cited above, argued that researchers should pay more attention to the design factors associated with their sampling. His comments on error margins, which he suggested should be widened by around 30%, corroborated Perry's earlier remarks. In addition, Collins argued that the commonly reported 3% error margin applies to estimates of each party's share, but not to estimates of one party's lead over another, where he suggested the error margin should be about ±6% (p. 9). This miscalculation of error margins led Collins to conclude that the media often report a level of certainty associated with poll estimates which the polls themselves simply cannot deliver.

6.3 Survey Instruments

Although the literature has documented several issues relating to question wording (see Chapter 3), it was necessary to select from these a limited number of factors which this research could then explore in greater detail. Since the overall objectives of the research relate to the testing of procedures which could reduce the proportion of undecided respondents, the treatments outlined in Figure 6.3a were used, and are

discussed in more detail in the remainder of this section.

Figure 4 Survey Design

Rapport Developing Questions	No Rapport Developing Questions
Traditional method	Traditional Method
Secret ballot method	Secret ballot method
Juster probability method	Juster probability method

6.3.1 Rapport Development

Throughout both the polling and question-wording literature, researchers have noted that respondents may be more forthcoming in their responses if they first have an opportunity to develop some kind of a relationship with the person interviewing them. Thus Sudman & Bradburn (1982) suggested: "*the opening questions should be easy and non-threatening and, if possible, salient to the respondents*" (p. 218).

Smith (1992) has suggested that introductory questions can do more than simply establish a rapport with respondents: "*A rapport effect argues that a more trusting and open exchange of information occurs after the interview has developed... less mentioned* ... [effects] ... *are learning effects. Learning effects suggest that respondents, in general, learn their role as respondents better as the interview unfolds and* [they] *become more familiar with response scales and other tasks.... this reduces measurement error*" (p. 177). Conversely, others have noted that the addition of these questions may bias the responses given. Thus while accepting that respondents may provide franker answers if they have some trust in the interviewer, some researchers have argued that a reduction in undecided responses *per se* does not necessarily result in an improvement in the accuracy of the survey estimates (see also Crespi, 1988).

Given the differing opinions that exist on this topic, this research also tested the extent to which rapport building or contextual questions reduced the proportion of undecided responses. Thus all treatments were tested using two groups of respondents. The first group was administered a questionnaire which began with the rapport building

questions, the second group answered a questionnaire which began directly with more personal (and arguably, more intrusive and threatening) behavioural questions. However, since information contained in these contextual questions was used to weight the data, all questionnaire versions contained the same rapport-developing questions and varied only in the order in which this information was presented.

Since some people prefer to keep private their political preferences and voting behaviour, the rapport-building questions began with very general questions exploring respondents' interest in the election, before turning to their views of the various parties' policies with respect to key economic and social issues. Questions exploring respondents' party identification, membership, registration, and past voting behaviour, then preceded the voting behaviour question.

These rapport-developing questions were gleaned both from surveys conducted by polling organisations and from practices reported in the literature (see Sudman & Bradburn, 1974; 1982; 1988), and were similar in design to many of the party identification questions asked by pollsters internationally. The social and economic factors examined were selected from lists, released by research companies, ranking New Zealanders' greatest concerns (see also Murphy, 1975). Although people's voting decisions hinge on personalities as well as on policies, it nevertheless seemed logical to examine these concerns since they were, apparently, shared by many New Zealanders and could thus help respondents to clarify their views about the different parties.

Although several researchers have noted the importance of candidate evaluations in voting decisions, it was not possible to include questions relating to the various candidates contesting each electorate since, even only three months prior to the election date, a number of the major parties had still not selected candidates in the electorates in which the research was conducted. While this information may have provided another means of classifying undecided respondents, the New Zealand electoral system has been less personality-based than in the United States, where the research documenting the importance of candidate evaluations was conducted[30].

---

30 Interestingly, Holm & Robinson (1978) reported that respondents' policy orientations bore a stronger relationship with their voting behaviour than did their candidate evaluations or party

6.3.2 Voting Intention

Many variations exist in the voting-intention questions asked, but a key issue researchers have to decide is whether to ask respondents about their past, present or future voting behaviour. In other words, respondents can be asked to consider one of at least three different questions:

1. *"If the election had been held yesterday, which party (or candidate) would you have voted for?"*

2. *"If the election were held today, which party (or candidate) would you vote for?"*

3. *"Thinking about the forthcoming election, which party (or candidate) do you plan to vote for?"*

The rationale underlying the first form, which is the question employed by many research houses in New Zealand, is that respondents will be better able to consider how they would have behaved than to project their behaviour into the future (see Murphy, 1978 for a discussion of questions employed by New Zealand pollsters). Yet although this assumption is held by a number of researchers, there appears to be no empirical support behind it. Bradburn & Mason (1964) suggested "*organisational folklore*" dominates much of the question wording literature and this phenomenon emerges frequently in the voting intention literature.

Form 2, used by some New Zealand research companies, and more widely adopted by United States' pollsters, attempts to capture respondents' current behavioural intentions. Some have argued that questions exploring retrospective behaviour may not capture shifts in opinion and intention, whereas a question that explores these issues in the present will produce more accurate responses (see Borelli, Lockerbie & Niemi, 1987).

Following these arguments further, researchers have argued that questions capturing present intentions will only reflect a snapshot in time, and perhaps cannot reasonably

---

identification.

be projected to predict the actual election outcome. Thus form 3 requires respondents to consider a specific event and their likely behaviour in relation to this event. However, Fenwick *et al* (1982) argued that questions which require respondents to project their behaviour result in higher proportions of undecided respondents than do either of the former two questions.

Since very little research has documented the relationship between the temporal location of the behaviour explored and the accuracy of responses, pollsters have little way of making empirically-based choices about which question form to use. In this research, the first form of the voting-intention question was used on the grounds that this would offer better comparisons between the results reported here and those reported by commercial polling organisations. Clearly if an alternative question had been used, different results may have been obtained. However, the decision to use the retrospective question recognised the need to be able to draw distinctions between current practice and modifications to this.

Thus the first two questionnaire versions investigated respondents' voting behaviour by asking them to consider how they would have voted, had the election been held yesterday. They differed only in the order in which interviewers administered the questions. (All questionnaires are reproduced in Appendix C.)

6.3.3 Confidentiality

In line with the New Zealand Market Research Society Code of Ethics, interviewers assured respondents that the information they provided would be treated as confidential and that no information would be reported in a way that could identify the individual who provided it. This statement of confidentiality was read after respondents' co-operation had been elicited, but prior to the commencement of the survey itself, and was repeated before administration of the demographic questions. Anonymity, however, (i.e., the absence of detail about individual respondents) presents some practical difficulties. For example, it prevents researchers from auditing the surveys and almost inevitably increases the potential for interviewer fraud. Furthermore, it obviates post-election interviews, an integral part of this study.

While it was therefore not practicable to offer respondents complete anonymity, it was

nevertheless possible to increase the level of confidentiality associated with the survey. The secret ballot technique discussed in detail in Chapters 2 and 3 enables respondents to record their views without having to state these directly to an interviewer. The research findings documented in Chapters 2 and 3 suggest that as the level of confidentiality associated with a survey increases, so the proportion of undecided responses decreases. Accordingly, a secret ballot was employed in some of the interviews to assess the extent to which it achieved such a reduction.

The questionnaire versions using the secret ballot technique all employed the retrospective voting-intention question. However, instead of administering this question directly, interviewers gave respondents a voting slip, analogous to an actual ballot paper, and asked them to tick the party for which they would have voted. Respondents then folded their voting slip and deposited it in the secret ballot box which their interviewer carried.

Like the versions involving direct administration of the retrospective voting-intention question, the secret ballot versions also differed according to whether the voting question was preceded or followed by the rapport-developing questions.

6.3.4 Relative versus Absolute Judgments

Researchers documenting undecided responses have suggested that differences exist between respondents who have no opinion on an issue, and those who have genuine difficulty in choosing between the options facing them. For example, Bogart (1967, cited earlier) suggested that the prevailing model, which assumes that people hold a single opinion, has serious flaws. Intuitively, it seems likely that the current format of voting-intention questions presents members of this latter group with some difficulty, since it requires them to make an absolute choice when, at best, they may only be able to indicate relative levels of support for the parties or candidates contesting their electorate.

In addition, a growing body of research has suggested that intention questions may elicit less accurate projections of behaviour than probability questions. Although voting-intention questions are not administered using a Likert-type scale, it is arguable that they represent an even more primitive type of intention question and that they are

analogous to the dichotomous intention questions discussed in Chapter 4.[31]

For this reason, a scale which enabled respondents to express their likely behaviour in relative rather than absolute terms was incorporated into the final two questionnaire versions. The evolution and development of the Juster Scale has been documented in Chapter 4, its final version, as shown in Figure 5 below, was used:

Figure 5 **The Juster Scale**

10	Certain, practically certain	(99 in 100)
9	Almost sure	(9 in 10)
8	Very probable	(8 in 10)
7	Probable	(7 in 10)
6	Good possibility	(6 in 10)
5	Fairly good possibility	(5 in 10)
4	Fair possibility	(4 in 10)
3	Some possibility	(3 in 10)
2	Slight possibility	(2 in 10)
1	Very slight possibility	(1 in 10)
0	No chance, almost no chance	(1 in 100)

The modified question now put to respondents: *"If the election had been held yesterday, how likely is it that you would have voted for the [A] Party?"*, was repeated to explore respondents' probability of voting for each of the four parties fielding candidates in every electorate, as well as their probability of voting for another (minor) party.

Since the Juster Scale is more complex scale than the standard intention questions, interviewers gave respondents an opportunity to familiarise themselves with it by asking them to indicate their likelihood of eating out at a restaurant in the next four weeks. Juster's work, and the extensive research undertaken by Brennan *et al* (see chapter 4) suggested that few respondents have major difficulties comprehending and

---

31 That is, the question: *"If the election had been held yesterday, which party (or candidate) would you have voted for?"* is akin to asking: *"If the election had been held yesterday, would you have voted for PARTY A...N?"* which is essentially the same as a dichotomous intention questions.

using the scale.[32] Similarly, interviewers, when debriefed, reported only a very small proportion of respondents appearing to have any difficulty using the scale.

In addition to using the Juster Scale to estimate their probability of voting for the various parties represented, all respondents used it to indicate their overall likelihood of voting. This question, designed to predict likely turnout, provided data used to weight respondents' party-support estimates and to discount the views of those respondents who indicated a lower probability of voting.

6.3.5 Demographic Questions

In addition to the opinion and behavioural information collected, respondents' demographic characteristics were also recorded. Although the research evidence relating demographic characteristics to actual voting behaviour remains somewhat contradictory, there have been some suggestions that associations exist between age and gender and partisan support, while still stronger evidence supports a link between education and voting behaviour. For these reasons, information was collected on these three demographic traits: age, education and gender.

Although age is most often collected as a categorical variable, to reduce any threat that might be associated with it and thus encourage respondents to provide the information sought, this approach reduces the data to an ordinal level. More recent research has suggested that asking respondents the year in which they were born both preserves the metric level status of the data while also being perceived as less threatening by respondents since it results in fewer refusals compared to ordinal versions of the question.

Education, on the other hand, was collected as a categorical variable and proved difficult to code as some respondents appeared to belong to more than one category. For example, some respondents who said they had left school after four years of secondary education, were now mid-way through a University degree, thus raising a

---

32 Brennan *et al* (1994) reported that although many respondents did not appear to have a sound understanding of probabilities, few experienced serious difficulties when using the Juster Scale. It should also be noted that Brennan *et al* required respondents to perform more complex and detailed tasks than was the case in this research. Moreover, despite deficiencies in respondents' understanding of probabilities, the aggregate results were generally accurate.

question about the group to which they most properly belonged. Where such anomalies emerged, respondents were placed in the higher of the two categories on the grounds that this best represented their current position. This decision also recognised that the difference between a partially and fully competed degree was less than the difference between secondary education only and a partially completed University degree.

In some cases, respondents' highest educational qualification had been obtained from a country other than New Zealand, When this occurred, interviewers were instructed to note details of the qualification and the respondents were allocated to a category later, after identification of the nearest New Zealand equivalent.

Gender was coded by the interviewers and only one interviewer reported an instance where a respondent's gender appeared indeterminate. (In this case, the interviewer's judgment was used to classify the respondent.)

6.3.6 Item Order Effects

As discussed briefly in Chapter 2, and in more detail in Chapter 3, item-order effects can bias the responses given to survey questions (Payne, 1951; Shulman, 1973; Kalton, Collins & Brook, 1978; Schuman & Presser, 1981). Bradburn & Mason suggested that researchers should pay particular attention to countering item order effects when investigating the average level of response for a particular population. For this research, interviewers rotated the order of certain response items relating to a number of questions and read response options starting at an asterisked or highlighted options. This practice ensured that each item was evenly rotated around every position in the list. While some researchers simply reverse the order of the list this does not recognise that the first and last items tend to suffer most from item-order bias; thus rotation of the list alone will not fully address this bias.[33] Details of the item-order procedures adopted can be seen in Appendix C, which contains copies of the questionnaires used.

---

[33] Although the order in which interviewers read the response options was rotated, the order of response options on the secret ballot slips remained alphabetical, since this approximated more closely the format of actual voting papers.

6.3.7 Piloting of the Survey Instruments

Since, as the research discussed in Chapter 3 makes clear, misunderstanding of survey questions can introduce significant bias, the piloting of the questionnaires took on an added importance. Traditionally, piloting is accomplished by administering the questionnaire to a small sample of people whose responses and reactions in general are then examined in more detail.

Thus each version of the questionnaire was administered to between four and six respondents who were then questioned in detail about what they understood the questions to mean. This technique, known as Belsen's Double-Back, is designed to check whether respondents' interpretation of a question concurs with what researchers intend it to mean. Although some differences in interpretation may seem trivial, researchers have noted that substantial discrepancies may exist, and that the potential for bias is thus large (see Belsen, 1968; Gendall, 1994). While all questions used in this survey were tested in this way, particular attention was paid to respondents' understanding of those questions which employed the Juster Scale, since this had not been employed in this context before and, consequently, the risk of respondent misunderstanding seemed greater.

No evidence of question misunderstanding appeared, although one of the Juster Scale response mechanisms, which asked respondents to allocate counters on a grid representing all the parties contesting the election, confused some people and was subsequently omitted from the final questionnaire versions[34].

Several researchers have argued that piloting the questionnaire on a small sample of respondents is the best method of evaluating the flaws of a particular instrument.

---

[34] This response device was included to anticipate the problem Hoek & Gendall (1993) detected, namely that those respondents answering questionnaires which employed the Juster Scale would allocate probabilities to the various parties which should, in theory, sum to their overall probability of voting, but which, in practice, did not. The method tested involved giving respondents counters corresponding to their overall probability of voting and asking them to allocate these counters among the different parties. By working with counters, respondents were forced to make trade-offs which they could avoid in purely verbal responses. However, a number of respondents found this task difficult to perform, and a correction factor, which recalculated respondents' party probabilities to sum to their probability of voting, was instead incorporated into the analysis.

However, others noted that respondents are ill-equipped to provide detailed or even very informative feedback about a questionnaire. Thus Converse & Presser (1986) noted: "*research on the pretest suggests that respondents are not very critical or sophisticated about survey questions, even when invited to be, and their counsel may not be a very good guide to practice*" (p. 53).

Hunt, Sparkman & Wilcox's (1982) research provided evidence to support this view. After exploring respondents' ability to identify five types of error deliberately incorporated into questionnaires they found respondents generally identified missing response categories though they were less able to recognise ambiguous or loaded questions. Nor did Reynolds, Diamantopoulous & Schlegelmilch's (1993) conclusions add support to the value of pre-testing questionnaires. They noted the absence of empirical guidelines for assessing pre-testing, but suggested that pre-tests should check for the presence of *double entendre*, ambiguity, and response categories, as well as for issues of relevance and attention. Although discussing both the protocol and debriefing approaches to pre-testing, Reynolds *et al* suggested neither was particularly good at detecting different types of problem.

In addition to the question of whether respondents are sufficiently knowledgeable about questionnaire design matters to offer insightful feedback, Payne (1951) queried whether researchers were in a position to collect and incorporate any suggestions respondents might make. He argued that because researchers normally cannot spend a great deal of time discussing a questionnaire with respondents, careful scrutiny of the questionnaire by themselves or their colleagues may produce more insightful suggestions: "*A critical reading of the questionnaire has been known to point out more problems than were detected in an earlier testing*" (p. 14).

As an addition, or even an alternative, to this type of pre-testing, Sudman & Bradburn (1982) described a technique they referred to as a "group mind", which employs well informed and experienced colleagues who critique the questionnaire. They noted: "*while the process is always a humbling experience for the writer, it is also a rapid, efficient method of improving the questionnaire*" (p. 283). The questionnaires used in this study were subjected to a rigorous group mind-appraisal by two colleagues who combined over 40 years' experience in survey research. Their suggestions at the early

stages of the questionnaire design process, and their comments throughout the revision process, helped ensure the versions pre-tested were close to final, and required only minor modifications.

6.4 Survey Procedure

The organisation and implementation of surveys requires researchers to make a number of administrative decisions about how the survey will be conducted. The following section outlines and explains these decisions as they affected the research described in the following chapters.

6.4.1 Interviewer Recruitment and Training

Interviewers were recruited from a variety of sources: New Zealand Employment Service (the state co-ordinator of employment); Student Job Search (a state-funded organisation which organises short term employment for tertiary students); newspaper advertising, and referrals from survey research companies. In all cases, it was specified that applicants should have had prior interviewing experience and a minimum education qualification of some tertiary study.

Applicants were interviewed by phone and were given full day training sessions prior to each round of interviewing. These sessions familiarised interviewers with the respondent selection criteria and procedures, and with the different questionnaire versions and showcards. Much of the day was taken up with role-plays, with interviewers administering the questionnaires to each other and devising strategies for dealing with problems they were likely to encounter. For example, they practised enjoining co-operation from a reluctant respondent and discussed how to differentiate between persuasion and coercion. In addition, they attempted to deceive other interviewers by giving obviously discrepant answers. This taught the interviewers to be alert to potential problems in respondents' answers and helped them to identify non-threatening and non-judgmental ways of resolving these.

Interviewers also discussed how to clarify responses which did not relate directly to any of the response categories provided. In particular, the importance of repeating the original question (rather than paraphrasing it in the interviewer's own words) was stressed. Again, training with "difficult" respondents helped to increase both the interviewers' familiarity with this situation and their ability to address it.

Interviewers were also given strict instructions to ensure that respondents' rights were both explained at the outset of the interview and protected during it. The need to maintain high ethical standards throughout was stressed.

At the end of the training session interviewers were supplied with:

36 numbered and 24 replacement questionnaires

A clipboard and pen

Showcards (relating to: interest in the election, the Juster Scale, education level)

A secret ballot box

A list of starting addresses and a map of their interview route

A letter establishing that they were a bona fide interviewer

A name tag

A full set of interviewing notes and instructions (see Appendix B)

6.4.2 Interviewer Auditing

Interviewers were advised that their work would be scrutinised and subject to audit procedures. Scrutiny of questionnaires during the data entry and verification resulted in a unique perspective on the data quality. In cases where the interview had not been properly administered (evidence of missing data from several questions without an accompanying explanation), the questionnaire was not coded. Where a pattern of mis-administration appeared, that interviewer's work was removed from the data-set pending further investigation. Two cases of interviewer fraud were detected in this way; ironically, both interviewers were currently employed by commercial research companies and the Department of Statistics.

In addition to this personal scrutiny of each questionnaire, a trained research assistant audited 15% of each interviewer's work by selecting one respondent at random from each questionnaire version and phoning that individual to audit the interview protocol. The assistant asked respondents whether they recalled participating in an interview and, depending on the questionnaire version used, then asked more specific questions about the interview itself. Although this audit exposed some minor flaws in

administration[35], it revealed no further instances of fraud. Respondents were also asked for their general impression of the survey and the interviewer; no negative comments emerged and the overwhelming majority of respondents appeared to have enjoyed the interview experience. Finally, all respondents were thanked for their co-operation with the survey and the audit.

6.4.3 Interview Timing

Several researchers have noted that the interview timing can affect the accuracy of the estimates obtained. Thus interviews conducted well before the election date typically prove less accurate, since respondents have more time to change their minds. For this reason, two rounds of interviews were conducted in each electorate: one some two to three months prior to the election date and the second within the designated campaign period, around two weeks prior to the election. The rationale for this was two-fold. First, having two measures of voting intention would allow the detection of any trends that may be present (although more measures would be required to assess the nature of the trend) and their likely effect on the election outcome. Second, any treatment showing a high level of accuracy, irrespective of timing, would appear to have a clear advantage over methods whose estimates varied considerably from the outcome. In other words, if a method was able to incorporate and account for changes in behaviour that may occur over time, it would appear to have some considerable advantages over the current methods. To examine this question, some assessment of the accuracy of the responses obtained by each version was necessary. The following section describes this assessment process.

6.4.4 Assessment of Accuracy

There are at least two means of assessing the accuracy of the survey estimates. First, the estimates can be compared to the actual election outcome, thus the predicted party support distribution can be set against the actual records pertaining to each electorate. However, this approach inevitably confounds sampling and non-sampling errors, complicating the overall assessment of the various treatments' accuracy.

The second option, that of re-interviewing some or all of the original respondents,

---

35 These minor flaws included interviewers having to return to their cars to collect their secret ballot boxes or showcards.

avoids the problem outlined above, but introduces additional difficulties such as the reliability of respondents' recall and the biases which might affect this. It is clear, even from this preliminary discussion, that neither approach to assessing the accuracy of the estimates is entirely straightforward. For this reason, both measures are included in Chapters 7 to 10 and are discussed in the light of their respective limitations. The remainder of this section outlines how the information relating to these two measures was collected.

Although it is not possible to obtain information about individuals' own voting behaviour, after each election, the Department of Justice publishes a summary containing details of the overall electorate turnout, together with a breakdown of the voting patterns across all the polling booths within each electorate. The relevant pages from this publication are included in Appendix D.

While details of aggregate behaviour are comparatively easy to obtain, the electoral and privacy legislation prevents the Department of Justice from providing details of individual respondents' behaviour. A sub-sample of respondents was therefore re-interviewed.

Although the original sample selection had, as far as was practicable, been randomised, the re-interviewed respondents were not chosen randomly. Instead, only those respondents interviewed initially by people who produced very high quality work were selected for re-interviewing. This was to ensure that variations in responses to different questionnaire versions could be attributed to the questions asked, rather than to variations in the interviewing. While it is never possible to eliminate interviewer bias (nor was it possible for the audit procedures to explore every type of bias), it is nevertheless possible to reduce it to a minimum, and the use of only the best interviewers' work represented an attempt to achieve such a reduction.

Of the 4040 respondents originally interviewed, 2402 were selected for re-interview by telephone and of these, 2147 were successfully re-interviewed, 81 refused to participate, and 174 could not be contacted (up to 15 contact attempts were made),

had shifted to an unknown address, or were ineligible for other reasons.[36] Overall, the response rate to the telephone phase was 89.4%.

The telephone interview was extremely brief. Interviewers asked whether respondents had voted and, if so, when they had decided on the party or candidate for whom they voted. Those respondents who stated they had made this decision on polling day were then asked whether they had decided prior to or after entering the polling booth. Finally, interviewers asked respondents to state the party for whom they voted.

As noted above, this procedure relies on respondents providing accurate reports of their behaviour, and a number of reasons exist why they may not do so. These reasons, together with other factors which may have affected the validity of the estimates, are discussed in the following section.

6.5 Potential Sources of Error

All survey research runs the risk of being affected by error, and the literature documents many of these errors in detail. Some issues, such as interviewer bias, can affect a wide range of surveys and constitute, in their own right, a detailed area of study (see Bradburn & Sudman, 1988, for an overview of these). Because the research into these more general errors is so prolific, this section explores only the more specific errors affecting the implementation and interpretation of this research.

6.5.1 Respondent Recall

Clearly any research which involves re-interviewing respondents to obtain measures of their actual behaviour depends upon those respondents providing truthful and accurate answers. While several researchers have noted that few respondents appear to set out deliberately to mislead interviewers, they have acknowledged that respondents may not always be in a position to tell the truth (see Bradburn & Sudman, 1988). The following sections examine this fundamental problem of respondents' ability to recall their behaviour, before turning to consider related errors.

---

36 Some respondents were ill at the time of the second interview, others were overseas, and some were deceased.

When the behaviour examined has occurred quite recently, respondents generally provide accurate recollections; however, when the behaviour has occurred some time ago, omission or telescoping effects may impair their ability to provide clear answers. Sudman & Bradburn (1974) described omission as occurring when respondents simply forget an event has happened, whereas telescoping occurs when respondents recall events as if they have occurred more recently than is actually the case. Both Sudman & Bradburn (1974) and Kalton & Schuman (1982) discussed the use of bounded and aided recall techniques which encourage respondents to construct mental time periods within which they attempt to recall specific behaviours.

Where behaviours have occurred quite infrequently (such as voting behaviour), problems of telescoping are typically less pronounced, although as the time-lag between the behaviour and the interview increases, problems of omission typically increase. This raises the possibility that some respondents may have inaccurately recalled, or failed to recall, their 1990 voting behaviour. Gauging the existence, or even the extent, of this problem remains difficult, since access to the actual records is denied. However, there is no *a priori* reason why this problem should not affect estimates associated with the various parties equally;[37] if this error is randomly distributed, its effects should cancel out. If, on the other hand, these effects are not randomly distributed, the weighting of each sub-sample by the voting distribution evident at the 1990 election should serve to reduce the scale of the problem.

In addition to the question of how researchers can define the period within which the behaviours recalled should have occurred, other researchers have examined differences between prior (or actual) past behaviour and recalled behaviour. Granberg & Holmberg (1986) specifically explored the relationship between prior behaviour and recalled behaviour as predictors of subsequent behaviour. Comparing the theories of behaviourism (where past behaviour is regarded as a strong predictor of future behaviour) and field theory (which is ahistorical, viewing behaviour as a function of respondents' life space), Granberg & Holmberg examined whether prior voting behaviour (obtained immediately after an election) or respondents' recalled behaviour

---

37 Neither the Alliance nor the NZ First parties were formally represented in the last election. Thus, technically, respondents could not have voted for these parties, and should not have recalled having done so.

(obtained 3 to 4 years after the election) best predicted their subsequent behaviour. Although their test was less concerned with predicting turnout than with specific preference patterns, Granberg & Holmberg found recalled behaviour a better predictor than prior behaviour (see also Marquette & Hinckley, 1988). However, since their experiment depended on self-reported behaviour it appears more a test of how preferences change over time than of behaviourist and field theories. Moreover, it is unclear how social desirability biases outlined by Tittle & Hill, 1967, and Silver, Anderson & Abramson (1986), would have affected the validity of this work.

These studies suggested that social desirability bias causes particular problems affecting turnout estimates. Furthermore, the biases these researchers noted appear to affect respondents' recollections of their past behaviour as well as their projections of future behaviours.

Also investigating this issue, Riedel & Dunne (1969) tested the hypothesis that people would vote in accordance with their predictions, and that the number changing their minds would be very small. Of the small sample interviewed, three quarters claimed to have voted as they had earlier said they would, and less than 5% said they had changed their minds in the polling booth. However, the reliance on self-reported rather than actual data makes it difficult to ascertain whether people's intentions did predict their behaviour accurately, or whether respondents simply wanted to appear insightful commentators on their own future actions.

Overall, dealing with self-reported data appears, inevitably, to incur a number of difficulties, as Parry & Crossley (1950) argued: "*verbal reports from the same subjects are not checks against outside data, they are as much a reliability measure as validity, and may be subject to the same kinds of inaccuracy on voting reports as the original pre-election questions*" (p. 68). The following section overviews the effects of social desirability bias alluded to above and discussed in detail in Chapter 3, and examines their specific implications for this research.

6.5.2 Social Desirability Bias

Social desirability bias occurs when respondents temper their answers to achieve some compromise between their actual behaviour and their perception of "appropriate"

behaviour. This bias can have several effects; Presser (1990) outlined the best known of these: "*One of the surest predictions in survey research is that more respondents will claim to have voted in post-election surveys than actually cast a ballot*" (p. 586). Several researchers have explored the extent to which this bias might affect self-reported data, and some have suggested that the relationship between respondents' reported and actual behaviour may be more tenuous than researchers initially thought. (For example, Parry & Crossley's 1950, work showed only one third of those interviewed correctly recalled having voted and, of these, many mis-reported the actual party or candidate for whom they voted.)

Tittle & Hill (1967) had earlier explored this problem by checking respondents' self-reported behaviour against voting records; they found 10% of those interviewed had inaccurately reported their behaviour. Attributing this to social desirability bias, Tittle & Hill suggested that removal of these mis-reporters from the sample improved the predictive validity of the attitude scales they had used.

Similarly, Silver, *et al* (1986) argued that social desirability bias presented a major problem affecting turnout calculations. After comparing claims of self-reported voting to actual voting records across four elections, Silver *et al* found between 20% to 30% of respondents mis-reported having voted when in fact they had not done so. Like Tittle & Hill, Silver *et al* concluded that this mis-reporting was due to social desirability bias. Further analysis revealed this bias bore no consistent relationship to respondents' demographic characteristics. Since mis-reporters did not appear to have similar traits, apart from their mis-reporting, Silver *et al* examined the context within which the interviews occurred and found that a reduction in mis-reporting occurred when a knowledgeable third party, such as a spouse, was present. However, this effect was not strong; thus although the presence of another person had some effect on the responses given, these effects were not sufficiently predictive of respondents' propensity to mis-report to enable the researchers to draw firm conclusions. Moreover, ensuring a knowledgeable third party presence during interviews presents researchers with some not inconsiderable practical difficulties.

While social desirability bias acts to inflate estimates of electoral participation, Adamany & Dubois (1973) documented a related but antithetical problem: under-

reporting. This phenomenon, which occurs during multiple elections, appears uniquely American. In essence, it posits that respondents who have voted in a presidential election may "forget" that, at the same time, they also voted in a Supreme Court election, supposedly because the latter has less salience. Since these situations do not arise in New Zealand elections, under-reporting appears unlikely to represent a serious source of error.

It is difficult to assess the extent to which social desirability bias may have inflated reports of past voting or estimates of future voting in this study. Since turnout at New Zealand General Elections tends to range between 80% to 90%, the potential for this bias to affect the estimates appears less than in the United States, where turnout can fall as low as 50%, depending on the election held. Overall, social desirability bias may have affected the turnout estimates reported in Chapter 7, but, although its direction is predictable, its effects are less so. The past pattern of electoral behaviour in New Zealand suggests that social desirability bias is unlikely to have seriously biased the turnout estimates, but empirical evidence to prove or disprove this hypothesis remains elusive.

Researchers have also identified a specific form of social desirability bias which particularly affects polling and post-election polls: bandwagon effects. The following section explores this effect in more depth.

6.5.3 Bandwagon Effects

Several decades ago Payne (1951) stated that: "*False assertions are made by voters and non-voters alike about having supported the candidate who won*" (p. 186; however, see also Ceci & Kain, 1982). Bandwagon effects recognise that people generally prefer to associate themselves with success rather than failure. In turn, this implies that some respondents will mis-report their own behaviour, rather than associate themselves with a "loser".

However, although more specific research into this phenomenon has occurred in recent years, researchers differ over the extent to which bandwagon effects bias poll estimates. Bradburn & Sudman (1988) suggested bandwagon effects should logically preponderate when one candidate has a substantial lead over others. Nevertheless,

they argued that, even in these circumstances, some respondents would "counter-bandwagon" to reduce the magnitude of any mandate candidates may believe voters have given them. Overall, Bradburn & Sudman concluded that the countervailing forces within electorates, and the closeness of many elections, have generated few strong bandwagon effects.

The polls taken around this election all suggested the National Party enjoyed a healthy lead over other contenders. However, the final election result remained unclear for several weeks as vote recounts took place to ascertain whether any one party had a clear mandate. The closeness of the final results suggest that bandwagon effects should have prompted few, if any, respondents to misreport their behaviour. Thus, according to Bradburn & Sudman's (1988) logic, the stimulus to mis-report should have been noticeably absent from this election. Yet while logic can suggest an expected outcome, there remains a paucity of empirical research to support this.

These sections have examined how internal factors might prejudice respondents' answers. Unfortunately, several external factors constitute other sources of error: the following section examines some of these and their effects.

6.5.4 Effects of Interviewing and Poll Publication on Subsequent Behaviour

Many studies predicting either turnout and party or candidate support appear to assume that the act of interviewing is a neutral process which leaves respondents unchanged by the experience. However, recent studies have suggested that this may not be the case, and there is growing evidence to suggest that people interviewed in pre-election studies have a higher probability of voting than those who are not. Thus Bogart (1967) asserted: *"The paradox of scientific method is that we change phenomena by measuring them"* (p. 335). In addition, researchers have suggested that the mere publication of poll results influences people's voting behaviour, a claim that has prompted widespread discussion about whether polls should be banned during the period immediately prior to an election (see Brudney, 1982; Skalaban, 1988; West, 1994). This section reviews the research relating to these, and other, claims about the potentially biasing effects of pre-election surveys.

Two decades ago Yalch (1976) asserted that people interviewed in pre-election polls

"cannot be considered equivalent to those not interviewed no matter how carefully the sample is selected" (p. 331; see also Clausen, 1968). He argued that the experience of being interviewed altered respondents' perceptions and knowledge and, consequently, influenced their subsequent behaviour. Kraut & McConahay (1973) had earlier argued that groups interviewed prior to an election had a much higher turnout rate (48%) than a matched control group (21%). Yalch sought to extend these findings by exploring whether these effects were observed across other elections, and by examining possible reasons for the difference. He also found that pre-election interviews stimulated turnout, but that the effect of this decayed rapidly, leading him to conclude that the interviews enhanced respondents' interest in and knowledge about the election, thus prompting their increased participation in the process.

While this explanation remains speculative, subsequent research by Greenwald, Carnot, Beach & Young (1987) generally confirmed Yalch's and Kraut & McConahay's conclusions. Greenwald *et al* investigated Sherman's (1963) notion that people will act in accordance with their predicted behaviour, even when these predictions grossly overestimate their initial likelihood of performing that behaviour. They tested this idea by interviewing students who had not yet registered to vote in an up-coming election, asking half whether they intended to register, and comparing their behaviour to that of the control group. Although the probability of registering to vote was much higher among the test group, actual registration was not significantly greater than that displayed by the control group. (See also Morwitz, Johnson & Schmittlein (1993) who, after examining Sherman's theory, reached the same conclusions.)

Bolstein (1991) also addressed this question by comparing the voting likelihood of pre-election poll respondents and non-respondents, and concluded that survey respondents and refusers both had a significantly higher probability of voting than did members of the group who could not be contacted. This latter group, in turn, was more likely to vote than those sample members who did not have a phone. These findings thus confirmed those advanced previously, though they also raise the possibility that even the opportunity to be interviewed may relate to people's turnout behaviour.

In summary, there appears to be a clear view emerging that people are affected by an interview experience. However, the exact process whereby this effect occurs, together with its precise consequences, is less clear. What is clear, though, is the possibility that bias may occur, especially in turnout estimates, which appear particularly vulnerable to the effects of interviewing.

Just as interviewing may affect respondents' behaviour, so too may its consequences, namely the publication of poll results. It is not within the scope of this thesis to detail the effects of outside sources on voters' behaviour. However, several European countries already ban the publication of poll results during the campaign run-up period and such restrictions have also been recently mooted in the United States, United Kingdom and New Zealand. The remainder of this section briefly overviews the key elements of this debate.

Ceci & Kain (1982) devised an elaborate experiment in which a sample of students was divided into different treatment groups, each of which was given differing poll information before details of group members' likely voting behaviour were elicited. Although they noted substantial shifts in opinion among those groups given inconsistent information, Ceci & Kain argued that these shifts were less a consequence of respondents modifying their original position than a movement away from the dominant candidate, whoever that was reported to be. However, these shifts were less marked among respondents who had earlier expressed strong preferences, leaving Ceci & Kain to conclude that weakly decided and undecided voters were most affected by external information (see also Shaiko, Dwyre, O'Gorman, Stonecash & Vike, 1991).

Although Ceci & Kain's original hypotheses concerned the effect of early televised results (a problem in countries with more than one time zone), Bowden (1987) took a broader view and claimed: "*The application of statistics to human affairs is in many instances compounded by the possibility that one is estimating parameters whose values may in turn depend on the published estimate, perhaps negating that estimate*" (p. 476). He argued that the publication of poll results was analogous to Heisenberg's uncertainty principle: that in observing a system, that system becomes changed. This suggests an ongoing spiral of change, where every estimate effectively invalidates itself once published (See also Skalaban, 1988; West, 1994).

Major challenges thus exist to the view of polls as unbiasing either in their construction or their ultimate effects. However, the extent to which these factors affect respondents remains difficult to generalise, and researchers are not yet in a position to estimate either the precise effects of an interview or the consequences of publication of poll results on respondents.

These errors also prove difficult to quantify in terms of this research, although Bradburn & Sudman's (1988) comments on countervailing forces raise the possibility that, for every undecided respondent who decides to support the party leading the polls, another opts to support an underdog. While perhaps it may be too optimistic to believe a complete neutralising occurs, it is nevertheless possible that this source of error may be less serious than some studies have suggested. Examination of the extent to which respondents interviewed several months prior to the election did not act in accordance with their intentions compared to those interviewed during the campaign period may offer some insights into the effects poll publication and other campaign activities may have had.[38]

6.5.5 Non-Response Error

The problem of non-response affects all survey types and has received detailed attention in the literature; clearly, any survey that achieves less than a 100% response rate may produce estimates affected by non-response error. While researchers have noted differences between early and late responders to a survey, the implications of these differences have been less obvious. The more important question, of the effect non-responders would have had on the survey estimates, has not received such detailed attention.

Hosie (1995) concluded that a 50% response rate could reasonably be seen as a minimum acceptable response rate, although her conclusion requires replication before it can be widely accepted. Bradburn (1992) also commented on non-response and suggested that its effects may not be as serious as sometimes thought, although it is only possible to establish the effects retrospectively, once predictions of behaviour have been compared to actual outcomes.

---

38 However, since more than one source of error may have affected the estimates at any one time, extracting the specific effects of these remains difficult.

The response rate for this study comfortably exceeded that proposed by Hosie as a minimum, however, since the validity of her conclusion requires further assessment, this fact may mean very little. Ultimately, until comparisons between a full sample and the sample achieved after the nominated number of callbacks is possible, the extent of non-response error will remain unknown. For this reason, its effects cannot be dismissed in this survey, although, since these cannot be quantified, it is not analysed in detail.

6.6 Summary

The chief potential errors affecting the research appear to be social desirability bias and the effects of interviewing, although neither of these appears likely to have had a profound effect on the estimates. The secret ballot method addresses the former problem; logically, if this bias were a serious problems, secret ballot estimates would prove consistently more accurate than those obtained via other methods. Likewise, if interviewing did affect respondents in the manner described above, the actual turnout among respondents should exceed that of the general population. It is more difficult to determine the effects of interviewing on levels of party support, since the research has tended to concentrate on turnout behaviour. A more likely problem is that poll results influenced undecided respondents, though whether this influence took the form of a bandwagon or counter-bandwagon effect remains difficult to predict. However, despite the lack of specific effects to note, the following chapter attempts to recognise and comment on these errors where they appear to have affected the estimates.

In summary, this chapter has outlined and discussed the various decisions made in the design and implementation of the research reported in Chapters 7 to 10 - decisions ranging from the sample selection procedures to the questionnaire design and interviewer training. In addition, it has noted the potential errors likely to have affected this research, and examined the extent to which they appear likely to present problems. The following chapters report on the results of this extensive design and implementation process.

CHAPTER 7

TURNOUT PREDICTION: RESULTS AND DISCUSSION

7.1 Introduction

This chapter reports the results of the research undertaken to investigate turnout prediction, and thus addresses the first objective outlined in Chapter 1, and re-stated in Chapter 5.

Earlier chapters examined whether polling inaccuracies result from pollsters' failure to differentiate between those respondents likely to cast their vote and those less likely to do so. Researchers have suggested that estimates based on samples of the general public (rather than on samples of likely voters) may be misleading if respondents who have a low probability of voting also differ in their voting intentions. Conversely, the potential for error would be small if both likely and unlikely voters shared similar intentions.

Whether error does exist can only be determined if researchers estimate respondents' likelihood of voting. The likelihood estimate can then be used to weight respondents' stated voting intention so that the final party support estimates reflect variations in the probability that respondents will vote.[39]

However, weighting in this way is only likely to improve estimates if the methods of predicting turnout are accurate. Although researchers have developed methods of modelling turnout, the techniques used are cumbersome, and do not appear to have been widely adopted by research companies. The challenge remains, therefore, to investigate whether a less complex, but still accurate, method of predicting voter turnout may be developed. The following sections in this chapter outline the results of one such attempt and discuss the implications these have for polling researchers.

---

39 For example, if supporters of Party A had a 90% probability of voting and supporters of Party B had a 75% probability of voting, predictions which did not take this variation in voting probabilities into account appear likely to be biased.

7.2 Turnout Prediction

Attempts to model and predict turnout have employed a variety of attitudinal and behavioural questions to develop a turnout scale. The results reported here are based on a single question which used a probability scale, the Juster Scale, to estimate the overall turnout for each electorate.

The validity of this alternative method of predicting turnout was assessed in two ways. First, respondents' estimates of their likely behaviour were compared to the actual outcome for each electorate surveyed. However, since this comparison introduces sampling error, a subset of the total sample was also examined. This subset of data contains details of respondents' predicted behaviour and their subsequent self-reported behaviour. While this approach eliminated the problem of sampling error which affects the first comparisons, it introduced the potential for other errors, such as social desirability bias, to affect the estimates. Neither data set thus provided comparisons uncomplicated by other factors, and each is interpreted in the light of the problems outlined above.

First, the Juster Scale turnout estimates from each survey were compared to the actual electoral turnout; these estimates, based on the total samples interviewed, are contained in Table 1.

Table 1: Juster Predictions and Actual Turnout

Electorate	Juster Estimate[1]	Standard Errors[2]	Actual Turnout[3]	Prediction Error
	%		%	%
Manawatu 1 (N=617)	86.6	1.75	85.7	+0.9
Miramar 1 (N=614)	84.4	2.02	88.4	-4.0
Western Hutt 1 (n=782)	87.1	1.57	88.6	-1.5
Total estimated turnout[4]	***86.0***	***1.24***	***87.6***	***-1.6***
Manawatu 2 (N=731)	89.7	1.50	85.7	+4.0[5]
Miramar 2 (N=665)	88.2	1.65	88.4	-0.2
Western Hutt 2 (N=753)	88.8	1.55	88.6	+0.2
Total estimated turnout[4]	***88.9***	***1.09***	***87.6***	***+1.3***

1. Estimates calculated across all six questionnaire versions tested within each electorate.

2. The standard errors were calculated using a design effect of 2.83. Harris (1978) and Koerner (1980) suggested the design effect is approximately given by:

$$\text{Deff.} = (1 + (b-1)\,p)$$

where b = the number of second stage sampling units (households) selected from each primary stage sampling unit.

p = the intra-class correlation co-efficient (rho), which measures the degree of homogeneity within primary sampling units, (Harris, p.13).

p increases as the members of each cluster become more homogeneous, however, as cluster size increases, homogeneity usually becomes smaller, leading to a smaller design effect.

Harris reported rho values of +0.08 and +0.06 associated with voting intentions for the Labour and Conservative parties in a pre-election poll which used clusters of 30. However, this study used clusters of six adjacent households, suggesting that rho is likely to be greater than the values Harris reported. Alsagoff, Esslemont & Gendall (1986) reported rho of 0.366 and a design effect of 2.60 associated with average cluster sizes of just over five. Further, they noted that rho virtually halved as the cluster size doubled. Harris' results, although they relate specifically to voting intention, appear less relevant to this situation because of the large differences in cluster size. Although Alsagoff *et al* examined the design effects associated with income variables, these are assumed to be analogous to voting intention variables for the purposes of calculating the design effect. Thus:

$$\begin{aligned}\text{Deff.} &= (1+ (b-1)\,p)\\ &= (1+ (5 * 0.366)\\ &= 2.83, \text{ giving a design factor of } 1.68.\end{aligned}$$

The standard errors reported in Tables 1 to 4 were calculated using the adjusted sample sizes i.e., the original sample/design effect.

3. Details of actual turnout were obtained from The General Election and Electoral Referendum 1993, published by the Department of Justice.

4. These estimates were produced by combining the overall probability of voting from each electorate and then comparing this to the total turnout across all electorates.

5. Probability of accepting the null hypothesis when it is false exceeds 5%.

Most of the estimates in Table 1 were within one standard error of the actual turnout and only one of the differences between estimated and actual turnout was significant at the 5% level. That is, the probability of accepting the null hypothesis (that the Juster Scale is not a good predictor of turnout) when it is false, is less than 5%.[40]

Two estimates had prediction errors of 4%, which were considerably larger than the other errors recorded. That these prediction errors occurred in different electorates, in different time periods, and in different directions, suggests they are more likely random occurrences than the result of a particular bias. Further evidence of this appears in the aggregate estimates (in italics) which suggest the individual prediction errors effectively cancel each other out. In other words, the Juster estimates showed no evidence of either a positive or a negative bias. The two overall estimates had prediction errors of around 1.5% and both estimates included the population value within their 95% confidence intervals. Overall, therefore, this table suggests that the Juster Scale has produced some very accurate estimates of turnout.

A possible explanation of the larger prediction errors observed for two estimates in Table 1 is sampling error. That is, differences between the samples on which the Juster predictions were made and the composition (and voting behaviour) of the total electorate. Thus Table 2 represents an alternative method of assessing the accuracy of the turnout estimates and compares respondents' predicted and self-reported behaviour.

---

40 More explicitly, the null hypothesis and the alternative hypothesis are as follows:

H_0 The difference between the Juster Scale based estimates of turnout and actual or self-reported turnout is not zero.

H_a The difference between the Juster Scale based estimates of turnout and actual or self-reported turnout is zero.

Table 2: Predicted and Self-Reported Turnout Behaviour[1]

Electorate	Predicted Behaviour	Standard Error	Self-Reported Behaviour	Prediction Error
	%		%	%
Manawatu 1 (n=300)	87.9	2.4	90.0	+2.1
Miramar 1 (n=258)	86.6	2.7	89.9	+3.2
Western Hutt 1 (n=458)	88.2	2.0	93.9	+4.7
Manawatu 2 (n=384)	91.5	1.8	94.0	+2.5
Miramar 2 (n=296)	89.1	2.2	93.2	+4.1
Western Hutt 2 (n=450)	90.3	1.8	93.1	+2.8

1. Estimates based on a sub-sample of respondents who participated in both a pre-election interview, which obtained details of their predicted behaviour, and a post-election survey, which explored their reported voting behaviour. Since no means of verifying respondents' claimed behaviour exist, the data are presented as respondents' *self-reported* behaviour rather than as their *actual* behaviour.

Table 2 also suggests that, when compared to reports of respondents' self-reported behaviour, the Juster Scale based estimates are reliable predictors. In only one case did the 95%confidence interval around an estimate of turnout not include the turnout figure based on respondents' self-reported behaviour. (In other words, the probability of accepting the null hypothesis when it was false exceeded 5% for this estimate.) All other errors lay between 2% to 4% and were within the estimated confidence intervals. The overall consistency between respondents' predicted and reported behaviour suggests that Juster Scale based estimates of what respondents say they will do bear a strong relationship to their self-reported behaviour.

Closer examination of the error reveals a much clearer trend as, unlike the errors displayed in Table 1, the errors contained in Table 2 occurred in one direction only. Such a consistent pattern suggests the presence of a uniform bias, rather than the effects of random error fluctuations, which could reasonably be expected to cancel out. Although the literature contains reports of many biases, social desirability bias seems most likely to have affected estimates of this variable for reasons explored more fully in section 7.3.2.

The estimates contained in Tables 1 and 2 were based on the total samples

interviewed or re-interviewed in each electorate. As well as examining the Juster Scale's ability to predict behaviour within each electorate, this research also explored whether the different questionnaire versions had any effect on the accuracy of the turnout estimates. Juster's initial work suggested that respondents will give more accurate predictions of their likely purchases of certain items if they first answer questions relating to their financial situation and current ownership or use of the goods under investigation. For this reason, the turnout results were also analysed according to whether the turnout question had been preceded by contextual questions, analogous to those Juster used when investigating purchases of consumer durables. In addition to this distinction, the results from versions 5 and 6 of the questionnaire are also presented separately to test whether more extensive experience with the probability scale enabled respondents to project their turnout behaviour more accurately[41]. Table 3 contains these results as they relate to the total sample.

---

41 Versions 5 and 6 of the questionnaire used the Juster Scale to elicit respondents' probability of voting for each of the parties represented before using it to obtain their overall probability of voting. Thus respondents administered either of these versions would have had more extensive experience of the Juster Scale before they declared their probability of voting (see Appendix C for copies of the questionnaires).

Table 3: Context Effects on Turnout Predictions

Context	Predicted Turnout	Standard Errors	Actual Turnout[1]	Prediction Error
	%		%	%
Contextual questions (N=1392)	87.8	1.16	87.6	+0.2
No contextual questions (N=1399)	86.6	1.18	87.6	-1.0
Contextual and Juster questions (N=690)	88.6	1.54	87.6	+1.0
No contextual but Juster questions (N=678)	88.1	1.57	87.6	+0.5

1. Since these predictions cover all three electorates, the actual turnout figure represents the sum of the total number of votes cast in each electorate divided by the sum of the total number of eligible voters in each electorate:

$$\frac{\Sigma \text{ (votes cast in each electorate)}}{\Sigma \text{ (eligible voters in each electorate)}}$$

As Table 3 shows, all estimates were within one standard error of the estimated turnout value. However, while the accuracy of the estimates was obvious, the effects of the different questionnaire versions were less immediately apparent. Although, as expected, the questionnaire version with the fewest contextual questions had the highest error, so too did the version with the greatest number of contextual questions. (However, since none of the differences between the turnout estimates was greater than one standard error, there is no evidence that the differences resulted from anything other than random sampling error.)

This suggests that neither the contextual questions nor additional experience with the Juster Scale enhanced respondents' ability to project their behaviour accurately. This finding is not consistent with the question wording literature, which suggests that contextual questions and greater experience with the scale should improve the accuracy of the estimates. However, research exploring contextual questions in opinion polls has noted that their effects may be more ambiguous than researchers have previously thought (see Crespi, 1988). Other factors which might explain this

apparent anomaly are discussed below.

First, the two questions exploring respondents' voting registration and past voting behaviour, which preceded the voting intention and turnout questions, even in the non-contextual versions, may have provided a context which enhanced respondents' predictions. That is, even those versions designed to have virtually no preliminary questions preceding the voting intention question may, in fact, have developed a sufficient context to dilute the effect of the test.

As noted, the questions preceding the voting intention question in the non-contextual versions both explored aspects of respondents' voting behaviour: their voting registration status and their past voting behaviour. These two questions were apparently sufficient to create similar results to the full contextual versions. Ultimately this finding suggests that questions relating to respondents' previous behaviour and their ability to perform the behaviour in question, may be just as effective at developing a context as the series of opinion and attitudinal questions included in the contextual versions. A more explicit comparison of the effects of contextual questions would thus have employed a questionnaire beginning directly with the voting turnout question. However, since the literature has clearly documented the higher refusal rates associated with questionnaires that open with sensitive questions, this option was not explored.

The explanations offered above remain speculative, and can only be determined by further experimentation. The main conclusion that may be drawn from Table 3 is that the Juster Scale provided consistently accurate predictions, irrespective of its position within the different questionnaires. However, since a large proportion of the New Zealand population do vote, it is important to balance this conclusion against the fact that the potential for error was small.

In addition to comparing the various predictions elicited via the different questionnaires to actual turnout, the effect of the different contexts on respondents' self-reported behaviour was also explored, and can be seen in Table 4.

Table 4: Context Effects on Predicted and Self-Reported Turnout Behaviour

Context	Predicted Behaviour	Standard Errors	Self-Reported Behaviour	Error
	%		%	%
Contextual questions (N=716)	88.7	1.52	91.3	+2.6
No contextual questions (N=715)	87.7	1.57	92.2	+4.5[1]
Contextual and Juster questions (N=360)	91.4	1.78	95.0	+3.6[1]
No contextual but Juster questions (N=355)	89.6	1.98	93.8	+4.2[1]

1. Probability of accepting the null hypothesis when it is false exceeds 5%.

Like Table 3, the lowest error in Table 4 was also associated with the version containing contextual questions, and only the Juster Scale turnout question. The questionnaires containing several Juster Scale questions had similar error levels, while the version with no contextual questions had the highest error.

Looking at both Tables 3 and 4, the lowest errors were associated with the questionnaire version containing attitudinal and opinion based contextual questions. This suggests the use of a context may have had some effect.

The two questionnaires which employed several Juster Scale questions displayed the same pattern: the version with the greatest number of contextual questions had the lowest error. However, if contextual questions improved the accuracy of turnout predictions, it would be logical for the version which contained both contextual and Juster Scale questions to have the lowest error. These results did not support this expectation. The questionnaire version which should, according to this reasoning, have provided the most accurate estimates actually resulted in an estimate which was more than two standard errors less than respondents' self-reported behaviour. For both this version, and the two non-contextual questionnaire versions, the probability of accepting the null hypothesis when it was false exceeded 5%.

This finding suggests contextual questions have a more ambiguous effect than first thought. In other words, having answered a number of questions about their likelihood of voting for various parties, respondents may have been less able, rather than more able, to correctly predict their likelihood of actually voting. Why this should be so is difficult to discern from these results, although it is possible that some respondents felt they could not allocate a lower probability to their overall likelihood of voting than they had allocated to the party questions that preceded this question.

While the effects of the first interview on responses to the second require further exploration, it is equally possible that social desirability bias contributed to the error patterns observed. Like Table 2, the errors in Table 4 also show that respondents consistently underestimated their reported behaviour. Although difficult to ascertain the exact source of this error, the literature suggests the "real" error is likely to be lower than the figures reported in both Tables 2 and 4 because of the effects of social desirability bias, discussed further in section 7.3.2.

Overall, Table 4 also suggests that neither the inclusion of contextual questions nor increased exposure to and use of the Juster Scale appears to have improved respondents' predictions of their turnout behaviour. Yet, although neither Table 3 nor Table 4 shows a clear association between the use of contextual questions and improved accuracy, they both suggest that the more limited the range of contextual questions used, the greater the corresponding error level. However, the lack of clear difference between the estimates resulting from the different treatments means that, while the Juster Scale itself has produced accurate turnout estimates, the best context in which to embed the turnout question remains more ambiguous.

7.3 Relationship with the Literature

The results reported in this chapter raise two issues: the accuracy and reliability of the Juster Scale as a predictor of turnout and, related to this, the effects of bias on the results. The following sub-sections examine the relationship between these issues and the research findings discussed in earlier chapters.

7.3.1 Accuracy of the Juster Scale

For several decades researchers have acknowledged the importance of predicting turnout and, more specifically, of identifying those respondents likely to cast a vote. More recently, several British researchers have implicated pollsters' failure to predict turnout accurately as a causal factor in their failure to predict the overall election outcome (see Collins, 1988; Jowell *et al*, 1993).

The turnout scales proposed by Perry (1979) and developed further by Traugott & Tucker (1984) appear to have provided accurate estimates of turnout, but both require the construction of an index based on up to nine variables. Thus neither of these methodologies combines simplicity, parsimony and accuracy.

The Juster Scale estimates presented in the preceding sections meet all three criteria; the scale therefore appears to provide both a viable alternative to methods previously proposed and an easily administered option for researchers not currently addressing this problem. However, it is also important to note that, in New Zealand at least, the great majority of the voting age population do turn out to vote, thus the potential for error is much lower than if the predictions were for behaviours performed by a smaller proportion of the population.

Clawson (1971) had earlier noted that the higher the level of prediction, the lower the error associated with that estimate. Likewise Gruber (1970) had suggested that probability scales may be of greatest value where the proportion of non-intenders is large. Testing the accuracy of the Juster Scale as a predictor of voting behaviour in New Zealand may not, therefore, constitute a particularly rigorous test of its accuracy. A more compelling test would occur if it was used to predict turnout in a situation which typically recorded lower levels of voter participation, such as referenda. Given the range of elections that occur in the United States, and the variation in participation in these, a series of experiments conducted during a sample of these elections would provide a better insight into the accuracy of the Juster Scale as a predictor of voter turnout.

While the aggregate estimates of turnout are accurate, comparison of respondents' Juster scores and self-reported behaviour reveals a consistent pattern of

underestimation which merits closer scrutiny. Figure 1 shows the allocation of Juster probabilities given by respondents who reported having cast their vote.

Figure 6: Proportion of Self-Reported Voters by Juster Scale Values

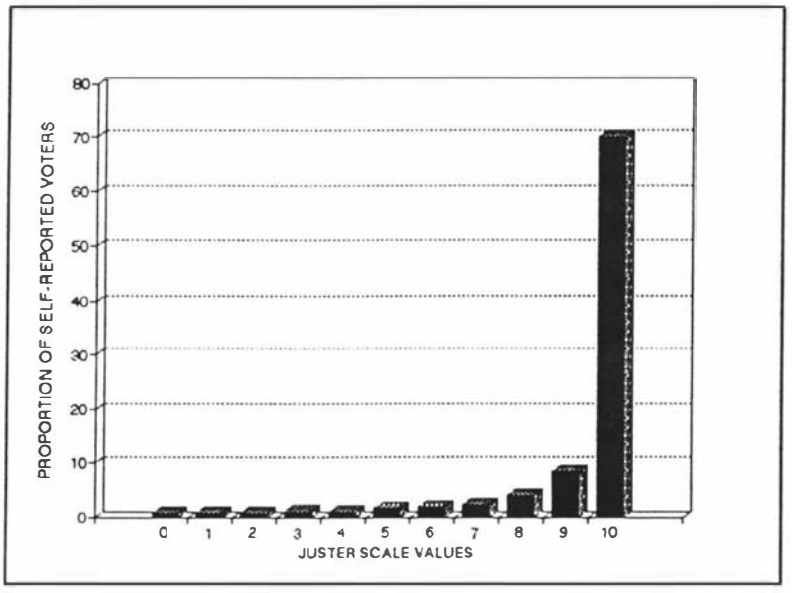
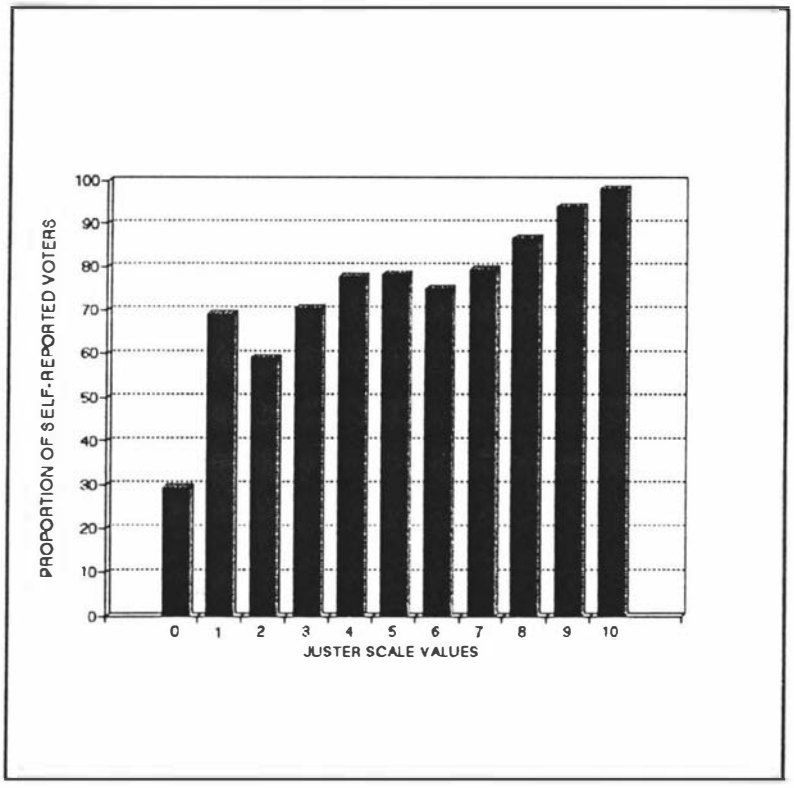


Figure 6 shows a clear trend across the probabilities: as the probability of voting given by respondents increased, so too did the proportion of respondents who reported voting. For example, 70% of those who voted gave a Juster probability of 10, around 8% gave a value of 9, 5% gave a value of 8, and so on. This finding is consistent with those reported by Juster (1966); Gabor & Granger (1972), and Pickering & Isherwood (1974).

Because so many eligible voters do go on to vote, the shape of the overall curve is more like a true "J" than the inverse "J" Juster described. However, since this difference arises simply from the behaviour examined (rather than from some characteristic of the scale itself) it is not discussed further here.

Juster (1966) had also noted that estimates from his scale tended to be biased downward, though, in this study, the social desirability bias, alluded to earlier, may have exaggerated any latent tendency the scale has toward underestimation. Examination of the pattern of respondents' voting behaviour by their Juster probabilities provides further evidence of underestimation, as Figure 7 shows.

Figure 7: Proportion of Respondents Nominating each Juster Value by Self-Reported Behaviour



This figure suggests that, with the exception of those respondents who allocated a 10 to their likelihood of voting, all other groups underestimated their likelihood of casting a vote. For example, 92% of those who gave a Juster value of 9 actually voted (or said they did).[42] However, since the overall proportion of respondents who allocated a score lower than 9 to their probability of voting was only approximately 20%, the potential for this underestimation to create a large error is low (as the small differences between predicted and actual or self-reported turnout confirm). Several earlier researchers had commented on this pattern of underestimation (see Juster, 1966; Gabor & Granger, 1972). However, the reasons advanced (difficulty in confining predictions to a particular time period, Pickering & Isherwood, 1974; changes in personal circumstances, Pickering, 1975) do not appear as relevant in this context.

---

42 While these results suggest a trend of under-estimation (i.e., a greater number of respondents voted than was predicted), it is also possible that social desirability bias prompted respondents to over-report their voting behaviour. In other words, it is possible that the scale over-estimated respondents' behaviour, but that social desirability bias masked the effects of this.

Juster argued that fluctuations in individuals' scores were unimportant since the estimates sought related to aggregate behaviour, and variations between what individuals said and did would cancel out. Comparison of the predicted and actual or self-reported estimates in this study supports Juster's assertion.

While these findings confirm earlier experiences with the Juster Scale, they also raise questions which must be addressed before the Juster Scale can be widely recommended as a turnout prediction instrument. The first, and most important of these is replication, since before any research finding can be generalised, it must first be tested, and proven, in a variety of situations. Clearly the Juster Scale has provided very accurate estimates of turnout in this study and has done so in five out of six distinct tests (see Table 1). Yet all test situations, or electorates, were characterised by high turnout, thus reducing the potential for error. Only tests of the scale in more diverse electoral environments will fully ascertain its reliability.

The second question also relates to assessment of the scale's accuracy. The results reported here examined accuracy by comparing the estimates either to the actual population value (where this was known) or to estimates of respondents' self-reported behaviour. While both of these comparisons show that the Juster Scale generally provided very accurate estimates, comparisons against the latter outcome variable suggest social desirability bias affected respondents' reported voting behaviour. Both Tables 2 and 4 show that self-reports consistently exceeded respondents' predicted behaviour. The presence of social desirability bias will exaggerate the level of underestimation, though the extent to which this might have occurred cannot be determined precisely from these data. The following section explores this problem in greater detail.

7.3.2 Effects of Social Desirability Bias

For some time, researchers have noted that respondents may not provide true reports of their own behaviour when asked what they did in a given situation. Such mis-reporting is particularly apparent if the behaviour in question is in some way socially desirable. All the research evidence relating to social desirability bias suggests that it acts to inflate the estimates of "approved" behaviours while decreasing the proportion of people who admit to performing certain misdemeanours. Sudman & Bradburn (1974) described this phenomenon as a conflict between being a good

respondent and telling the truth, and appearing to the interviewer to have behaved in a socially desirable or responsible way. Ultimately, researchers can only determine the extent of this bias by comparing respondents' self-reported behaviour to a known value. Unfortunately, comparisons such as this are not possible in New Zealand. As a consequence, this research relied on self-reported behaviour estimates, and, in so doing, almost certainly introduced social desirability bias.

The literature records several instances where respondents have over-reported their actual voting behaviour, and researchers have noted that reports of electoral participation appear especially vulnerable to the effects of social desirability bias (see Parry & Crossley, 1950; Bradburn *et al*, 1978; Kalton & Schuman, 1982, for example). The findings reported here appear consistent with these earlier reports. In every case, respondents' self-reported behaviour exceeded both their predicted turnout and the actual turnout figures. The probability of all six estimates (in Table 2) or all four estimates (in Table 4) being over-estimates are .0156 and .0625, respectively.

Since the election around which these interviews occurred had an extremely close result, it is possible that respondents were more susceptible to social desirability bias than would have been the case had the election had a more clear-cut outcome. This may provide an additional motive for over-reporting actual turnout, but it does not help quantify the effects of the bias. However, it seems reasonable to propose that respondents' actual behaviour is almost certainly lower than their self-reported turnout, and consequently that the Juster Scale estimates are more accurate than the errors in Tables 2 and 4 suggest.

Despite evidence in the literature that developing of a rapport with respondents, via contextual questions, may reduce any perceived need on respondents' part to appear to have acted in a socially desirable manner, these results suggest the effects of contextual questions are more ambiguous. Smith (1992) acknowledged that rapport developing questions could have contradictory effects. He noted that while they may help respondents to provide more accurate estimates, they may also stimulate respondents' interest in the topic and thus lead them to exaggerate their behaviour.[43] While it is not possible from these results alone to determine how social desirability

---

[43] Crespi (1980) also noted pollsters' ambivalent views on the effects of contextual questions.

might have worked, or even whether it affected these estimates, the phenomenon Smith described could have affected the estimates derived from the two Juster-based questionnaires. Other things being equal, these questionnaires should have produced estimates at least as accurate as the traditional contextual versions. That the estimates of turnout increased as the number of contextual questions increased suggests respondents' interest in the topic may have been stimulated by earlier questions.[44] However, these interpretations remain speculative.

Overall, social desirability bias appears to have led respondents to inflate their reported voting behaviour. Consequently, the "real" errors associated with the estimates in Tables 2 and 4 are likely to be lower than those reported. However, in the absence of actual voting records, the precise level of bias remains unknown.

7.4 Summary

The previous sections have examined the accuracy of the Juster Scale when used to predict voter turnout. When compared to either the actual electorate turnout or to respondents' self-reported behaviour, the Juster Scale-based predictions were consistently accurate. However, since turnout is traditionally high in New Zealand, the task of estimating the proportion of respondents likely to vote is more straightforward than where the behaviour of interest is performed by a smaller proportion of the population. For this reason, the Juster Scale should also undergo testing in contexts where turnout is much lower than 80% of the eligible electorate.

Examination of the effects of the various questionnaire contexts on turnout estimates suggested that those versions with fewer contextual questions resulted in the highest error level. Among those versions associated with lower error levels, no consistent pattern pointing to the superior accuracy of one version appeared. Thus it seems likely that all versions created a context of varying length and detail, and that the comparisons made were consequently between contexts of varying lengths rather than between the presence and absence of a context.

---

[44] Both Tables 3 and 4 showed the questionnaire version which produced the highest turnout estimate contained both Juster and contextual questions. The next highest estimate was associated with the Juster version, then the direct contextual version, and, finally, the non-contextual version.

All tables based on respondents' self-reported behaviour suggested evidence of bias. If, as suggested, this bias arises from social desirability pressures, it appears likely to have exaggerated the difference between respondents' predicted and actual behaviour, and thus inflated the errors associated with these estimates.

Overall, these results suggest that the Juster Scale can provide accurate turnout estimates without the more cumbersome analyses associated with other methods. However, although these results represent six discrete tests of the Juster Scale, the behaviour examined may well have lent itself to more straightforward prediction. The following sections, which examine predictions of party support thus constitute a more rigorous test of the Juster Scale's performance.

CHAPTER 8

CONTEXT EFFECTS ON PREDICTED VOTING BEHAVIOUR: RESULTS AND DISCUSSION

8.1 Introduction

As well as predicting which respondents will vote, researchers must also predict the party for whom these respondents will vote. Chapter 2 discussed some of the explanations put forward to account for discrepancies between predicted and actual election outcomes, and noted the increasing emphasis placed on undecided respondents. Many possible explanations for undecided responses exist. Recently researchers have explored the effects of questionnaire design and question wording on respondents' predicted behaviour, with the aim of assessing how survey instruments may assist respondents to clarify their likely behaviour. This work has examined issues such as respondents' understanding of the survey questions, their familiarity with the behaviour explored, and their overall knowledge of the topic under investigation. In general, this line of enquiry has pointed to the importance of establishing a relationship with respondents and has highlighted the role of questionnaire design in this process. More specifically, researchers have noted the importance of the number and sequence of questions within a questionnaire, and have suggested that these may provide a context which helps respondents to recall events or circumstances that, in turn, enhance their ability to predict their own behaviour.

While this argument has an intuitive logic, the effects of contextual questions remain more ambiguous. Crespi (1988) and Smith (1992), among others, have pointed out that stimuli designed to aid respondents' recall may also increase their interest in the topic and so may introduce bias into the estimates.

The results reported in this chapter relate to the two questionnaire versions designed as a test of contextual effects. First, the results themselves are presented; the following sections then discuss these results in the context of previous work before suggesting how pollsters might incorporate the findings in their own work.

8.2 Context Effects on Voting Intention Estimates

All treatments[43] explored in this study were tested with and without an extensive set of contextual questions preceding the voting intention questions. The purpose of this was twofold. First, it tested whether providing a series of contextual questions lowered the proportion of respondents giving undecided responses. Second, since a reduction in the proportion of undecided respondents is only of value if it improves the accuracy of the estimates, this approach also allowed a comparison of respondents' predicted behaviour with either the overall electorate behaviour or with their self-reported behaviour.

Table 5 contains the aggregate voting intention estimates for the contextual and non-contextual questionnaire versions in all six surveys. As both questionnaire versions resulted in a proportion of undecided and non-voting responses, this table contains two sets of estimates. First, it shows the raw percentages, including undecided respondents, non-voters and refusals). Second, it contains the percentages resulting after these groups had been re-allocated proportionately across the parties.[44]

---

[43] That is, the three treatments outlined in Chapter 6: the "vote yesterday" question administered directly; the same question administered via a secret ballot, and the Juster-based party probability questions. The results in this chapter are based on the two questionnaire versions which examined only the effects of contextual questions.

[44] Although many methods of dealing with undecided respondents exist (see Chapter 2 for a more detailed discussion of these), Crespi (1988) noted that the majority of pollsters employ proportional allocation.

Table 5: Context Effects on Voting Intention Estimates

	Contextual Questionnaire Version (N=713)[1]				Non-Contextual Questionnaire Version (N=715)[1]			
Party Support	**Raw Estimates**	**Undecideds Allocated**	**Actual Result[2]**	**Error**	**Raw Estimates**	**Undecideds Allocated**	**Actual Result[2]**	**Error**
	%	%	%	%	%	%	%	%
Alliance	8.4	10.6 *(1.9)*[3]	17.5	-6.9	10.2	15.1 *(2.2)*[3]	17.5	-2.4
Labour	31.7	39.9 *(3.1)*	39.6	+0.3	27.0	40.0 *(3.1)*	39.6	+0.4
National	31.8	40.0 *(3.1)*	36.0	+4.0	25.6	37.9 *(3.1)*	36.0	+1.9
NZ First	6.2	7.8 *(1.7)*[3]	4.3	+3.5	4.3	6.4 *(1.5)*	4.3	+2.1
Other	1.4	1.7 *(0.8)*	2.6	-0.9	0.3	0.4 *(.001)*	2.6	-2.2
None	6.7	---	---	---	7.5	---	---	---
Undecided	13.0[4]	---	---	---	22.8[4]	---	---	---
Refused	0.8	---	---	---	2.2	---	---	---
TOTAL	**100.0**	**100.0**	**100.0**	**3.1**[5]	**100.0**	**100.0**	**100.0**	**1.8**[5]

1. Because examining the results in the context of each electorate would have resulted in small cell counts, these results are aggregated across the six surveys and represent the total number of responses elicited using versions 1 and 2 of the questionnaire.

2. The actual results were calculated by summing the actual vote distributions across all three electorates; because no interviewing was conducted in Whitby, part of the Western Hutt electorate, the results from polling booths located within Whitby were excluded from this aggregation (see Chapter 6 for further discussion of this). The standard errors associated with each estimate are in parentheses.

3. Differences between the estimates and the actual results were significant at $p < 0.05$.

4. Difference between the two proportions of undecided respondents significant at $p < 0.001$.

5. Mean absolute error.

Looking first at the effect of the two questionnaire versions on the number of undecided respondents, the contextual version clearly resulted in significantly fewer undecided responses ($p < .001$). However, this reduction in the number of undecided respondents did not translate into a higher level of accuracy. Instead, these results suggest the questionnaire version which included a set of rapport building questions prior to the voting intention question resulted in a higher overall error level than the version which began more directly (3.1% compared to 1.8%).

The non-contextual version correctly predicted that Labour would receive the highest proportion of the votes while the contextual version suggested the race was really too close to suggest a clear outcome. Given the error margins associated with the estimates, both versions provided accurate predictions of the party that would win the greatest share of the vote.

While in most cases the probability of accepting a false null hypothesis did not exceed 5%, some of the errors observed were non-trivial, particularly those relating to the contextual version. By contrast, the non-contextual version resulted in smaller errors and consequently was more accurate.

This finding is contrary to the research hypothesis which posited that the provision of contextual questions would lead respondents to provide more accurate estimates. However, since the results presented in Table 5 are aggregate results, it is possible that the combined data obscure differences that might otherwise have emerged. For this reason, Table 6 contains the mean absolute errors relating to each individual electorate. (Appendix E contains the full tables from which these errors were calculated.)

Table 6: Context Effects on Mean Absolute Errors: Electorate Level

Electorate	Contextual Version		Non-Contextual Version	
	N	Mean Abs. Error	N	Mean Abs. Error
Manawatu 1	102	5.5	112	4.9
Miramar 1	116	3.4	112	1.9
Western Hutt 1	134	3.3	132	1.8
Manawatu 2	128	5.5	126	3.8
Miramar 2	102	6.2	110	4.3
Western Hutt 2	124	2.9	129	2.4

Table 6 shows that, in every electorate, administration of the non-contextual questionnaire version resulted in a lower overall error than the version which included contextual questions. Although the small cell counts effectively preclude testing for statistically significant differences, the chance of finding lower errors consistently associated with the same version of the questionnaire in all six electorates is low ($p<.1$).[45]

Yet despite the lower mean absolute errors associated with the non-contextual version, the predictions based on the non-contextual questionnaires did not always put the party that won each electorate ahead in the polls (see Appendix E). Nor did the non-contextual version display a superior predictive ability; both versions correctly predicted some outcomes, but failed to predict others. Because the cell counts relating to these analyses are small, the mean absolute errors are used as the basis of comparisons.

Not unexpectedly, the results from Table 6 confirm the overall findings reported in Table 5. Although in each electorate the contextual questionnaire produced fewer

---

45 The test used to establish this was:

$$b(x, n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

Where: n = total number of trials
x = the number of "successes" observed during n trials

undecided respondents, it also produced higher mean absolute errors. This finding is contrary to the research hypothesis and much of the literature outlined in earlier chapters, which posited that the contextual questions would lead respondents to provide more accurate predictions. While the contextual questions reduced the proportion of undecided responses, they did not necessarily help respondents to predict their behaviour correctly. These results do not, therefore, support the assumption that a reduction in the proportion of undecided respondents will lead to an improvement in the accuracy of the estimates.

As noted in Chapter 7, it is possible that the voter registration and past voting behaviour questions which preceded the voting intention question (and which were initially thought to provide a non-contextual introduction) in fact created a context which helped respondents to clarify their likely behaviour. However, it is more difficult to explain why this minimal context should have produced more accurate estimates than did the detailed context. Section 8.4 explores possible explanations in greater detail.

Overall, Tables 5 and 6 provide strong evidence to support the hypothesis that contextual questions reduce the proportion of undecided responses given. However, although the expected reduction in undecided respondents occurred, the anticipated increase in the accuracy of the estimates did not, and neither table supported the view that reducing the number of undecided respondents would result in more accurate estimates.

An alternative approach to assessing the effects the two questionnaires had is to compare respondents' predicted and self-reported behaviour. Chapter 7 noted the problems caused by social desirability bias, though this would seem less likely to affect party support estimates. However, it is possible that other influences, such as a bandwagon effect, noted in Chapters 2 and 6, may have biased these results. Accordingly, neither set of results can be argued as presenting a clear or unbiased view of respondents' actual behaviour.

Table 7 contains details of respondents' predicted and self-reported behaviour.

Table 7: Context Effects on Voting Intention: Predicted versus Self-Reported Behaviour

CONTEXTUAL VERSION (N=329)

Predicted Voting	Alliance	Labour	National	NZ First	Other	None	Don't Know
	%	%	%	%	%	%	%
Alliance (n=31)	77.4	19.4	0.0	0.0	0.0	0.0	3.2
Labour (n=111)	7.2	78.4	9.0	0.9	0.9	0.0	3.6
National (n=111)	3.6	9.9	77.5	0.9	1.8	0.0	6.3
NZ First (n=19)	21.1	21.1	0.0	52.6	0.0	0.0	5.3[2]
Other (n=4)	0.0	0.0	25.0	0.0	75.0	0.0	0.0
None (n=16)	12.5	31.3	31.3	0.0	6.3	12.5	6.3[2]
Don't Know (n=37)	18.9	21.6	29.7	8.1	5.4	0.0	16.2[2]

PROPORTION CORRECTLY PREDICTED: 76.1%[1]

NON-CONTEXTUAL VERSION (N=326)

Predicted Voting	Alliance	Labour	National	NZ First	Other	None	Don't Know
	%	%	%	%	%	%	%
Alliance (n=35)	60.0	22.9	5.7	2.9	0.0	0.0	8.5
Labour (n=77)	5.2	84.4	6.5	0.0	3.9	0.0	0.0
National (n=97)	8.2	8.2	74.2	1.0	0.0	0.0	8.2[2]
NZ First (n=15)	20.0	20.0	20.0	33.3	0.0	6.7	0.0
Other (n=0)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
None (n=19)	15.8	26.3	42.1	0.0	0.0	0.0	15.8
Don't Know (n=83)	18.1	42.2	28.9	2.4	2.4	0.0	6.0

PROPORTION CORRECTLY PREDICTED: 72.8%[1]

1. The proportion correctly predicted was calculated by establishing the proportion of respondents who predicted they would vote for one of the four major parties and who did so.
2. Table may contain small rounding errors.

Table 7 suggests that the contextual version resulted in more accurate estimates than the non-contextual version, though the proportion of respondents whose voting behaviour was correctly predicted did not differ significantly between questionnaires. Examination of specific party support estimates shows both versions produced more accurate estimates of support for the Labour and National parties, which both have well established electoral histories. The contextual version produced more accurate estimates for the two new parties (Alliance and NZ First), which suggests that contextual questions may be of greater value when respondents need to consider comparatively unfamiliar options, (though the small cell counts make this suggestion rather tentative).

Very few of the respondents who said they would not vote for any of the parties in fact reported not voting.[46] This lends support to the view that respondents may declare themselves non-voters or undecided to protect the privacy of their opinions. Equally, it is possible that social desirability bias affected these estimates, as respondents may not have wished to admit to non-voting.

For both versions, the pattern of party support reported by respondents who predicted they would vote for none of the parties was similar to the overall pattern of predicted voting. That is, allocating the declared non-voters proportionately would not have affected the predicted voting estimates. However, the undecided respondents' pattern of voting was less similar to the overall predicted pattern, suggesting that proportional allocation of this group may reduce the accuracy of the overall estimates. Since the cell counts relating to the declared non-voting and undecided groups were small, this conclusion also remains speculative.

Overall, these results do not confirm the earlier impression that the non-contextual version produced more accurate estimates; instead they suggest there were no real differences between the estimates produced by the two questionnaire versions. Further

---

46 Respondents who said they would not vote for any party actually had the following probabilities of voting:

	Probability of Voting
Contextual version (n=16)	57.5%
Non-contextual version (n=19)	53.7%

evidence of this comes from the results relating to each individual electorate.

Analogous to Table 6, Table 8 presents the results of respondents' predicted and self-reported behaviour in each electorate. The cell counts here are clearly small and the purpose of the table is simply to examine the pattern of errors, in particular to assess whether these occurred in the same direction as that noted in Table 7.

Table 8: Context Effects of Mean Absolute Errors: Predicted versus Self-Reported Behaviour - Electorate Level

Electorate	Contextual Versions		Non-Contextual Version	
	N	Mean Abs. Error	N	Mean Abs. Error
Manawatu 1	45	2.1	45	4.4
Miramar 1	38	2.6	35	16.7
Western Hutt 1	73	2.8	74	1.9
Manawatu 2	56	4.6	65	5.8
Miramar 2	45	5.3	42	3.5
Western Hutt 2	72	3.1	66	3.0

As a subset of the data presented in Table 7, not surprisingly Table 8 also suggests there are, on average, no differences between the two versions. Although in some instances the contextual version resulted in lower errors, this was not always so, and the results did not reveal a consistent pattern.

Tables 5 and 7 reveal that the only estimates that consistently exceeded respondents' reported behaviour related to the National Party. Bandwagon effects suggest respondents associate themselves with "winners" and distance themselves from "losers". Technically, the National Party won the election insofar as it remained in government. However, some would argue this victory was pyrrhic, since the balance of power remained in doubt for some weeks after the election. Given this scenario, some National Party voters may have distanced themselves from a rather disappointing outcome by mis-reporting their behaviour. Without recourse to actual voting records, this explanation remains speculative, however, it highlights the ambiguity associated with both sets of results.

The pattern of error in Tables 5 and 6 pointed to the superior accuracy of the non-

contextual version and supports the view that pollsters should avoid the use of contextual questions and proportionately allocate undecided respondents. Although the results in Tables 7 and 8 do not contradict this conclusion, they suggest that the difference between versions was minimal and that pollsters would, on average, obtain similar error levels regardless of whether they used contextual questions.

8.3 Effects of Weighting Procedures

All the estimates presented thus far have been based on unweighted data. Common practice among researchers is to weight the data to correspond to known population parameters. Demographic variables, such as age and gender are typically used for weighting, although there is no strong evidence pointing to a relationship between these variables and voting behaviour (see Murphy, 1975 or Crespi, 1988, for a discussion of this). Nevertheless, to enable an assessment of this practice, these data were also weighted so the age-gender groups matched those found in the three electorates. In addition, the data were separately weighted so respondents' reported voting behaviour in the 1990 election matched the actual electorate results. Finally, the data were also separately weighted by respondents' overall likelihood of voting. Table 9 contains the mean absolute errors that resulted after these various weighting procedures were implemented. (Full Tables, together with specific details of the weighting procedures, appear in Appendix H.)

Table 9: Effects of Weighting Procedures on Mean Absolute Errors -Contextual Effects

Electorate	Contextual Unweighted	Contextual Demographics Weighted	Contextual Probability Weighted	Contextual Weighted by Past voting	Non-Contextual Unweighted	Non-contextual Demographic Weighted	Non-Contextual Probability Weighted	Non-Contextual Weighted by Past Voting
Manawatu 1	5.5	6.4	4.6	3.6	4.9	5.1	5.1	4.7
Miramar 1	3.4	3.9	3.1	3.4	1.9	2.8	2.7	4.2
Western Hutt 1	3.3	2.7	3.0	3.5	1.8	3.0	1.9	1.4
Manawatu 2	5.5	6.0	6.1	5.0	3.8	5.5	3.6	4.7
Miramar 2	6.2	6.4	6.1	5.8	4.3	4.9	3.8	2.6
Western Hutt 2	2.9	3.6	4.0	2.2	2.4	2.1	2.0	1.8
Overall Mean Absolute Error	4.5	4.8	4.5	3.9	3.2	3.9	3.2	3.2

Ironically, Table 9 reveals that weighting by demographic characteristics actually increased the mean absolute error over that obtained from the unweighted data in ten of the twelve comparisons. Overall, weighting respondents' predictions by their probability of voting resulted in the same error levels, while weighting the data to reflect previous electoral results produced similar or lower error levels. Since this election saw the entry of two new political groups, weighting by past voting behaviour may have failed to incorporate the volatility brought about by this increased choice, and may prove even more effective in a more stable political environment.

While the effects of these weighting procedures need to be examined in other contexts, this table suggests researchers might consider other weighting variables when exploring how to improve the accuracy of their estimates. In particular, it would seem useful to investigate further the effects of weighting by past voting behaviour (once mainstream electoral choice has stabilised) and by respondents' probability of voting.

8.4 Relationship with the Literature

Research discussed in Chapter 3 concurred that questionnaires containing few or no channelling questions to lead respondents through to questions on future voting behaviour generally resulted in higher proportions of undecided respondents. The allocation of these undecided respondents to parties in the same proportions as respondents who had expressed an opinion has been the commonest way of dealing with this problem. However, if these undecided respondents did not subsequently behave in the same way as those who expressed an opinion, the estimates made would be biased. Clearly, the larger proportion of undecided respondents, the greater the potential for bias would be.

To address this problem more directly, many researchers have used contextual questions which enable respondents to develop some familiarity and confidence with the survey topic. Labaw (1980), for example, suggested researchers develop a set of attitudinal and opinion questions, together with questions eliciting details of respondents' past behaviour, on the grounds that the former would provide a context for questions relating to future behaviour, while the latter would be the best predictor of this behaviour. Sudman & Bradburn (1982) also recommended beginning questionnaires with general, non-threatening questions, and then moving from these to more specific questions which addressed the behavioural issues directly.

In addition to discussing general structural issues of questionnaire design, numerous researchers have commented on problems which stem from respondents' understanding of the question or survey topic. Many have dwelt on the problems caused by respondents who believe they know little about the research topic, or who feel it bears little relevance to them personally, and who may subsequently give more uncertain and less accurate responses than more interested or knowledgeable respondents. In addressing this problem, Sudman & Bradburn (1982) commented on the provision of contextual questions and concluded that these may increase the saliency of a particular topic and clarify the issues explored in a survey. Their finding reflected earlier work by Blakenship (1943) and Sigelman (1981), both of whom concluded that higher levels of indecision resulted when key questions, were placed early in a questionnaire than when they appeared later in the questionnaire. The findings reported here confirm these conclusions and lend support to Strack's (1992) view that provision of contextual questions primes respondents and increases the proportion who express a definite response.

However, as noted, although the proportion of respondents offering a prediction certainly increased, the accuracy of those predictions did not. The questionnaire version containing a more detailed introduction to the questions of interest produced less accurate estimates (when compared to the actual election results) than the versions which began more directly. This pattern was consistent across all electorates. Husbands (1987) earlier hinted at the mixed blessing brought about by contextual questions when he outlined both sides of the "warm start/cold start" debate. More recently, Smith (1992) noted that contextual questions may result in biased estimates.

Crespi (1988) also discussed the error associated with contextual questions. He noted that several pollsters did not precede the voting intention questions they used with contextual questions on the grounds that these could shape respondents' perceptual environment and introduce error into their responses. Thus while it is difficult to explain why the expected pattern of results did not appear, it is clear that others have also sought to explain similarly anomalous results. This suggests that the issue of rapport development may be more ambiguous than first thought.

Although the contextual versions of the questionnaire contained a set of questions designed to create a channel leading to the core behavioural questions, the non-

contextual version was not devoid of preliminary questions. This raises the possibility that the questions exploring respondents' voting registration status and past voting behaviour may have formed a sufficient context to enable respondents to clarify their likely future behaviour.

These results also raise the question of whether a more detailed context may have confused respondents and confounded their choice instead of clarifying their behaviour (as Crespi, 1988 implicitly suggested). In other words, consideration of a range of electoral and party support issues may have clouded respondents' judgement and led them to provide less accurate predictions. By contrast, the two behaviourally oriented questions (investigating respondents' voting registration status and past voting behaviour) used in the non-contextual version may have honed respondents' judgement, thereby improving the accuracy of their responses.

However, this explanation does not fit well with the accumulated evidence about general questionnaire design effects. Moreover, it would appear logical for any confusion to be manifested in an increased proportion of undecided responses, and the results show quite clearly that this did not occur. While it is possible that, instead of simply confounding respondents' choice (an effect likely to lead to an increase in the proportion of undecided responses), the more detailed context actually biased respondents' predictions, these data do not show how or why this occurred.

An alternative explanation is that events occurring between the interview and the election date affected the accuracy of respondents' predictions, a problem noted by several researchers (see Worcester, 1984; or Crespi, 1988, for example). Though, if this were so, the observed errors should have been randomly distributed. Since the likelihood that the observed pattern of errors occurred by chance is very low, this explanation seems unlikely. Furthermore, if unanticipated events had intervened to affect respondents' predicted behaviour, it would seem logical that the estimates resulting from the first round of interviewing would have greater errors than those produced by the later round. Table 6 shows this was not the case (only two of the mean absolute errors declined in the second round of interviews), thus raising

additional questions about the likely influence of unanticipated events.[47]

Crespi (1988) noted that a number of US pollsters deliberately did not precede the voting intention question with contextual questions because they thought it would introduce error into the voting intention estimates. If this is so, another explanation for these results may be that respondents use polls to send messages to politicians. That is, instead of acting simply as passive conveyers of information to interviewers, respondents may manipulate the information they provide so as to convey a signal to political parties. Anecdotal evidence suggests this behaviour occurs already, though the extent to which it affects poll estimates is difficult to quantify.[48] However, it is possible that respondents who first considered a range of policy issues may have felt more aggrieved by perceived failings on the part of some parties and so allocated what amounted to a protest vote, which reflected their attitude, but not their likely behaviour.

The public have become increasingly aware of the weight politicians place on political polls (despite their stated views to the contrary), thus registering dissatisfaction with a party or parties' performance via a poll has become an effective means for the public to communicate with politicians. In comparison, respondents asked to recall only whether they had confirmed their voting registration and to state their voting behaviour at the last General Election may be less likely to have considered the same range of policy issues prior to indicating their likely behaviour. Their responses are therefore likely to have been shaped more by their previous behaviour than by their views on the inadequacies or otherwise of different parties' policies.

This explanation has an intuitive appeal and some support from other details gathered in the follow-up telephone survey. In addition to exploring whether respondents voted and for whom they cast their vote, the post-election survey also investigated when

---

47 Although it must be acknowledged that the campaign period, during which the second round of interviewing took place, may have introduced additional volatility and perhaps explains why the expected decrease in error did not eventuate.

48 Letters to the editor and talk-back radio suggested the public recognised that political parties used opinion polls to shape, rather than to reflect, public opinion. Indeed, some people claimed to have countered politicians' efforts by providing deliberately misleading information to interviewers.

respondents decided upon the party or candidate they would support in the General Election. Interestingly, those people who predicted they would vote for one of the minor parties gave significantly less accurate predictions than those who stated they planned to support either of the two major parties.[49] Again, this supports the suggestion that some respondents may have deliberately mis-reported their views or likely behaviour.

Notwithstanding these explanations, Tables 7 and 8, which examined respondents' predicted and self-reported behaviour, suggest that sampling error may also have affected the results presented in the preceding tables, although the data do not enable an assessment of the specific effect sampling error may have had. Overall, the contextual version of the questionnaire did not perform as well as originally hypothesised.

Some researchers have noted that occasionally respondents develop an excessive rapport with interviewers. In these instances, respondents who have volunteered opinions which appear to indicate support for a certain party may feel that to declare an intention to vote for another party would be inconsistent with their earlier responses. Rather than risk appearing fickle, they may align their responses even if this means their answers no longer reflect their likely behaviour (see Bradburn & Sudman, 1988, and Smith, 1992). Thus an alternative explanation for some of the inconsistencies apparent in these results may be respondents' desire to present themselves as thoughtful, intelligent and consistent providers of information.

By contrast, another effect of an excessive rapport is respondent fatigue, and Herzog & Bachman (1981) noted a phenomenon they described as straight-line responding. When this occurs, respondents nominate the same option to a set of questions, irrespective of their own views on the questions put to them.

While it is not possible to ascertain which, if any, of these explanations may be most relevant, other researchers have also noted that contextual questions may not

---

[49] Two thirds of those who said they would have voted for the Alliance in fact reported having done so compared to 80% of Labour and National supporters, and only a third of NZ First supporters.

necessarily improve the accuracy of resulting estimates. Although these results appear to contribute to the ambiguity already surrounding this issue, this in itself may prove a substantive finding. The fact that neither questionnaire version produced dramatically different results, and that most of the population values fell within the confidence intervals surrounding the estimates, raises the hopeful possibility that the estimates may be more robust than hitherto thought. However, this possibility depends on the rigorous calculation of confidence intervals, as Collins (1988) and Jowell *et al* (1993) have already noted.

8.5 Summary

In spite of a strong research tradition pointing to the beneficial effects of contextual questions, this research did not confirm the posited link between contextual questions and an increased accuracy in the resulting estimates. Several explanations for this apparent anomaly have been advanced, although the data do not allow identification of the most likely explanation.

Overall, this research suggests that, while provision of contextual questions reduces the proportion of undecided respondents, this reduction does not necessarily mean the resulting estimates are more accurate. Inclusion of rapport developing questions therefore had little discernible effect on the overall accuracy of the estimates, although there was some very tentative evidence to suggest it helped supporters of recently formed political parties clarify their preferences.

As far as pollsters are concerned, these results represent ambivalent outcomes. While neither questionnaire version consistently produced more accurate estimates, this may indicate that the estimates are robust and are independent of variations in the question context. However, before any firmer conclusions can be drawn, it is important to examine the effects of contextual questions when combined with other treatments. Chapters 9 and 10 present the results of this further investigation.

Chapter 9

SECRET BALLOT EFFECTS: RESULTS AND DISCUSSION

9.2 Introduction

Asking respondents who they plan to vote for represents, to some at least, a potential invasion of their privately held political views. Rather than reveal their opinions or likely behaviour to a comparative stranger, these respondents may declare they are undecided, state that they do not plan to vote, or refuse to answer the survey.

Although contextual questions reduce the proportion of undecided respondents to between 10% to 15% of the total sample, a group of this size may still introduce error if the behaviour of its members deviates dramatically from that of respondents who did express an opinion or intention. That an undecided group of this size remains even after contextual questions are employed, suggests indecision may arise for a number of reasons, not all of which were addressed by the question sequence examined in the previous chapter. Whether respondents' indecision stems from a genuine inability to choose between the different parties or candidates, or whether it arises from a desire to maintain the privacy of their beliefs, is difficult to discern from the data presented thus far.

The research reported in this chapter investigated the extent to which the undecided group comprises privacy-conscious individuals. Respondents were administered the voting intention questions using a secret ballot which allowed them to express their voting intentions independently of the interviewer. As in Chapter 8, the research explores first how this treatment affected the size of the undecided group, before turning to examine its effects on the accuracy of the estimates produced. The chapter then compares the secret ballot results to the literature documenting confidentiality treatments, and outlines some implications for researchers.

9.2 Effects of Secret Ballots on Voting Intention Estimates

Two versions of the questionnaire used a secret ballot to collect respondents' answers to the voting intention question. When administering this question, interviewers gave respondents the equivalent of a ballot paper, which they marked and placed in a secret ballot box independently of the interviewer. The secret papers differed from authentic ballot papers in that they only specified party options, rather than party and candidate

details, as real ballot papers do. In addition, they made no provision for undecided responses or for people who did not plan to vote for any party, although respondents could volunteer these options if they wished. While this latter difference may have made it more difficult for undecided or non-voting respondents to express their intentions, structuring the ballot paper in this way closely simulated actual voting behaviour.

The two questionnaire versions used to explore this treatment differed according to whether contextual questions preceded the voting intention questions and presentation of the secret ballot. In this way, these questionnaires paralleled those discussed in the preceding chapter and differed only in the administration of the voting intention question. Table 10 presents the overall results produced by these questionnaires, and contains both the raw percentages and those resulting from the re-allocation of undecided and non-voting respondents.

Table 10: Effects of a Secret Ballot on Voting Intention Estimates

	Contextual Questionnaire Version (N=679)[1]				Non-Contextual Questionnaire Version (N=683)[1]			
Party Support	**Raw Estimates**	**Undeclded Allocated**	**Actual Result[2]**	**Error**	**Raw Estimates**	**Undeclded Allocated**	**Actual Result[2]**	**Error**
	%	%	%	%	%	%	%	%
Alliance	12.1	12.8 *(2.2)*	17.5	-4.7[3]	13.2	14.6 *(2.3)*	17.5	-2.9
Labour	40.5	42.7 *(3.2)*	39.6	+3.1	35.1	38.8 *(3.1)*	39.6	-0.8
National	33.6	35.5 *(3.1)*	36.0	-0.5	32.1	35.4 *(3.1)*	36.0	-0.6
NZ First	7.5	7.9 *(1.7)*	4.3	+3.6[3]	8.3	9.3 *(1.9)*	4.3	+5.0[3]
Other	1.0	1.1 *(0.7)*	2.6	-1.5	1.8	1.9 *(0.9)*	2.6	-0.7
None	2.4	---	---	---	4.1	---	---	---
Undecided	1.9[4]	---	---	---	4.2[4]	---	---	---
Refused	1.0	---	---	---	1.2	---	---	---
TOTAL	**100.0**	**100.0**	**100.0**	**2.7[5]**	**100.0**	**100.0**	**100.0**	**2.0[5]**

1. Because examining the results in the context of each electorate would have resulted in very small cell counts these results are aggregated across the six surveys and represent the total number of responses elicited using versions 3 and 4 of the questionnaire.
2. The actual results were calculated by summing the actual vote distributions across all three electorates; because no interviewing was conducted in Whitby, part of the Western Hutt electorate, the results from polling booths located within Whitby were excluded from this aggregation (see Chapter 6 for further discussion of this).
3. Probability of accepting the null hypothesis when it is false exceeds 5%.
4. Difference between undecided groups significant at $p < 0.05$.
5. Mean absolute error.

Table 10 shows quite clearly the dramatic reduction in the undecided group that resulted from the use of a secret ballot. While Table 5 showed undecided groups of 13% and 22.8%, these declined to 1.9% and 4.2%, respectively, when the secret ballot was used. Whether this reduction was because of the secret ballot, or because of the omission of an explicit undecided response option, is difficult to determine. However, the contextual version of the secret ballot questionnaire elicited a significantly lower proportion of undecided respondents than the non-contextual version. This suggests that context effects continued to have some effect on the number of undecided respondents even when the overall proportions of undecideds were very low.

However, the decline in the proportion of undecided respondents was not paralleled by an increase in the accuracy of the estimates obtained (although both correctly predicted that the Labour Party would gain the highest share of vote). Neither of the secret ballot versions had a significantly lower error than their more directly administered counterparts and, in some cases, the error was slightly higher. This suggests that respondents who might otherwise have declared themselves undecided did provide a response. Yet the use of a secret ballot to reduce the proportion of undecided responses provided estimates that were no more accurate than those resulting from proportional allocation of undecided respondents (as shown in Table 5).

Within Table 10, the contextual version again resulted in a higher overall error than did the non-contextual version, affirming the pattern noted in Table 5. Some possible explanations for this unexpected finding have been explored in section 8.4, and Table 10 sheds no new light on why the contextual version did not elicit more accurate predictions. Closer examination of Table 10 suggests the largest errors tend to be associated with the new parties, perhaps pointing to the difficulty of predicting behaviours never before undertaken (such as voting for a party that had not previously contested a General Election), despite the provision of contextual or confidentiality devices.

The probability of accepting the null hypothesis (that the difference between the estimates and the actual results was zero) when it was false, was less the 5% in all but three cases. However, despite falling within the 5% acceptance level, some errors remained quite large (around 3%) and, if the secret ballot methodology is to gain

acceptance, these estimates should ideally be much smaller.

Examination of the refusals in Tables 5 and 10 shows all treatments resulted in a similar sized group of respondents who refused to answer the voting intention question. That the proportion of refusals remained consistent in both Tables 5 and 10, suggests the existence of a small hard-core group whose members do not wish to reveal their views, irrespective of the level of confidentiality offered to them. Fortunately, the size of this group is small, therefore it appears unlikely to have had any substantive effect on the estimates.

Overall, these results confirm that dramatic reductions in the size of the undecided group are possible when the confidentiality associated with respondents' answers is increased. However, none of the treatments examined thus far has resulted in more accurate estimates than direct administration of the voting intention question. In order to examine whether this pattern occurred uniformly across the results, Table 11 presents the mean absolute errors relating to each electorate. (Appendix F contains full tables.)

Table 11: Secret Ballot Effects on Mean Absolute Errors - Electorate Level

Electorate	Contextual Version		Non-Contextual Version	
	N	Mean Abs. Error	N	Mean Abs. Error
Manawatu 1	101	5.9	104	4.3
Miramar 1	96	1.9	97	3.0
Western Hutt 1	131	3.9	128	3.7
Manawatu 2	117	6.2	119	4.4
Miramar 2	110	3.6	110	4.4
Western Hutt 2	124	3.3	125	3.4

Unlike Table 6, which showed a consistently higher level of error associated with the

contextual version of the questionnaire, Table 11 shows a more random error pattern, with neither version consistently superior to the other. This suggests that researchers could employ either secret ballot treatment without affecting the overall level of error associated with their estimates.

The secret ballot error levels exceeded those resulting from the directly administered non-contextual treatment. In fact, this latter treatment produced lower mean absolute errors in every electorate. Thus, while the differences between the overall mean absolute errors shown in Tables 6 and 11 were not significant, it is unlikely that the direct non-contextual version would produce smaller errors in every electorate by chance alone ($p < .0003$).

Furthermore, Appendix F reveals that the estimates from either questionnaire version only sporadically predicted the winning party. Although the small cell counts increase the error associated with those estimates, the secret ballot estimates often deviated sharply from the actual results.

Once again, these results suggest that a reduction in the size of the undecided group does not increase the accuracy of the sample estimates. On the contrary, the results presented here and in Table 6 show that reductions in the proportion of undecided respondents are paralleled by decreases in the accuracy of the survey estimates. While it is difficult to argue that a causal relationship between the number of undecided respondents and the accuracy of the estimates exists, these results do raise some important questions. In particular, they suggest that the use of contextual questions, or the addition of a more sophisticated confidentiality device, such as a secret ballot, may effectively coerce respondents into providing an answer. If respondents do feel pressured in this way, they may provide any answer, simply to comply with the request made of them. This does not imply that these respondents deliberately mis-report their behaviour, but instead suggests that the treatments tested thus far may be less helpful than initially thought. Section 9.4 explores these ideas further.

The errors arising from the second round of interviewing were again greater than those associated with the earlier surveys. This finding also queries the view that, as the time lag between the interview date and the election date decreases, the estimates' accuracy will increase. Thus if unanticipated events were responsible for the errors

noted, it would seem logical for the errors to decrease as the time between the interview and the election decreased. In neither Table 6 nor Table 11 are the results reported consistent with this expectation. While unanticipated events may have affected the accuracy of respondents' predictions, the tables suggest other factors contributed more to the errors they contain.

Overall, Table 11 confirms the conclusions drawn from Tables 5, 6 and 10, namely that neither of the treatments (contextual questions nor the use of a secret ballot) appears to have improved the accuracy of the sample estimates. However, these tables all compared respondents' predicted behaviour to the actual electorate's behaviour. As noted earlier, this measure of accuracy is not perfect, thus the following tables examine a sub-sample of respondents and compare their predicted and self-reported behaviour. Table 12 contains the first results pertaining to this analysis.

Table 12: Effects of a Secret Ballot on Voting Intention: Predicted versus Self-Reported Behaviour

CONTEXTUAL VERSION

	Self-Reported Behaviour						
Predicted Voting	**Alliance**	**Labour**	**National**	**NZ First**	**Other**	**None**	**Don't Know**
	%	%	%	%	%	%	%
Alliance (n=43)	62.8	23.3	7.0	0.0	0.0	0.0	7.0[2]
Labour (n=125)	8.0	69.6	16.8	1.6	0.8	0.0	3.2
National (n=115)	3.5	9.6	74.8	0.9	3.5	0.0	7.9[2]
NZ First (n=25)	12.0	28.0	28.0	20.0	4.0	0.0	8.0
Other (n=5)	0.0	20.0	20.0	0.0	60.0	0.0	0.0
None (n=6)	0.0	33.3	50.0	0.0	0.0	0.0	16.7
Don't Know (n=6)	50.0	0.0	16.7	0.0	0.0	0.0	33.3

PROPORTION CORRECTLY PREDICTED: 66.6%[1]

NON-CONTEXTUAL VERSION

	Self-Reported Behaviour						
Predicted Voting	**Alliance**	**Labour**	**National**	**NZ First**	**Other**	**None**	**Don't Know**
	%	%	%	%	%	%	%
Alliance (n=44)	56.8	25.0	2.3	6.8	2.3	0.0	6.8
Labour (n=123)	6.5	78.0	8.9	1.6	0.8	0.8	3.3[2]
National (n=108)	6.5	12.0	71.3	0.9	1.9	0.0	7.4
NZ First (n=23)	21.7	30.4	8.7	26.1	4.3	0.0	8.7[2]
Other (n=8)	25.0	0.0	12.5	12.5	25.0	0.0	25.0
None (n=10)	30.0	30.0	30.0	0.0	0.0	0.0	10.0
Don't Know (n=16)	25.0	50.0	6.3	6.3	0.0	0.0	12.5[2]

PROPORTION CORRECTLY PREDICTED: 68.5%

1. The proportion correctly predicted was calculated by establishing the proportion of respondents who predicted they would vote for one of the four major parties who in fact went on to vote for the party they had earlier specified.
2. Table may contain small rounding errors.

Table 12 suggests that the non-contextual version elicited slightly more accurate predictions from respondents, although the difference between the two treatments is not statistically significant. Closer examination of the accuracy of the voting intention predictions suggests that both versions produced more accurate estimates for the Labour and National parties, although the Alliance estimate produced by the contextual version was also similar to these. Neither version produced accurate predictions for the New Zealand First party, and the non-contextual version was especially poor in this respect. These results support the tentative hypothesis advanced earlier, that the inclusion of contextual questions may help respondents to clarify their behaviour with respect to new options. However, these results, and those presented in Table 7, suggest that although the contextual versions may result in more accurate estimates of minor party support, the level of accuracy still has considerable room for improvement.

Table 12 has several parallels with Table 7. First, all versions of the questionnaire examined thus far have produced more accurate estimates for the two major parties than for the new parties contesting the election. Again, this finding points to the difficulty of predicting new behaviour patterns.

Although the proportion of respondents declaring themselves undecided, or claiming to have no intention of voting was very small, both of these groups showed a different pattern of voting behaviour compared to those who had earlier stated an intention. These results raise the possibility that routine proportional allocation of these groups may introduce error into the estimates.

Examination of the proportion of respondents correctly classified shows that both secret ballot versions resulted in a lower level of accuracy than the equivalent questionnaires employing a more direct method of administering the voting intention questions. This finding further supports the idea that a reduction in the proportion of undecided respondents is not necessarily followed by an increase in the accuracy of the resulting estimates.

To investigate further the pattern of errors detected, Table 13 presents the mean absolute errors relating to respondents' predicted and self-reported behaviour across each electorate. The cell counts involved are very small, thus the results in Table 13

are interpreted cautiously.

Table 13: Secret Ballot Effects on Mean Absolute Errors: Predicted versus Self-Reported Behaviour

Electorate	Contextual Version		Non-Contextual Version	
	N	Mean Abs. Error	N	Mean Abs. Error
Manawatu 1	47	2.5	40	6.4
Miramar 1	41	2.6	45	9.3
Western Hutt 1	67	4.7	70	2.1
Manawatu 2	56	7.6	58	4.8
Miramar 2	49	6.3	45	3.6
Western Hutt 2	65	6.5	75	4.7

Table 13 suggests errors at the electorate level are randomly distributed as neither version of the questionnaire emerged as consistently more accurate. Overall, the table confirms that there was no evidence to support a superior performance by the contextual questionnaire version. Furthermore, the level of accuracy does not appear to improve as the interviewing moves closer to the actual election date. Again, this is inconsistent with Crespi's (1988) view that predictions increase in accuracy as the time lag between the interviews and the election date decreases. Conversely, as acknowledged earlier, it is also possible that electoral volatility increases as the election date draws nearer.

In summary, the tables presented in this section show that use of a secret ballot technique dramatically reduces the proportion of undecided responses given. However, it remains unclear whether this reduction occurred as a consequence of the increased confidentiality provided by the secret ballot or because respondents had no explicit undecided option presented to them. Although respondents administered the question directly also had no explicit undecided option presented to them, it is possible that the more tangible depiction of voting options on the ballot slip may have

inhibited at least some respondents from classifying themselves as undecided. However, the net consequence of this reduction was not the hypothesised improvement in the accuracy of the voting intention estimates. Moreover, the contextual questions, designed to clarify respondents' decision making, appear to have had the opposite effect. Possible explanations of these findings are advanced in section 9.4.

9.3 Effects of Weighting Procedures

As in Chapter 8, the estimates presented in Tables 10 to 13 are based on unweighted data. This section explores the effects of the weighting procedures described in Chapter 8 on the responses to these questionnaire versions. Table 14 contains the mean absolute errors which resulted from the application of the various weighting factors. (Appendix H contains the full tables from which these mean absolute errors were calculated.)

Table 14: Effects of Weighting Procedures on Mean Absolute Errors - Secret Ballot Effects

Electorate	Contextual Unweighted	Contextual Demographics Weighted	Contextual Probability Weighted	Contextual Weighted by Past Voting	Non-Contextual Unweighted	Non-contextual Demographic Weighted	Non-Contextual Probability Weighted	Non-Contextual Weighted by Past Voting
Manawatu 1	5.9	6.4	6.2	5.9	4.3	5.3	4.2	4.1
Miramar 1	1.9	3.0	1.7	1.9	3.0	4.5	2.6	3.0
Western Hutt 1	3.9	4.6	3.5	3.9	3.7	5.1	3.6	3.7
Manawatu 2	6.2	7.0	6.3	6.2	4.4	4.8	4.3	4.4
Miramar 2	3.6	3.6	3.7	3.3	4.4	4.8	4.6	4.3
Western Hutt 2	3.3	2.7	3.1	3.3	3.4	3.2	3.4	3.4
Overall Mean Absolute Error	4.1	4.6	4.1	4.1	3.9	4.6	3.8	3.8

Table 14 reveals exactly the same pattern of results as that found in Table 9. That is, weighting to correct age-gender imbalances in the sample increased the mean absolute error in nine of the twelve comparisons. That Table 14 confirms this pattern raises serious doubts about the effectiveness of weighting based on these two demographic variables, in this specific context. Neither of the other two weighting variables (respondents' overall probability of voting or the voting distribution recorded in the last general election) significantly decreased the error when compared to the error associated with the unweighted estimates. However, weighting by respondents' probability of voting resulted in small decreases in the mean absolute errors (see Appendix H), and this procedure, together with weighting the sample to reflect the most recent election's voting distributions, appears to merit closer scrutiny once the newer parties become better established.

Overall, however, Table 14 shows that weighting generally did not improve the accuracy of the estimates, and since, weighting increases the margin of error associated with the estimates, it seems logical to dispense with it until it is clearly associated with a decrease in error levels.

9.4 Relationship with the Literature

Earlier chapters documented and discussed the belief that increases in the confidentiality associated with respondents' answers would lead many of those who might otherwise have given an undecided response to provide a more accurate description of their views or likely behaviour (see Perry, 1979; Brusati, 1987; Crespi, 1988). As noted in the literature, the secret ballot method dramatically reduced the proportion of undecided respondents, and the use of contextual questions saw the group decrease in size further still.

However, despite reducing the proportion of people declaring themselves undecided, or who gave a false report of intended non-voting, the secret ballot method did not improve the overall accuracy of the estimates in this study (as the mean absolute errors reported in Tables 5 and 10 show). Perry (1979) clearly recognised this possibility when he noted that changes in the party support distributions were often greater than the reduction in size of the undecided group. However, despite this, he argued that only a tiny proportion of respondents would give invalid responses. It is difficult to tell, from his published work, whether Perry meant few respondents would

deliberately mis-report their intended behaviour, or whether he thought the level of accuracy would improve greatly. The data presented in this chapter provide no way of ascertaining whether respondents deliberately mis-reported their behaviour, though they show quite clearly that the expected improvement in the accuracy of the estimates did not come about.

Perry (1979) also commented on where questions employing a secret ballot device should appear in a questionnaire. In particular, he noted that the use of secret ballots later in an interview, after interviewers had established a rapport with respondents, resulted in more accurate estimates than when the ballot questions were presented earlier.

However, the results reported here show no difference in the proportions of respondents refusing to complete the secret ballot slip, regardless of where it appeared in the questionnaire. Unlike Perry, who noted a higher level of refusal when the secret ballot was introduced early in the questionnaire, these results suggest that, irrespective of its position in the questionnaire, the secret ballot version resulted in only around 1% of respondents refusing to state their voting intentions. Since Table 5 shows that neither of the direct methods of administering the voting intention question resulted in more than a 2% refusal rate, this problem, at least, seems unlikely to affect the final estimates greatly, whatever approach is used.

Nor does Perry's argument that interview behaviour which simulates actual voting behaviour, will produce more accurate estimates, appear to be supported by these results. Given the apparent similarity between the secret ballot procedure and casting a vote, this finding is difficult to explain. One possibility, which further research could explore, is whether respondents' perceive secret ballot polling to be analogous to actual voting behaviour, since it is possible that the assumed relationship exists only in the minds of researchers.

If respondents who state they do not plan to vote for any party are regarded not as intending non-voters, but as polite refusers, the effects of a secret ballot are more pronounced. Comparison of Tables 5 and 10 shows a decline of 4.3% (6.7% to 2.4%; $p < .05$) when a secret ballot was used in conjunction with rapport developing questions, and 3.6% (7.5% to 4.1%; $p < .05$) when it was administered directly.

Overall, the reductions in the size of the undecided group, of 11.1% (from 13% to 1.9%) in the contextual versions, and 18.6% (from 22.8% to 4.2%), in the non-contextual versions are quite unambiguous, and are clearly consistent with the expected pattern. Yet regardless of these effects, the fact remains that reducing the number of undecided respondents to a negligible proportion of the sample did not improve the accuracy of the resulting estimates. The remainder of this section explores why the hypothesised results did not eventuate.

Recent work conducted by Sanchez & Morchio (1992) may provide some insights into these apparently discrepant results. Their work investigated the effects of detailed probing, used to ascertain the direction in which declared undecided respondents might lean. While noting that probing might prompt some individuals to voice an opinion, or state their intention, they also concluded that respondents' choice, in these circumstances, involved a substantial degree of guesswork.

Although Sanchez & Morchio's work examined these effects as they occurred in telephone interviewing, it is possible that contextual questions, or the secret ballot method, or the combined effect of these, created an effect analogous to extensive probing. Interviewers administering the voting intention question via a secret ballot did not probe undecided respondents, since they did not know how respondents had completed their ballot paper. However, while interviewers could not probe, the ballot paper may have performed this task since it contained neither a non-voting nor an undecided option. Though speculative, it is possible that the lack of these options pressured respondents into selecting one of the parties named as the party for whom they planned to vote.

Respondents could have stated that they were undecided, or that they did not intend to vote, by ticking the "other" option and nominating an alternative intention. However, in the context of the ballot slip, the "other" category was clearly designed to enable respondents who planned to vote for a minor party to include their preferences. Using this option to specify not an alternative voting intention, but a lack of voting intention, or complete indecision, may have appeared to violate the norms of helpful respondent behaviour. While defining these norms is difficult, it seems likely that respondents could perceive that writing something other than the name of another party not already listed, would not meet the interviewer's, or the researcher's

expectations. In order to remain co-operative, respondents may have specified a voting intention, even though this may have been only vaguely formed in their mind.

As with the explanations advanced in Chapter 8, this reasoning is speculative, and there is no way of verifying what respondents had in their mind as they completed the survey. Nor does the comparatively thin body of literature on confidentiality measures offer an obvious, or even a plausible, explanation. While Singer's work suggests confidentiality assurances may improve responses where the data sought are sensitive, she did not explore secret ballots specifically. Even the work on conversational logic which might be brought to bear here, remains at a rudimentary state (see Schwarz *et al* 1991; Gendall, Carmichael & Hoek, under review), and cannot be applied readily to this situation.

Overall, these data do not allow a more precise definition of the extent to which the problem Sanchez & Morchio discussed affected these estimates, if it affected them at all. Irrespective of the possible explanations advanced, the fundamental problem of whether reducing the size of the undecided group will improve the accuracy of the predictions obtained has not been solved by the use of a secret ballot technique. Unless the results obtained using the secret ballot technique are significantly more accurate than those elicited via more direct questioning, the technique remains cumbersome and expensive to administer, disadvantages which do not appear to be outweighed by improvements in the accuracy of the estimates.

Alternatively, it is possible that a secret ballot was not an appropriate way of increasing the confidentiality associated with the interview. However, even if other measures, such as a more lengthy verbal assurance, were employed, the research suggests these measures have only minor effects and, in some cases, are counter-productive (Singer *et al*, 1992; Singer *et al,* 1995).

In the light of the results reported in Chapters 8 and 9, it seems prudent to examine whether the treatments examined thus far are the most sensible options to pursue. An additional explanation therefore, is that neither contextual effects nor secret ballots will improve respondents' predictions if the voting intention question itself is unable to capture the complexity of respondents' answers. That is, requiring an absolute

choice of respondents may not be the optimal method of eliciting voting intentions. Chapter 10 examines this question, and a method of addressing it, in more detail.

9.5 Summary

This chapter examined whether the use of a secret ballot would reduce the size of the undecided group and, in so doing, improve the accuracy of the estimates produced. While there was no doubt that the secret ballot methodology resulted in significant reductions in the undecided group, (when compared to estimates resulting from direct administration of the voting intention question) the data again provided no evidence to suggest this was followed by a decrease in the errors associated with the resulting voting intention estimates. This finding corroborated the conclusion put forward in Chapter 8 and suggests these treatments have more ambiguous effects than have hitherto been recognised.

Although potential explanations of these results were outlined, the data provide no firm evidence to support the hypothesis that the equivalent of extensive probing may have led respondents to guess at their likely intentions, thereby diminishing the accuracy of these resulting estimates. However, they do confirm Singer *et al*'s (1995) findings that assurances of confidentiality may have little overall effect, even where the data sought are sensitive.

Overall, the secret ballot methodology would appear to have little to offer pollsters. Reliant on face-to-face interviewing for its administration, the secret ballot incurs high fieldwork costs which are not repaid in terms of improved estimates. Based on these results, employing a secret ballot technique in this context would appear unlikely to result in any demonstrable advantages for pollsters.

CHAPTER 10

RELATIVE PROBABILITIES: RESULTS AND DISCUSSION

10.1 Introduction

Previous chapters have documented two attempts to reduce the proportion of undecided respondents. The first explored the effects caused by contextual questions, which aimed to enhance respondents' recall and help them to consolidate their intentions. The second attempted to increase the confidentiality associated with respondents' answers to achieve the same end. While these treatments, whether administered singly or in unison, significantly lowered the proportion of undecided respondents when compared to the direct, non-contextual questionnaire, none increased the accuracy of the resulting estimates.

Given that these findings are inconsistent with the dominant views expressed in the literature, it seems timely to turn to examine a more fundamental issue, such as whether the question itself contributes to the inaccuracies noted. If this is so, reducing the proportion of undecided respondents by using contextual or confidentiality devices may make no greater contribution than guesswork would, and may even detract from the estimates' accuracy.

Although simple proportional allocation of undecided and non-voting respondents has produced the most accurate estimates to date, this procedure has, in other circumstances, resulted in quite inaccurate predictions. Even if these inaccuracies had not occurred, both pollsters and researchers would presumably prefer a method which eliminated the need to reallocate these groups, and which simultaneously improved their predictions. Since neither changes in the context surrounding the questions, nor the means by which interviewers administer the questions, appear to have achieved this end, logic suggests that the question format itself may be flawed.

As phrased, the voting intention question requires respondents to make an absolute choice, a fact which appears likely to contribute to the proportion of undecided respondents. An alternative to reducing this group by either providing a more helpful route to the key behavioural questions, or a more sophisticated confidentiality mechanism, is therefore to change the question from one requiring an absolute answer, to one which enables respondents to express their likely behaviour in relative terms.

This chapter therefore explores the effect a probability scale had on respondents' ability to predict their behaviour. The chapter begins by comparing respondents' predicted behaviour to the actual electorates' behaviour before examining the sub-samples' predicted and self-reported behaviour. Next, the effects of weighting are discussed before the chapter turns to explore these findings in the context of earlier work.

10.2 Effects of a Probability Scale on Voting Intention Estimates

The questionnaires used to explore the effects of a probability scale on respondents' estimates collected details of respondents' probability of voting for each of the major parties represented within the electorates surveyed. Thus instead of nominating the party for whom they would have voted, had the election been held yesterday, respondents used the Juster Scale to indicate the chance that they would have voted for each of the parties contesting the election, had the election been held yesterday.[49]

The sum of the probabilities respondents allocated to each party should have equalled their overall probability of voting. However, in many cases the sum of respondents' party probabilities was greater than one (a logical impossibility). The estimates have therefore been re-calculated to ensure the overall electorate estimates did not exceed 100%. (Table 19, presented in section 10.3, includes a specific re-calculation of the estimates weighted by respondents' overall probability of casting a vote.)

The two questionnaire versions tested differed according to the placement of contextual questions, and so replicated the treatment discussed in Chapters 8 and 9. Table 15 contains the aggregate results relating to these questionnaires. Because this method effectively eliminated the undecided group, the format of Table 15 varies from that of the corresponding tables presented in Chapters 8 and 9.

---

[49] The order in which the parties were presented to respondents was rotated to avoid item-order bias.

Table 15: Effects of a Probability Scale on Voting Estimates

	Contextual Questionnaire Version (N=679)[1]				Non-Contextual Questionnaire Version (N=678)[1]			
Party Support	**Raw Estimates**	**Re-Allocated to 100% base**	**Actual Result[2]**	**Error**	**Raw Estimates**	**Re-Allocated to 100% base**	**Actual Result[2]**	**Error**
	%	%	%	%	%	%	%	%
Alliance	19.6	16.6 *(1.7)*	17.5	-0.9	19.1	16.7 *(1.7)*	17.5	-0.8
Labour	42.5	35.9 *(2.2)*	39.6	-3.7	41.2	35.9 *(2.2)*	39.6	-3.7
National	34.9	29.4 *(2.3)*	36.0	-6.6[3]	33.6	29.3 *(2.3)*	36.0	-6.7[3]
NZ First	15.6	13.2 *(1.6)*	4.3	+8.9[3]	14.6	12.7 *(1.6)*	4.3	+8.4[3]
Other	5.8	4.9 *(1.0)*	2.6	+2.3[3]	6.2	5.4 *(1.0)*	2.6	+2.8[3]
TOTAL	**118.4**	**100.0**	**100.0**	**4.5[4]**	**114.7**	**100.0**	**100.0**	**4.5[4]**

1. Because examining the results in the context of each electorate would have resulted in very small cell counts, these results are aggregated across the six surveys and represent the total number of responses elicited using versions 5 and 6 of the questionnaire. The figures in parentheses are the standard errors associated with the estimates.
2. The actual results were calculated by summing the actual vote distributions across all three electorates; because no interviewing was conducted in Whitby, part of the Western Hutt electorate, the results from polling booths located within Whitby were excluded from this aggregation (see Chapter 6 for further discussion of this).
3. Probability of accepting the null hypothesis when it was false exceeds 5%.
4. Mean absolute error.

Table 15 shows how similar the results from the contextual and non-contextual versions of the questionnaire are. The fact that both have the same mean absolute error suggests that the contextual questions had no obvious effect on the accuracy of the estimates. Both also correctly predicted that the Labour Party would gain highest proportion of the votes cast.

The mean absolute errors (both 4.5%) suggest this method was less accurate than the other treatments tested. More detailed examination of the error patterns shows that, unlike the other questionnaires, the Juster Scale questionnaires provided very accurate estimates of the Alliance's share of vote. However, both Juster Scale questionnaires underestimated the incumbent National government vote while overestimating that of its philosophical counterpart, the NZ First Party. The methods also underestimated the main Labour opposition's share of vote, while overestimating that which minor parties would receive. The probability of accepting the null hypothesis (that the difference between the estimates and the actual results was zero) when it was false exceeded 5% for several of these estimates. As a consequence, the mean absolute error associated with these versions is considerably higher than that found in Tables 5 and 10.

The data themselves do not reveal why these errors occurred, though it is possible that this method made it easier for respondents to express anti-incumbent feelings. That is, by allocating probabilities greater than zero to the minor parties, respondents may have vented their frustration with government in general, a quite different task to expressing their probabilities of voting for the various parties. Although speculative, these, and other possibilities, are explored further in section 10.4.

Table 16, below, examines the extent to which this pattern of under-estimation of the major parties and over-estimation of minor parties was manifested in individual electorates. (Appendix G contains the full tables from which the mean absolute errors in Table 16 were taken.)

Table 16: Effects of a Probability Scale on Mean Absolute Errors - Electorate Level

Electorate	Contextual Version		Non-Contextual Version	
	N	Mean Abs. Error	N	Mean Abs. Error
Manawatu 1	105	6.5	97	6.9
Miramar 1	98	6.6	96	5.9
Western Hutt 1	126	5.0	131	4.4
Manawatu 2	124	5.1	116	4.2
Miramar 2	112	6.9	112	5.5
Western Hutt 2	125	4.7	126	4.6

Table 16 confirms that the mean absolute errors presented in Table 15 were not a consequence of a varied pattern of error in the different electorates. In virtually every instance (three exceptions were noted in the twenty four comparisons; $p < .0003$), the estimates produced using the Juster Scale were less accurate than those elicited via the absolute choice question (irrespective of its position in the questionnaire or its method of administration). This suggests that the relative choice question based on the Juster Scale was significantly less accurate than the absolute choice questions discussed in Chapters 8 and 9. However, despite the higher error levels, Appendix G reveals that for all but one prediction each, the Juster Scale questionnaires correctly predicted the party that would win each electorate, even if the magnitude of the win was not accurately estimated.

Examination of the detailed electorate results (contained in Appendix G) revealed a similar pattern in virtually every electorate: over-estimation of the minor parties, in particular the NZ First Party, and corresponding under-estimation of the major parties, especially the National Party. This pattern of error differs from that affecting the other treatments, which was more randomly distributed, and suggests this method of predicting voting behaviour may contain a bias working against the major parties in favour of the minor parties, which was not detected during the piloting work.

However, the individual electorate data also contained a point of similarity with the other treatments. The data lent further support to the idea that, in this situation at least, contextual questions do not improve the accuracy of estimates. Indeed, five of the six estimates resulting from the contextual questionnaire were less accurate than those resulting from the non-contextual version. Although no one electorate result is significantly less accurate, the chance that five of the six errors would be in the same direction approaches significance ($p = .094$).

In contrast to the results reported in earlier chapters, these results suggest that the timing of the interview may have had some effect on the accuracy of the estimates within the different electorates. Three of the six errors decreased slightly and two recorded a sharp decrease between the rounds of interviewing. While this in itself does not constitute strong support for the role played by unanticipated events, it does represent a variation from the previously reported findings.

As expected, the results in Table 16 confirm those contained in Table 15: both suggest the relative choice question has not provided the increase in accuracy hypothesised. However, it is possible that sampling error contributed to the observed errors. Though difficult to quantify the precise effects of sampling error, the following tables, which examine respondents' self-reported behaviour, provide an alternative assessment of the questionnaires' accuracy. Because respondents did not provide a single stated intention, as they did in response to the questionnaires discussed in Chapters 8 and 9, Table 17 examines the mean absolute errors associated with respondents' predicted and self-reported behaviour. For comparative purposes, the mean absolute errors are examined together with those associated with Tables 7 and 12.

Table 17: **Effects of a Probability Scale on Voting Intention: Predicted versus Self-Reported Behaviour**

CONTEXTUAL VERSION (N=342)

Predicted Voting	Predicted	Self-Reported	Error
	%	%	%
Alliance	16.9 *(2.5)*	18.2	-1.3
Labour	33.6 *(3.0)*	39.2	-5.5
National	32.3 *(3.3)*	37.7	-5.5
NZ First	12.8 *(2.2)*	3.4	+9.4[1]
Other	4.4 *(1.3)*	1.5	+2.9[1]
TOTAL	*100.00*	*100.00*	*4.9*

NON-CONTEXTUAL VERSION (N=333)

Predicted Voting	Predicted	Self-Reported	Error
	%	%	%
Alliance	17.2 *(2.5)*	19.2	-2.1
Labour	35.0 *(3.2)*	37.7	-2.7
National	30.5 *(3.3)*	37.3	-6.8
NZ First	12.8 *(2.3)*	3.6	+9.2[1]
Other	4.4 *(1.2)*	1.9	+2.4[1]
TOTAL	**100.0[2]**	**100.0[2]**	**4.6**

1. Probability of accepting the null hypothesis when it was false exceeds 5%.
2. Table may contain small rounding errors.

This table shows a similar overall error pattern to that noted earlier: the parties which attracted a higher proportion of the vote are consistently under-estimated while those with lower levels of support are over-estimated. This pattern is identical in both sets of results, suggesting that the contextual questions did little to correct or reduce this bias.

Looking more specifically at the individual party estimates and the associated prediction errors, Table 17 shows both versions predicted the Alliance and Labour share of vote quite accurately. The non-contextual version also predicted Labour's share of vote accurately, but neither version provided accurate predictions for the other parties.

Unlike Tables 7 and 12, the estimates for the two major parties were not more accurate than the estimates for other parties. This suggests that factors other than the recognised difficulty of predicting new behaviours were responsible for these errors.

Both Juster questionnaire versions resulted in greater mean absolute errors than did the other treatments, thus adding weight to the overall conclusion that these questionnaires resulted in less accurate estimates than the other questionnaires.[50]

To investigate further the distribution of errors, Table 18 contains the mean absolute errors that resulted after respondents' predicted and self-reported behaviour in each electorate were compared. Again, as can be seen from Table 18, the cell counts are very small, and the results are only presented to assess whether a particular pattern of error is apparent.

Table 18 **Effects of a Probability Scale Measures on Mean Absolute Errors: Predicted versus Self-Reported Behaviour**

Electorate	Contextual		Non-Contextual	
	N	Mean Abs. Error	N	Mean Abs. Error
Manawatu 1	47	4.4	46	5.0
Miramar 1	39	6.0	34	6.5
West Hutt 1	75	8.0	71	4.0
Manawatu 2	64	4.2	62	5.2
Miramar 2	49	7.9	46	8.0
West Hutt 2	68	5.1	74	5.4

Although the errors resulting from both the contextual and non-contextual versions were similar, the contextual version, with one notable exception, tended to result in

50 The contextual version errors from the direct and secret ballot questionnaires were 2.5% and 2.6% compared to the 4.9% reported here, while the errors from the non-contextual versions of these questionnaires were 4.0% and 3.0% compared to the 4.6% reported here.

lower mean absolute errors. Again, this pattern of results approaches significance ($p = .094$).

Comparison of the errors from the first and second rounds of interviewing suggested that most errors increased, some considerably. Overall, there was no evidence to suggest that the later round of interviewing resulted in more accurate estimates. This conflicts with the trend apparent in Table 16, but is more consistent with the results reported in Chapters 8 and 9.

In summary, the tables presented thus far provide strong evidence that the relative probability questions have elicited less accurate predictions. The reminder of this section examines possible explanations for this result.

The estimates presented in Tables 15 to 19 were not weighted in any way. If respondents who allocated a high probability of voting for a new party had a lower than average likelihood of casting their vote, the estimates would over-state minor parties' share of the vote. Weighting respondents' party support estimates by their probability of voting would address this problem and section 10.3, which examines weighting procedures, explores this idea in greater detail.

Second, respondents may not have made the trade-offs the question implicitly required of them. Instead of providing probabilities for each party which summed to their overall probability of voting, many respondents allocated party probabilities that summed to well over one. This suggests using a constant-sum technique to administer this question (for example, requiring respondents to allocate counters among the various parties instead of simply assigning verbal probabilities) would provide a more effective means of ensuring such a trade-off occurred.

The lack of consistency between respondents' party probabilities and their overall probability of voting suggests that instead of providing their probability of voting for the various parties, respondents may have answered a different (unintended) question, perhaps relating to their support for the various parties' policies. Although this explanation remains speculative and unproven, it would account for the over-estimates of support for the minor parties, which almost certainly have some appealing policies or characteristics, and the under-estimation of support for the incumbent opposition

and government, about whom respondents may have felt jaded.

An additional explanation may be the sympathy often extended to individuals or groups seen as underdogs. Thus respondents, perhaps reasonably certain of their likely voting behaviour, saw no harm in allocating a greater than zero probability to some of the minor parties. Although there is no substantive evidence to suggest this occurred in this research, the literature has identified an "underdog" phenomenon (see Ceci & Kain, 1982). Clearly if this behaviour had occurred, the consequence of a cumulative set of greater than zero probabilities allocated to minor parties would be an overestimation of their actual support. Again, while speculative, this explanation nevertheless accounts for the pattern of results observed and fits well with previously reported findings. This explanation also suggests respondents are not making the specific trade-offs between parties required of them although this could also be remedied through use of a constant sum procedure.

Overall, the Juster Scale based method resulted in the highest error levels with an apparent bias toward over-estimation of minor parties. Of the various explanations put forward to explain this finding, differing probabilities of voting amongst supporters of the various parties appears most plausible. The following section explores the effect of weighting the sum of each respondents' party probabilities by their overall probability of voting.

10.3 Effects of Weighting Procedures

As noted, the data presented in Tables 15 to 18 were unweighted. Table 19 compares the effects of three different weighting procedures to the unweighted mean absolute errors. Appendix H contains full details of the weighting procedures.

Table 19: Effects of Weighting Procedures on Mean Absolute Errors - Probability Scale Effects

Electorate	Contextual Unweighted	Contextual Demographics Weighted	Contextual Probability Weighted	Contextual Weighted by Past voting	Non-Contextual Unweighted	Non-Contextual Demographic Weighted	Non-Contextual Probability Weighted	Non-Contextual Weighted by Past Voting
Manawatu 1	6.5	6.8	6.2	6.5	6.9	6.7	5.8	5.0
Miramar 1	6.6	6.6	4.6	5.4	5.9	6.4	4.1	3.8
Western Hutt 1	5.0	5.0	3.8	3.4	4.4	3.8	3.0	4.2
Manawatu 2	5.1	5.1	4.7	3.2	4.2	4.5	4.0	2.4
Miramar 2	6.9	6.8	4.9	5.1	5.5	5.8	3.8	3.4
Western Hutt 2	4.7	4.9	2.7	2.1	4.6	4.8	2.9	3.6
Overall Mean	5.8	5.9	4.5	4.3	5.3	5.3	3.9	3.7

Table 19 contains similar results to Tables 9 and 14. First, the errors resulting from demographic weighting are either the same as, or higher than, those associated with the unweighted estimates. That all three treatments have shown this trend suggests that weighting voting intention data to conform to a population's age-gender ratio may introduce more error than it removes.

Not surprisingly, weighting respondents' predictions by their overall probability of voting decreased the mean absolute errors, especially in the non-contextual version. Since many respondents allocated party probabilities which summed to more than one (or, more accurately, to more than their probability of voting) and the scaling mechanism employed in Tables 15 to 18 was crude, this finding was predictable.

Weighting the data to reflect the pattern of voting at the 1990 General Election also decreased the mean absolute errors. Overall, the findings re-iterate the patterns and conclusions advanced earlier: weighting the data to conform to age-gender ratios is not supported by these findings, however, weighting by more behavioural measures, such as respondents' likelihood of voting, or their past voting behaviour, appears to offer more promise and merits more detailed investigation.

The remainder of this chapter explores the implications of these results in the light of earlier research, summarised and discussed in Chapters 2 and 4.

10.4 Relationship with the Literature

When Bogart (1967) argued that people may hold incongruent and inconsistent opinions simultaneously and that undecided responses may result from respondents' inability to rank order their preferences or opinions, he highlighted an inherent weakness of most voting intention questions. Although his idea does not appear to have been explored specifically in the context of voting behaviour, it provides a very plausible explanation of why undecided responses have plagued polling research.

Further, Bogart's argument relates to a more methodological issue over which researchers have disagreed: whether intention or probability scales provide the most accurate estimates of future behaviour. Since voting intention questions are analogous to dichotomous intention questions, they arguably compound respondents' difficulty in ranking the various behaviours they might perform. Despite the popularity of

intention scales in the past, therefore, more recently researchers have queried the posited intention-behaviour link, and have instead argued that intention may bear only a tenuous relation to subsequent behaviour (Foxall, 1982). This challenge implies that the intention based questions traditionally used by pollsters may be fundamentally flawed.

However, irrespective of the arguments advanced, or the available empirical evidence, these results show clearly that respondents' probabilities were not more accurate than the intentions data collected using other versions of the questionnaire. The results presented in Tables 15 to 18 reveal a consistent over-estimation of the smaller parties at the expense of the larger parties. This suggests the scale employed may contain a bias which contributes to the errors noted. The remainder of this section explores this issue further and suggests measures which may ameliorate its effects.

Juster (1966) noted a tendency for his scale to under-estimate actual purchase, though his work concentrated on one behaviour (such as the purchase of a car) rather than on a series of related but mutually exclusive behaviours (such as voting for political parties). In Juster's work, therefore, under-estimation of car purchases did not necessarily result in over-estimation of washing machine purchases, whereas under-estimation of support for one political party necessarily means over-estimation of support for another. That is, when using the Juster Scale to predict their likelihood of voting for a particular party, respondents must weigh this up against their chances of voting for all other parties represented. The pattern of error noted suggests some respondents, at least, did not compare their likelihoods in this way. A constant-sum approach would make this implied trade-off more explicit and remains an option for remedying this problem.

Examination of the probability scores allocated by respondents shows many people gave probabilities which summed to more than one, or to more than respondents' stated probability of voting. Clearly, respondents' probability of voting for each of the parties should sum to their overall probability of voting. That it so often did not, raises several issues.

First, respondents may have found it too difficult to allocate probabilities, an issue

debated by Hogarth (1975) and Seymour *et al* (1994), among others. The task required of them may have been too taxing and the inaccuracies in the results may therefore represent the inability of some respondents to come to terms with the scale itself. However, while it is possible that some respondents experienced difficulties, no interviewers reported more than trifling problems administering the scale. Furthermore, Justers' work, as well as that conducted more recently by Day *et al* (1991) and Brennan *et al* (1994), suggests the scale provides more accurate estimates than intention scales. If a sizeable proportion of respondents struggled to interpret and use the scale as required, it seems unlikely that replication studies would continue to confirm its predictive superiority over intention scales.

The second issue is the possibility that respondents answered a different question to the one the interviewer administered. Instead of recording their likely voting behaviour, respondents may have given some indication of their relative support for different parties and their policies. Although no qualitative work was conducted to test this view, it is plausible that respondents did not adequately trade-off the different competing behaviours. Respondents who had a high probability of voting for Party A, but who also supported some of Party B's policies, may have allocated it a greater than zero probability to reflect this support, even if their actual probability of voting for Party B was zero.

This explanation also recognises that respondents may use polls to send messages of approval or disapproval to the parties contesting the election. The data show that the party consistently over-supported was the New Zealand First party, and that this support came at the expense of the National Party. Since the New Zealand First party evolved because of disaffection with the National government, expression for support for it would represent an obvious way of registering disapproval of the government. Again, although this interpretation was not explored empirically, it is consistent with the patterns that emerged consistently in the results.

An additional explanation arises from comments Juster (1966) and Pickering & Isherwood (1974) made about the accuracy of probability scales when used to explore behaviours occurring within a compressed time frame. They suggested respondents may have provided inaccurate estimates because of difficulties in confining their

predictions to the specified time period. However, collecting predictions of consumers' likely purchase behaviours differs from likely voting behaviour as the former has only a loose time frame compared to the latter, where the behaviour explored can only occur on a specific day. Thus it seems unlikely that respondents in this survey would have experienced similar difficulties in envisaging the time-frame within which their behaviour would occur. Furthermore, the question posed investigated voting behaviour retrospectively, thus reducing the likely effects any problem of envisaging the relevant time period may have had.

All researchers involved in the development and refinement of probability scales acknowledge the problems caused by unanticipated events, which occur between the interview and the time at which the behaviour explored is expected to occur. However, the majority of the results presented in this and previous chapters have shown that the earlier estimates were typically more accurate than those taken closer to the election date. While this does not eliminate the influence of unanticipated events, it does sugest that the relationship between these and the accuracy of the estimates maybe more complex than previously thought.

The question of whether unanticipated events intervene between the interview and the predicted behaviour again raises the question of whether polls provide "snapshots in time" or whether they attempt to predict specific outcomes. Juster's original argument in support of his scale depended not on the scale's ability to predict individuals' behaviour, but on the accuracy of the aggregate estimates it produced. If the effects caused by unanticipated events occur randomly, they could be expected to disrupt certain individual's predictions (i.e., those specifically affected by the events), but those effects should average out across the entire sample, leaving the aggregate estimate of behaviour unaffected. However, if the effects caused by unanticipated events do not occur randomly, such averaging cannot be expected. In the case of election campaigns, political advertising and policy promulgation intensifies during the official campaign period, however, the similarity of the estimates obtained prior to and during this campaign period suggests this activity had little real effect on respondents' behaviour. While unanticipated events cannot, therefore, be ruled out as sources of error, their likely effect appears quite limited, providing further support for the view that polls, like other surveys exploring behaviour, are concerned with predicting certain outcomes.

Gabor & Granger (1972) suggested that the object of the behaviour (whether high or low ownership) may also explain inaccuracies in the estimates. Although respondents do not have ownership of political parties in the same way as they own consumer durables, they nevertheless have patterns of behaviour associated with political parties and these are arguably analogous to their behaviour patterns with respect to consumer durables. Gabor & Granger found that the estimates associated with high ownership items were more accurate than those associated with low ownership items. However, while respondents gave quite accurate probabilities of voting Labour, their estimates of voting for the other well established and dominant party, the National Party, were less accurate. The frequency with which respondents have performed a particular behaviour does not, therefore, appear to account for all the error observed.

More recent research (see Day *et al*, 1991; Brennan *et al,* 1994) has also noted that the Juster Scale has a tendency to over-estimate actual purchases, especially where the brands explored are not well established. Again, this is analogous to the results reported here, which showed consistent over-estimation of the minor parties contesting the election for the first time. Roshwalb (1975) (in his discussion of the heuristics Tversky & Kahneman, 1973 & 1974, suggested people adopt to simplify the task of assessing the likelihood of uncertain events occurring) noted that people overestimated events with a low probability of happening and underestimated those with a high probability of occurring. Why this should occur, Roshwalb did not hypothesise, however, his discussion encapsulates comments made independently by Juster (1966), Gabor & Granger (1972) and Pickering & Isherwood (1974), and helps explain the pattern of results observed in this study.

While many researchers have noted similar error patterns to those observed in this study, few have actively researched means of remedying these. The most obvious solution, a constant-sum approach, has been examined by Hamilton-Gibbs *et al* (1991) who reported it produced more accurate estimates than the direct question method of administering the Juster Scale. That this was so is unsurprising, since a constant-sum approach forces a direct trade-off between competing alternatives. Panagakis (discussed in Crespi, 1988) also experimented with a constant-sum treatment, though his work attempted to measure respondents' attitudes rather than their behaviour, and he does not appear to have validated his estimates by comparing them to behavioural

data.

Given the range of explanations put forward to explain the larger than expected errors in the Juster Scale based estimates, a constant-sum approach appears to provide a means of resolving the potential problems noted, if indeed these were responsible for the errors.

Overall, the results reported here display similar characteristics and patterns to those noted in earlier research investigating purchase behaviour. While the mean absolute errors were greater than hypothesised, some explanations for this emerged. The final table presented suggests that two weighting procedures improved the accuracy of the estimates considerably, and brought the mean absolute errors in to line with those associated with the contextual and confidentiality treatments. Before dismissing this method of predicting voting behaviour, it would seem logical and prudent to explore further the constant-sum method of administering the question. Further exploration of weighting variables based on respondents' behaviour also merits close attention.

10.5 Summary

Although several arguments pointed to failings in the voting intention question, the alternative explored and reported on in this chapter provided significantly less accurate estimates than the treatments documented earlier. Of the explanations examined, failure to weight respondents' estimates by their probability of voting clearly accounted for a considerable proportion of the error noted. In addition, the method of administering the probability questions may have provided respondents with an opportunity to report something other than their probability of voting for a particular party.

Given its generally strong performance when used to predict other behaviours, the Juster Scale would appear to merit further consideration before being rejected as a potential tool in political opinion polling. In particular, future research could explore its administration using a constant-sum technique.

CHAPTER 11
CONCLUSIONS

11.1 Conclusions

This research explored a new method of estimating voter turnout and examined how three methods of asking or administering voting intention questions affected the proportion of undecided respondents and the resulting survey estimates.

The first research objective was to explore whether an eleven-point purchase probability scale, the Juster Scale, could provide accurate estimates of voter turnout. Irrespective of the measure used to assess accuracy, the results show that the Juster Scale provided consistently accurate estimates of turnout. However, since a high proportion of New Zealanders do vote, the situation in which the scale was tested may not have provided a rigorous test of its accuracy. Clearly, where a high proportion of a sample perform a behaviour, the potential for error when estimating this proportion is likely to be lower than when estimating the proportion performing a low incidence behaviour. Consequently, although the Juster Scale performed well in this context, its predictive ability will be better established in situations where turnout is lower than 80%. Though there is need for replication research, these findings are promising and offer the potential for improving turnout estimates.

The remainder of the thesis explored the related difficulty of predicting people's voting behaviour. In particular, the research examined whether various question wording or administration treatments, designed to reduce the proportion of undecided respondents, would produce more accurate estimates.

The first of these treatments involved the use of contextual questions. Although the general question wording literature suggests that contextual questions improve respondents' ability to answer subsequent questions, the literature dealing specifically with voting intention questions is more ambivalent. At least some political polling researchers have noted that contextual questions may lead respondents to provide less accurate answers than they would otherwise have done.

The results reported here confirm that contextual, or rapport-developing, questions will significantly reduce the proportion of undecided respondents. However, respondents

asked to discuss their interest in politics and their views on various policy issues prior to stating their voting intention consistently provided less accurate voting intention estimates. Instead, the questionnaire versions containing a minimal context - two behavioural questions - produced more accurate voting intention estimates. However, rather than avoiding the use of contextual questions completely, pollsters could consider using questions designed to explore pre-cursor or historical behaviours (such as respondents' registration status, or their past voting behaviour) before eliciting respondents' voting intentions, since these produced the most accurate estimates in this study. Overall, therefore, while contextual questions had a dramatic effect on the proportion of undecided respondents, they did not improve the accuracy of the estimates obtained.

The second research objective was to examine whether enhancing the confidentiality associated with respondents' answers would reduce the proportion of undecided respondents and improve the accuracy of the resulting estimates. Evidence presented in the literature suggested that the improved confidentiality offered by secret ballots would reassure privacy-conscious individuals who might otherwise refuse to answer the question or declare themselves undecided. The results in Chapter 9 confirm that secret ballots can reduce the number of undecided respondents to a negligible proportion of the sample, especially when used in conjunction with contextual questions. However, despite this reduction in the size of the undecided group, the accuracy of the voting intention estimates did not improve. Consequently, the third research hypothesis, that the accuracy of voting intention estimates elicited using a secret ballot would be greater than those elicited by more direct questions, was rejected.

The final research objective explored whether analysing respondents' probability of voting for each party would produce more accurate estimates than analysis of their voting intentions. The Juster Scale was used to elicit these probabilities because of its superior performance in estimating consumers' purchase behaviour. This method allowed respondents to express their behaviour in relative terms and so eliminated the undecided group completely. Because of this total elimination of the undecided group, this method was expected to produce the most accurate estimates.

However, contrary to expectations, the Juster Scale estimates had the highest mean

absolute errors. This failure to predict voting behaviour accurately appeared to be due to an inherent bias in the question which resulted in the consistent over-estimation of minor party support and a corresponding under-estimation of the major parties, especially the incumbent National party. One explanation of this bias is that respondents answered a broader question, relating to overall party support, instead of addressing the specific issue of their likely behaviour that the question was designed to explore. However, this is only a tentative interpretation of these results, not a substantive conclusion, and the results provide further support for the conclusion that elimination of the undecided group will not necessarily result in more accurate estimates.

As well as analysing the unweighted estimates, the research also explored the effects on the data of various weighting procedures. In almost every case, weighting to ensure the samples matched the age-gender distribution of the electorate from which they were drawn decreased the accuracy of the resulting estimates. However, weighting the data by respondents' probability of voting, or weighting to ensure the past voting behaviour of the sample reflected the previous election result, produced similar or more accurate estimates. Given that this pattern appeared in all six surveys conducted, it would seem logical to defer the use of age-gender weights until their relationship with voting behaviour is more apparent. However, these results suggest that the use of more behaviourally-oriented weighting procedures merits further investigation.

Overall, the results show that the expected relationship between a reduction in the proportion of undecided respondents and an improvement in the accuracy of the survey estimates did not eventuate. Further explanations of why the expected pattern did not appear are discussed in sections 11.2 and 11.3.

11.2 Limitations of the Research

The limitations of this research fall into three general categories: the adequacy of the test, potential sources of error, and the design of the treatments, all of which are examined in detail below.

11.2.1 Adequacy of the Test

Whether the testing of a hypothesis is adequate depends largely on the rigour of the test employed. In the case of predicting voter turnout, the test was less rigorous than desirable because a high proportion of the New Zealand population votes in each General Election. As noted in earlier chapters, the reliability of the Juster Scale as a predictor of turnout requires further exploration in situations characterised by lower turnout ratios.

However, these findings provide strong support for such replication research and offer some promise, despite their limitations. Furthermore, the research tested the Juster Scale on six independent data sets, and so goes some way towards satisfying Ehrenberg & Bound's (1993) criterion that conclusions should be based on results evidenced across many sets of data as opposed to those arising from a single set. Of the six turnout estimates, only one was sufficiently different to the actual turnout to raise the probability of accepting the null hypothesis (that the Juster Scale was a poor method of predicting turnout) when it was false above 5%.

The second reservation about the adequacy of the test applied affects all four research objectives. Ideally, respondents' estimated behaviour would have been compared to their actual behaviour, for only this comparison would have enabled a true assessment of the various treatments' effects. Since access to this information was denied, the only remaining options were to compare the survey estimates to the actual electorate results, and to compare respondents' predicted behaviour to their self-reported behaviour. Unfortunately, neither of these comparisons is completely adequate. While the former necessarily entails some consideration of sampling error, the latter measure is prone to social desirability bias and bandwagon effects, the possible consequences of which are discussed below.

11.2.2 Sources of Error

Numerous researchers have documented the effects of social desirability bias, which occurs when respondents over-report behaviours such as voting (and under-report behaviours such as drinking or smoking). Chapter 7 included a discussion of this problem, which appears to have exaggerated the error in tables comparing respondents' predicted turnout to their self-reported behaviour. Although the net effect of social desirability bias is that the turnout estimates may be more accurate than the tables

suggest, the data do not allow a precise calculation of the effects social desirability bias may have had.

Some respondents may fail to provide truthful responses because of a phenomenon known as the bandwagon effect, which is manifested in more people claiming to have supported successful candidates than actually voted for them. However, the literature concerning these effects is rather ambivalent, and there was no strong evidence of bandwagon effects having influenced these results.

A more intractable limitation is the effect the experience of being interviewed had on respondents' subsequent behaviour. Several researchers have argued that Heisenberg's Uncertainty Principle may also apply to survey research. In other words, that the very measurement process may change the phenomena being measured.

Although the main reported effect of interviewing has been to increase respondents' turnout behaviour, there was no evidence that respondents interviewed in this study were, as a consequence, more likely to cast a vote. The results presented in Table 1 show a random pattern of error. These results are not surprising, since it seems logical that this problem will be greatest where turnout is traditionally low and the potential to affect respondents' behaviour is correspondingly greater.

Since respondents' actual voting behaviour is unknown, it is not possible to assess how their interview experience affected their actual party support behaviour. However, as there is no evidence to suggest it affected their turnout behaviour, it seems reasonable to conclude that any effect on their party support behaviour would be slight.

Another potential source of error is the response rate to the face-to-face survey. Fewer than 60% of those respondents originally sought or contacted actually participated in the survey, a fact which raises the possibility that non-response error distorted the survey estimates. Without contacting and interviewing the 40% of people who could not, or would not, participate, the true extent of non-response error will remain unknown. However, Gendall & Davis (1993), and more recently Hosie (1995), have suggested that estimates based on a 50% response rate do not change significantly when the remaining respondents' views are included in the estimates. The implication of these findings is that with the 60% response rate achieved, it is

unlikely that the estimates reported would be seriously affected by non-response error.

11.2.3 Design of the Research

The remaining research limitation is endemic to methodological research of this type and relates to the selection of the test variables. The question wording variations explored in this research were chosen after an extensive perusal of the literature. However, all methodological research must address the question of whether alternative variables or treatments might have provided more interesting results. It is not possible to establish *a priori* whether the contextual questions chosen were the most appropriate ones, although their widespread use in analogous questionnaires suggests other researchers believe them suitable. Similarly, there were a number of alternatives to secret ballots which could have enhanced the confidentiality of respondents' answers. Finally, the Juster Scale can also be administered in different ways. Thus at each stage of the research, the selection of certain treatment variables necessarily precluded others. While all researchers must establish some boundaries by deciding which specific variables to test, these decisions always contain some subjective elements. Consequently, the conclusions reported here relate to one interpretation and expression of specific treatments, and further research is required to investigate variations on the themes explored here. Section 11.3 develops these future research directions more fully.

11.3 Future Research

Replication research designed to investigate the Juster Scale's accuracy as a predictor of turnout has already been discussed in sections 11.1 and 11.2. The remainder of this section therefore examines how future research might extend knowledge of the relationship between question wording, the proportion of undecided respondents, and the accuracy of voting intention estimates.

Chapter 8 revealed that, when only specific behavioural questions preceded the voting intention questions, the resulting estimates were more accurate than when a mixture of attitudinal and behavioural questions preceded the voting intention question. Aside from further corroboration of this finding, future research could explore the effects additional questions about antecedent behaviours would have on the estimates. For example, more specific details of respondents' voting history and party affiliation

could be sought.

It is always possible that the particular attitude and opinion questions used to provide a context for the behavioural questions contained flaws that ultimately produced the errors observed. Although the widespread use of these questions suggests other researchers have not found this to be so, subsequent research could also investigate whether employing other attitudinal questions produces more accurate final estimates.

Despite the dramatic reduction in the proportion of undecided respondents when it was used, the secret ballot did not produce more accurate voting intention estimates. This naturally raises the question of whether it is the most appropriate means of increasing the confidentiality of a poll. Although this interview method is most directly analogous to the behaviour investigated, it may have forced respondents to state an intention where one did not exist. For this reason, further research could explore alternative confidentiality devices, such as the more extensive reassurances Singer and her colleagues have reported using. Variations of the secret ballot slip, such as including a specific undecided category on it, could also be explored. Future research could also amalgamate aspects of direct and secret ballot administration of the questions, by using a secret ballot only with respondents who declared themselves undecided when administered the voting intention question directly.

An alternative possibility is that respondents did not view the secret ballot interview as analogous to voting. Future research could employ cognitive laboratories as a means of identifying interview characteristics which respondents believe closely mirror the actual voting situation.

Although this research did not establish the superior predictive ability of a probability-based method of predicting voting behaviour, the possible explanations provided for this are plausible and also merit further investigation. Further research could therefore explore the use of a constant-sum approach and, in particular, could extend Brennan *et al*'s work which examined a constant sum method of collecting probabilities by telephone interview. Further refinement to the question itself, to make it explicit that the question seeks respondents' likely behaviour and not their philosophical sympathies, would also seem useful. In addition, the temporal setting of the question

could be varied to allow comparison of retrospective and projected behavioural estimates.

Although pursuing all these lines of enquiry would be interesting, further exploration of alternative questions, especially Juster Scale-based questions appears to offer most potential. However, until these are consistently and empirically established as constituting a superior prediction method, researchers should continue to use intention questions in conjunction with behavioural contextual questions and undecided respondents should be allocated proportionately.

11.4 Summary

The results presented in earlier chapters suggest that the Juster Scale can provide accurate turnout estimates. However, the results did not support the hypothesised relationship between the proportion of undecided respondents and the accuracy of the estimates. In fact, the questionnaire which resulted in the highest proportion of undecided respondents also had the lowest mean absolute error.

The use of specific behavioural questions prior to the voting intention questions was associated with more accurate results than the use of more wide-ranging contextual questions. Neither a secret ballot questionnaire nor a Juster Scale-based questionnaire resulted in more accurate estimates.

Several potential errors were discussed; of those whose effects can be gauged, none appeared to have seriously affected the estimates reported. The chapter then presented further lines of enquiry and placed particular emphasis on further experimentation with question wording.

This thesis has made two contributions to election prediction knowledge. First, it has identified a promising means of estimating turnout which is less complex and more parsimonious than other methods proposed. Second, the comprehensive review of factors thought to be related to the proportion of undecided respondents has confirmed the existence of these relationships, but has queried their effect on the accuracy of the resulting estimates. While the posited relationship between the level of undecided respondents and the accuracy of the estimates was not confirmed, the research identified a number of contributing factors which merit more detailed examination.

Further investigation of these, as well as replication of this study in different circumstances, will continue to add to knowledge of the factors affecting polling accuracy and may, in time, lead to the development of *" the psephological equivalent of laws of behaviour*" (p. 256) for which Jowell *et al* (1993) called.

REFERENCES

Abrams, M. (1970). The opinion polls and the 1970 British General Election. *Public Opinion Quarterly*, 34, 317-324.

Adamany, D. & Dubois, P. (1973). The "forgetful" voter and an under-reported vote. *Public Opinion Quarterly*, 37, 227-231.

Adams, J. (1956). An experiment in question and response bias. *Public Opinion Quarterly*, 20, 593-598.

Alsagoff, J., Gendall, P. & Esslemont, D. (1986). *The precision of omnibus survey estimates in New Zealand*. Market Research Centre Report No. 46.

Anderson, J. (1985). A theory for attitude and behaviour applied to an election survey. *Behavioural Science*, 30, 219-229.

Antaki, C. (1985). Attribution and evaluation in ordinary explanations of voting intention. *British Journal of Social Psychology*, 24, 141-151.

Axelrod, J. (1968). Attitude measures that predict behaviour. *Journal of Advertising Research*, 8 (1), 3-17.

Bartels, L. & Broh, A. (1989). The polls - A review. *Public Opinion Quarterly*, 53, 563-589.

Bassili, J. (1993). Response latency vs uncertainty as indexes of the strength of voting intentions in a CATI survey. *Public Opinion Quarterly*, 57, 54-61.

Becker, S. (1954). Why an order effect? *Public Opinion Quarterly*, 18, 271-278.

Belson, W. (1968). Respondent understanding of survey questions. *Polls, 3, 1-13.*

Belson, W. (1981). *The Design and Understanding of Survey Questions.* Aldershot, England: Gower.

Beniger, J. (1979). Avoiding interpersonal comparison of preference intensity: A rank distance approach to attitudinal scaling. *Political Methodology,* 6 (3), 247-274.

Benson, L. (1941). Studies in secret-ballot technique. *Public Opinion Quarterly*, 5, 79-82.

Bishop, G. (1990). Issue involvement and response effects in public opinion surveys. *Public Opinion Quarterly*, 54, 209-218.

Bishop, G., Oldendick, R. & Tuchfarber, A. (1980a). Experiments in filtering political opinions. *Political Behaviour*, 2, 339-369.

Bishop, G., Oldendick, R., Tuchfarber, A. & Bennet, S. (1980b). Pseudo-opinions on political affairs. *Public Opinion Quarterly*, 44, 198-209.

Bishop, G., Oldendick, R. & Tuchfarber, A. (1984). What must my opinion be if I just told you "I Don't Know"? *Public Opinion Quarterly*, 48, 510-519.

Blakenship, (1943). Consumer and opinion research. *Harper: New York.*

Bogart, L. (1967). Non opinion, don't know, and maybe no answer. *Public Opinion Quarterly*, 31, 331-345.

Bogart, L. (1985). Polls and public opinion. *Society*, 23, 16-25.

Bolstein, R. (1991). Comparison of the likelihood to vote among preelection poll respondents and nonrespondents. *Public Opinion Quarterly*, 55, 648-650.

Borelli, S., Lockerbie, B. & Niemi, R. (1987). Why the Democratic-Republican partisanship gap varies from poll to poll. *Public Opinion Quarterly*, 51, 115-119.

Bowden, R. (1987). Repeated sampling in the presence of publication effects. *Journal of the American Statistical Association*, 82 (3980), 476-484.

Bradburn, N. (1992). Presidential address: A response to the nonresponse problem. *Public Opinion Quarterly*, 56, 391-397.

Bradburn, N. & Mason, W. (1964). The effect of question order on responses. *Journal of Marketing Research*, 1, 57-61.

Bradburn, N. & Sudman, S. (1979). *Improving interview method and questionnaire design.* San Francisco: Jossey-Bass.

Bradburn, N. & Sudman, S. (1988). *Polls and surveys: Understanding what they tell us.* Jossey Bass: San Francisco.

Bradburn, N., Sudman, S., Blair, E. & Stocking, C. (1978). Question threat and response bias. *Public Opinion Quarterly, 42, 221-234.*

Brennan, M. & Esslemont, D. (1994). The accuracy of the Juster Scale for predicting purchase rates of branded, fast moving consumer goods. *Marketing Bulletin*, 5, 47-52.

Brennan, M. Hini, D. & Esslemont, D. (1994). Obtaining purchase probability data via telephone surveys: A preliminary test of two techniques. *Marketing Bulletin*, 5, 64-70.

Brook, D. & Upton, G. (1974). Biases in local government elections due to position on the ballot paper. *Applied Statistics*, 23, 414-419.

Brudney, J. (1982). An elite view of the polls. *Public Opinion Quarterly*, 46, 503-509.

Brusati, E. (1987). Electoral polls in Italy. *European Research*, August, 152-157.

Buchanan, W. (1986). Election predictions: An empirical assessment. *Public Opinion Quarterly*, 40, 222-227.

Butler, D. (1994). *The opinion polls and the 1992 General Election: A report to the Market Research Society.*

Byrnes, J. (1964). Consumer intentions to buy. *Journal of Advertising Research*, 4, 49-51.

Cahalan, D. (1968). Correlates of respondent accuracy in the Denver validity survey. *Public Opinion Quarterly*, 32, 607-621.

Campbell, J. & Wink, K. (1990). Trial-heat forecasts of the presidential vote. *American Politics Quarterly*, 18, 251-269.

Ceci, S. & Kain, E. (1982). Jumping on the bandwagon with the underdog: The impact of attitude polls on polling behaviour. *Public Opinion Quarterly*, 46, 228-242.

Chandler, M. (1988). Models of voting behaviour in survey research. *Synthese*, 76, 25-48.

Chapman, R. & Palda, K. (1983). Electoral turnout in rational voting and consumption perspective. *Journal of Consumer Research*, 9, 337-346.

Clancy, K. & Garsen, R. (1970). Why some scales predict better. *Journal of Advertising Research*. 10 (5), 33-38.

Clausen, A. (1968). Response validity: vote report. *Public Opinion Quarterly*, 32, 588-606.

Clawson, J. (1971). How useful are 90-day purchase probabilities? *Journal of Marketing*, 35, 43-47.

Collins, M. (1982). Comments on Kalton & Schuman's "Effect of the question on survey responses: A review". *Journal of the Royal Statistical Association*, 145, Part 1, p.61.

Collins, M. (1988). Lessons from the polls. Paper presented at the Market Research Society conference, Brighton, March 16-18.

Converse, J. (1976). Predicting no opinion in the polls. *Public Opinion Quarterly*, 40, 515-30.

Converse, J. & Presser, S. (1986). *Survey questions: Handcrafting the standardised questionnaire*. Newbury Park, Calif.: Sage.

Converse, P. (1966). The concept of a normal vote. In *Elections and the political order*, ed. Campbell, A. New York: Wiley.

Converse, P. (1974). Comment on Burnham's "Theory and voting research". *American Political Science Review*, 68, 1024-1027.

Converse, P. & Traugott, M. (1986). Assessing the accuracy of polls and surveys. *Science*, 234, 1094-1098.

Coombs, C. & Coombs, L. (1976). 'Don't know': Item ambiguity or respondent uncertainty? *Public Opinion Quarterly*, 40, 497-514.

Crespi, I. (1988). *Pre-election polling: Sources of accuracy and error*. New York: Russell Sage Foundation.

Crespi, I. & Morris, D. (1984). Question order effect and the measurement of candidate preference in the 1982 Connecticut elections. *Public Opinion Quarterly*, 48, 578-591.

Crewe, I. (1992). A nation of liars? The failure of the polls in the British election of 1992. Paper presented at the annual meeting of the American Political Science Association, Chicago, September 3-6.

Curtice, J. & Jowell, R. (1992). What has been learned from British electoral surveys since 1950. Market Research Society Conference papers, pp. 83-95.

Daves, R. & Warden, S. (1993). Wimps no more: Guidance for allocating undecideds in pre-election polls. Paper presented at the American Association of Public Opinion Research Conference, St. Charles, Illinois, May 20-23.

Day, D., Gan, B., Gendall, P. & Esslemont, D. (1991). Predicting purchase behaviour. *Marketing Bulletin*, 2, 18-30.

Day, R. & Becker, K. (1984). Preelection polling in the 1982 Illinois gubernatorial contest. *Public Opinion Quarterly*, 48, 606-614.

Downs, R. (1957). *An economic theory of democracy*. Harper & Row: New York.

Driver, J. & Foxall, G. (1986). How scientific is advertising research? *International Journal of Advertising*, 5, 147-160.

Duncan, O. & Schuman, H. (1980). Effects of question wording and context: An experiment with religious indicators. *Journal of the American Statistical Association*, 75, 269-275.

Duncan, O. & Stenbeck, M. (1988). No opinion or not sure? *Public Opinion Quarterly*, 52- 513-525.

Ehrenberg, A. & Bound, J. (1993). Predictability and prediction. *Journal of the Royal Statistical Society*, 156 (Part 2), 167-206.

England, L. (1987). The Market Research Society initiative. *Journal of the Market Research Society*, 29, 391-404.

Faulkenberry, G. & Mason, R. (1978). Characteristics of non-opinion and no opinion response groups. *Public Opinion Quarterly*, 42, 533-543.

Feick, L. (1989). Latent class analysis of survey questions that include don't know responses. *Public Opinion Quarterly*, 53, 525-547.

Felson, M. & Sudman, S. (1975). The accuracy of presidential preference primary polls. *Public Opinion Quarterly*, 39, 232-236.

Fenwick, I., Wiseman, F., Becker, J. & Heiman, J. (1982). Classifying undecided voters in pre-election polls. *Public Opinion Quarterly*, 33, 383-391.

Ferber, R. (1956). The effect of respondent ignorance on survey results. *Journal of the American Statistical Association*, 51, 576-586.

Ferber, R. & Piskie, R. (1965). Subjective probabilities and buying intentions. *The Review of Economics and Statistics*, 47, 322-325.

Fife-Schaw, C. & Breakwell, G. (1990). Predicting the intention not to vote in late teenage: A UK study of 17- and 18-year olds. *Political Psychology*, 11, 739-755.

Fishbein, M. & Ajzen, I. (1975) *Belief, Attitude, Intention and Behaviour*. Reading, M.A.: Addison-Wesley.

Fishbein, M. & Coombs, F. (1974). Basis for decision: An analysis of voting behaviour. *Journal of Applied Social Psychology*, 4, 95-124.

Fowler, F. (1992). How unclear terms affect survey data. *Public Opinion Quarterly*, 56, 218-231.

Foxall, G. (1982). *Consumer Choice*. MacMillan: Hong Kong.

Francis, J. & Busch, L. (1975). What we now know about "I don't know". *Public Opinion Quarterly*, 39, 207-218.

Fredericks, A. & Dossett, D. (1983). Attitude-behaviour relations: a comparison of the Fishbein-Ajzen and Bentler-Speckart models. *Journal of Personality and Social Psychology*, 45, 501-512.

Gabor, A. & Granger, C. (1972). Ownership and acquisition of consumer durables: Report on the Nottingham Consumer Durables Project. *European Journal of Marketing*, 6, 234-248.

Gan, B., Esslemont, D. & Gendall, P. (1985). *A test of the accuracy of the Juster Scale as a predictor of purchase behaviour*. Market Research Centre Report No. 45.

Gendall, P. (1994). If the answer was "People who have AIDS get much less sympathy than they deserve", what was the question? *Marketing Bulletin*, 5, 1-12.

Gendall, P., Assendelft, E. & Hoek, J. (1991). The stability of responses to forced choice questions. *Marketing Bulletin*, 2, 41-46.

Gendall, P., Carmichael, V. & Hoek, J. (under review). A test of the conversational logic analysis model of question order effects.

Gendall, P. & Davis, P. (1993). Are callbacks a waste of time? *Marketing Bulletin*, 4, 53-57.

Gendall, P., Esslemont, D. & Day, D. (1991). A comparison of two version of the Juster Scale using self-completion questionnaires. *Journal of the Market Research Society*, 33 (3), 257-263.

Gendall, P. & Hoek, J. (1990). A question of wording. *Marketing Bulletin*, 1, 25-36.

Gendall, P., Hoek, J. & Blakeley, M. (1992). Estimating a socially undesirable behaviour. *Marketing Bulletin, 3, 1-8.*

Gilljam, M. & Granberg, D. (1993). Should we take "Don't Know" for an answer?, *Public Opinion Quarterly*, 57, 348-357.

Goldfield, E., Turner, A., Cowan, C. & Scott, J. (1978). Privacy and confidentiality as factors in survey response. In *American Statistical Association Proceedings of the Social Sciences Section*, 1977, Part 1, Washington DC: American Statistical Association.

Granberg, D. & Holmberg, S. (1986). Prior behaviour, recalled behaviour, and the prediction of subsequent voting behaviour in Sweden and the US. *Human Relations*, 39, 135-148.

Granberg, D. & Holmberg, S. (1991). Self-reported turnout and voter validation. *American Journal of Political Science*, 35, 448-459.

Green, D. & Schikler, E. (1993). Multiple-measure assessment of party identification. *Public Opinion Quarterly*, 57, 503-535.

Greenwald, A., Carnot, C., Beach, R. & Young, B. (1987). Increasing voting behaviour by asking people if they expect to vote. *Journal of Applied Psychology*, 72, 315-318.

Gross, E. (1964). The effect of question sequence on measures of buying interest. *Journal of Advertising Research*, 4, 40-41.

Gross, S. & Niman, C. (1975). Attitude-behaviour consistency: A review. *Public Opinion Quarterly*, 39, 308-368.

Groves, R. Cialdini, R. & Couper, M. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 475-495.

Gruber, A. (1970). Purchase intent and purchase probability. *Journal of Advertising Research*, 10 (1), 23-27.

Haley, R. & Case, P. (1971). Testing thirteen attitude scales for agreement and brand discrimination. *Journal of Marketing*, 43, 20-32.

Hamilton-Gibbs, D., Esslemont, D. & McGuiness, D. (1992). Predicting demand for frequently purchased items. *Marketing Bulletin*, 3, 18-23.

Harris, P. (1978). The effect of clustering on costs and sampling errors of random samples. *Journal of the Market Research Society*, 19, 112-122.

Heald, B. (1970). The relationship of intentions-to-buy consumer durables with levels of purchase. *British Journal of Marketing*, 4, 87-97.

Herzog, A. & Bachman, J. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, 45, 549-559.

Hess, T. (1994). Random individual samples within households. Paper presented to 47th ESOMAR Marketing Research Congress, Davos, Switzerland, 11-14 September, 1994.

Himmelweit, H. Jaegar, M. & Stockdale, T. (1978). Memory for past vote: implications of a bias in recall. *British Journal of Political Science*, 8, 365-384.

Hippler, H. & Schwarz, N. (1989). "No opinion" filters: A cognitive perspective. *International Journal of Public Opinion Research*, 1, 77-87.

Hoek, J. & Gendall, P. (1993). A new method of predicting voting behaviour. *Journal of the Market Research Society*, 35, 361-371.

Hoek, J. & Gendall, P. (1995). Classifying undecided respondents: A comparison of three techniques. Paper presented to the American Association of Public Opinion Research Conference, May 18-21, Fort Lauderdale, Florida.

Hoek, J. & Gendall, P. (1996). Public opinion polling: Supporter or subverter of democracy? In *Dangerous Democracy* ed. J. McGregor, Dunmore: Palmerston North.

Hogarth, R. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, 70 (350), 271-289.

Hoinville, G. & Jowell, R. (1978). *Survey Research Practice*. London: Heinemann.

Holm, J. & Robinson, J. (1978). Ideological identification and the American voter. *Public Opinion Quarterly*, 33, 235-246.

Hosie, J. (1995). Practical implications of non-response bias in sample surveys. Masterate thesis, Department of Marketing, Massey University.

Hunt, S., Sparkman, R. & Wilcox, J. (1982). The pretest in survey research: Issues and preliminary findings. *Journal of Marketing Research*, 19, 269-273.

Husbands, C. (1987). The telephone study of voting intentions in the June 1987 General Election. *Journal of the Market Research Society*, 29 (4), 405-411.

Isherwood, B. & Pickering, J. (1975). Factors influencing individual purchase of motor cars in Great Britain. *Oxford Bulletin of Economics and Statistics*, 37, 227-249.

Jowell, R., Hedges, B., Lynn, P., Farrant, G. & Heath, A. (1993). The polls - A review. *Public Opinion Quarterly*, 57, 238-263.

Juster, T. (1960). Prediction and consumer buying intentions. *American Economic Review*, 50 (2), 604-622.

Juster, T. (1966). *Consumer buying intentions and purchase probability: An experiment in survey design*. Occasional Paper 99, National Bureau of Economic Research: New York.

Juster, T. & Wachtel, P. (1971). Anticipatory and objective models of durable goods demand. *The American Economic Review*, 61, 564-579.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.

Kalton, G., Collins, M. & Brook, L. (1978). Experiments in wording opinion questions. *Journal of the Royal Statistical Society Series C* 27: 149-161.

Kalton, G., & Schuman, H. (1982). The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society A* 145, Part 1, 42-73.

Kassarjian, H. & Nakanishi, P. (1967). A study of selected opinion measurement techniques. *Journal of Marketing Research*, 4, 148-153.

Katosh, J. & Traugott, M. (1981). The consequences of validated and self-reported voting measures. *Public Opinion Quarterly*, 45, 519-535.

Klein, L. & Lansing, J. (1955). Decisions to purchase consumer durable goods. *Journal of Marketing*, 20 (2), 109-132.

Koerner, R. (1980). The design factor - an under-estimated concept? *European Research*, November, 266-272.

Kohut, A. (1986). Is there a crisis of confidence? *Public Opinion Quarterly*, 33, 1-10.

Kraut, A. & McConahay, R. (1973). How being interviewed affects voting: An experiment. *Public Opinion Quarterly*, 16, 381-398.

Kraut, A. Wolfson, A. & Rothenberg, A. (1975). Some effects of position on opinion survey items. *Journal of Applied Psychology*, 60, 774-776.

Labaw, P. (1980). *Advanced questionnaire design.* Abt. Books: Cambridge MA.

Lampert, S. (1978). A new approach to pre-election polling. *Public Opinion Quarterly*, 42, 259-264.

Lampert, S. & Tziner, A. (1985). A predictive study of voting behaviour using Lampert's pollimeter. *Social Behaviour and Personality*, 13, 1-9.

Laurent, A. (1972). Effects of question length on reporting behaviour in the survey interview. *Journal of the American Statistical Association*, 67, 298-305.

Lau, R. (1994). An analysis of the accuracy of "trial heat" polls during the 1992 presidential election. *Public Opinion Quarterly*, 58, 2-20.

Levy, M. (1983). The methodology and performance of election day polls. *Public Opinion Quarterly*, 47, 54-67.

Lewis-Beck, M. & Rice, T. (1982). Presidential popularity and presidential vote. *Public Opinion Quarterly*, 46, 534-537.

Lockerbie, B. & Borrelli, S. (1990). Question wording and public support for Contra aid, 1983-1986. *Public Opinion Quarterly*, 54, 195-208.

Manski, C. (1990). The use of intentions data to predict behaviour: A best-case analysis. *The Journal of the American Statistical Association*, 85 (412), 934-940.

Marquette, J. & Hinckley, K. (1988). Voter turnout and candidate choice: A merged theory. *Political Behaviour*, 10, 52-76.

Marsh, C. (1985). Predictions of voting behaviour from a pre-election survey. *Political Studies*, 33, 642-648.

Martin, L. & Harlow, T. (1992). "Basking & Brooding: The motivating effects of filter questions in surveys. In *Context effects in social and psychological research*, ed. Schwarz, N. & Sudman, S., Springer-Verlag, New York, pp. 81-95.

Mavros, D. (1987). The dawning of election polls in Greece. *European Research*, August, 166-170.

McFarland, S. (1981). Effects of question order on survey responses. *Public Opinion Quarterly*, 45, 208-215.

Miller, M. (1952). The Waukegan study of voter turnout prediction. *Public Opinion Quarterly*, 16, 381-398.

Miniard, P., Obermiller, C. & Page, T. (1983). A further assessment of measurement influences on the intention-behaviour relationship. *Journal of Marketing Research*, 20, 206-212.

Mitau, G., Thorson, S. & Johnson, Q. (1969). Aggregate data in election prediction. *Public Opinion Quarterly*, 33, 96-99.

Mitofsky, W. (1981). The 1980 pre-election polls: A review of disparate methods and results. Paper presented at the annual meeting of the American Statistical Association, Detroit, 10-13 August.

Morrison, D. (1979). Purchase intentions and purchase behaviours. *Journal of Marketing*, 43, 65-74.

Morwitz, V., Jonson, E. & Schmittlein, D. (1993). Does measuring intent change behaviour? *Journal of Consumer Research*, 20, 46-61.

Mullet, G. & Karson, M. (1985). Analysis of purchase intent scales weighted by probability of actual purchase. *Journal of Marketing Research*, 22, 93-96.

Murphy, B. (1975). Political opinion polling in New Zealand. In *New Zealand politics: A reader*, ed. S. Levine, Cheshire Publishing: Sydney, pp. 163-167.

Murphy, B. (1978). Polling and the election. In *New Zealand at the polls*, ed. H. Penniman, American Enterprise Institute for Public policy Research: Washington, pp. 168-180.

Neter, J. & Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association*, 59, 18-55.

Newman, B. & Sheth, J. (1985). A model of primary voter behaviour, *Journal of Consumer Research*, 12, 178-187.

Noelle-Neumann, E. (1970). Wanted: Rules for wording structured questionnaires. *Public Opinion Quarterly*, 34, 191-201.

Norrander, B. (1991). Explaining individual participation in presidential primaries. *The Western Political Quarterly*, 44, 640-655.

Opatow, L. (1991). Some thoughts about how interview attempts affect survey results. *Journal of Advertising Research - Research Currents*, 31, February- March, RC6-RC9.

Parry, H. & Crossley, H. (1950). Validity of responses to survey questions. *Public Opinion Quarterly*, 14, 61-80.

Payne, S. (1946). Some opinion research principles developed through studies of social medicine. *Public Opinion Quarterly*, 10, 93-98.

Payne, S. (1950). Thoughts about meaningless questions. *Public Opinion Quarterly*, 14, 687-696.

Payne, S. (1951). *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.

Pearl, D. & Fairley, D. (1985). Testing the potential for nonresponse bias in sample surveys. *Public Opinion Quarterly*, 49, 553-560.

Pearlin, L. (1961). The appeals of anonymity in questionnaire response. *Public Opinion Quarterly*, 25, 640-647.

Perreault, W. (1975). Controlling order-effect bias. *Public Opinion Quarterly*, 35, 544-551.

Perry, P. (1960). Election survey procedures of the Gallup Poll. *Public Opinion Quarterly*, 24, 531-542.

Perry, P. (1973). A comparison of the voting preferences of likely voters and likely nonvoters. *Public Opinion Quarterly*, 37, 99-109.

Perry, P. (1979). Certain problems in election survey methodology. *Public Opinion Quarterly*, 43, 312-325.

Petrocik, J. (1991). An algorithm for estimating turnout as a guide to predicting elections. *Public Opinion Quarterly*, 55, 643-647.

Pickering, J. (1975). Verbal explanations of consumer durable purchase decisions. *Journal of the Market Research Society*, 17, 107-113.

Pickering, J. & Isherwood, B. (1974). Purchase probabilities and consumer durable buying behaviour. *Journal of the Market Research Society*, 16 (3), 203-226.

Pickering, J. & Isherwood, B. (1975). Determinants of expenditure on consumer durables. *Journal of the Royal Statistical Society, A* 138 (Part 4), 504-530.

Presser, S. (1990). Can changes in context reduce vote over-reporting in surveys? *Public Opinion Quarterly*, 54, 586-593.

Ptacek, C. & Ross, I. (1979). Propensity-to-buy ratings. *Journal of Advertising Research*, 19, 43-47.

Ravinder, H. (1992). Bias in aggregations of subjective probability and utility. *Journal of the Operations Research Society*, 43 (6), 621-627.

Ray, J. (1984). Combining demographic and attitude variables to predict vote. *The Journal of Social Psychology*, 122, 145-146.

Reynolds, N. Diamantopoulos, A. & Schegelmilch, B. (1993). Pretesting in questionnaire design: A review of the literature and suggestions for further research. *Journal of the Market Research Society*, 35, 171-182.

Riedel, J. & Dunne, J. (1969). When the voter decides. *Pubic Opinion Quarterly*, 33, 619-621.

Roper, B. (1983). Some things that concern me. *Public Opinion Quarterly*, 47, 303-309.

Roper, B. (1984). What bothers me about opinion polls. *Across the Board*, 21, 2-6.

Roshwalb, I. (1975). A consideration of probability estimates provided by respondents. *Journal of Marketing Research*, 12, 100-103.

Rugg, D. & Cantril, H. (1942). The wording of questions in public opinion polls. *Journal of Abnormal Social Psychology*, 37, 469-495.

Salmon, C. & Nichols, J. (1983). The next-birthday method of respondent selection. *Public Opinion Quarterly*, 47, 270-276.

Sanchez, M., Morchio, G. (1992). Probing "Don't Know" answers: effects on survey estimates and variable relationships. *Public Opinion Quarterly*, 56, 454-474.

Schuman, H. (1986). Ordinary questions, survey questions, and policy questions. *Public Opinion Quarterly*, 50, 432-444.

Schuman, H. (1992). Context effects: State of the past/state of the art. In *Context effects in social and psychological Research*, ed. Schwarz, N. & Sudman, S., Springer-Verlag, New York.

Schuman, H., Kalton, G. & Ludwig, J. (1983). Context and contiguity in survey questionnaires. *Public Opinion Quarterly*, 47, 112-115.

Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys: experiments on question form wording and context.* New York: Academic Press.

Schuman, H. & Presser, S. (1977). Question wording: an independent variable in survey analysis. *Sociological Methods and Research*, 6 (2), 151-170.

Schuman, H. & Presser, S. & Ludwig, J. (1981). Context effects on survey responses to questions about abortion. *Public Opinion Quarterly*, 45, 216-223.

Schwarz, N. (1990). What respondents learn from scales: the informative function of response alternatives. *International Journal of Public Opinion Research*, 2, 276-285.

Schwarz, N. & Strack, F. & Mai, H. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, 55, 3-23.

Schwarz, N. & Sudman, S. (1992). *Context effects in social and psychological research.* Springer-Verlag: New York.

Seymour, P., Brennan, M. & Esslemont, D. (1994). Predicting purchase quantities: Further investigation of the Juster Scale. *Marketing Bulletin*, 5, 21-36.

Shaffer, W. (1974). Simple and inexpensive election prediction: A practical alternative. *The Western Political Quarterly*, 27, 506-515.

Shaiko, R., Dwyre, D., O'Gorman, M., Stonecash, J. & Vike, J. (1991). Pre-election polling and the non-response bias issue. *International Journal of Public Opinion Research*, 3, 86-99.

Shamir, J. (1986). Pre-election polls in Israel: Structural constraints on accuracy. *Public Opinion Quarterly*, 50, 62-75.

Sharp, L. & Frankel, J. (1983). Respondent burden: A test of some common assumptions. *Public Opinion Quarterly*, 47, 36-53.

Sheppard, B., Hartwick, J. & Warshaw, D. (1988). The Theory of Reasoned Action: a meta-analysis of part research with recommendations for modifications and future research. *Journal of Consumer Research*, 15, 325-343.

Sherman, (1983). Social explanation: The role of timing, set and recall on subjective likelihood estimates. *Journal of Personality and Social Psychology*, 44, 1127-1143.

Shulman, A. (1973). A comparison of two scales on extremity response bias. *Public Opinion Quarterly*, 37, 407-412.

Sieger, K. (1990). Opinion research in East Germany: A challenge to professional standards. *International Journal of Public Opinion Research*, 2 (4), 323-343.

Sigelman, L. (1981). Question order effects on presidential popularity. *Public Opinion Quarterly*, 45, 199-207.

Sigelman, L. & Presser, S. (1988). Measuring public support of the new Christian Right: The perils of point estimation. *Public Opinion Quarterly*, 52, 325-337.

Silver, B., Anderson, P. & Abramson, B. (1986). The presence of others and over-reporting of voting in American national elections. *Public Opinion Quarterly*, 50, 228-239.

Singer, E. (1978a). The effect of informed consent procedures on respondents' reactions to surveys. *Journal of Consumer Research*, 5, 49-57.

Singer, E. (1978b). Informed consent: Consequences of response rate and response quality in social surveys. *American Sociological Review*, 43, 144-162.

Singer, E. Hippler, H. & Schwarz, N. (1992). Confidentiality assurances in surveys: reassurance or threat? *International Journal of public Opinion Research*, 4, 256-268.

Singer, E., von Thurn, D. & Miller, E. (1995). Confidentiality assurances and response. *Public Opinion Quarterly*, 59, 66-77.

Skalaban, A. (1988). Do the polls affect elections? Some 1980 evidence. *Political Behaviour*, 109, 136-150.

Smith, T. (1987). The art of asking questions. *Public Opinion Quarterly*, 51, S95-S108.

Smith, T. (1992). Thoughts on the nature of context effects. In *Context effects in social and psychological research*, ed. Schwarz, N. & Sudman, S., Springer-Verlag, New York.

Sopariwala, D. & Roy, P. (1990). Opinion polling in India. *Journal of the Market Research Society*, 32, 173-186.

Sparrow, N. (1993). Improving polling techniques following the 1992 General Election. *Journal of the Market Research Society*, 35, 79-89.

Strack, F. (1992). Order effects in survey research: Activation and information functions of preceding questions. In *Context effects in social and psychological research*, ed. Schwarz, N. & Sudman, S., Springer-Verlag, New York.

Straits, B. (1990). The social context of voter turnout. *Public Opinion Quarterly*, 54, 65-73.

Stubbs, R. & Hulton, P. (1976). Yea-saying: Myth or reality in attitude research? ESOMAR Congress on fieldwork, sampling and questionnaire design. Venice, 1976, 447-473.

Stapel, J. (1968). Predictive attitudes. In *Attitude Research on the Rocks*, eds. L. Adler and I. Crespi. American Marketing Association: Chicago.

Sudman, S. & Bradburn, N. (1974). *Response effects in surveys: A literature review and synthesis.* Aldine: Chicago.

Sudman, S. & Bradburn, N. (1982). *Asking questions: A practical guide to questionnaire design.* San Francisco: Jossey-Bass.

Tauber, E. (1981). Utilisation of concept testing for new-product forecasting: traditional versus multi-attribute approaches. In *New Product Forecasting*, eds. Y. Wind, V. Manjan & R. Cardozo Lexington, MA: Lexington Books.

Terris, F. (1949). Are poll questions too difficult? *Public Opinion Quarterly*, 13, 314-319.

Theil, H. & Kosobud, R. (1968). How informative are consumer buying intentions surveys? *Review of Economics and Statistics*, 50, 50-59.

Tittle, C. & Hill, R. (1967). The accuracy of self-reported data and prediction of political activity. *Public Opinion Quarterly*, 31, 103-106.

Tobin, (1959). On the predictive validity of consumer attitudes and intentions. *Review of Economics and Statistics*, 41 (1), 1-11.

Traugott, M. (1987). The importance of persistence in respondent selection for preelection surveys. *Public Opinion Quarterly*, 51, 48-57.

Traugott, M. & Tucker, C. (1984). Strategies for predicting whether a citizen will vote and estimation of electoral outcomes. *Public Opinion Quarterly*, 48, 330-343.

Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

van Westendorp, P. (1987). Political opinion polling in the Netherlands: Defeats and victories. *European Research*, August, 158-160.

Voss, S., Gelman, A. & King, G. (1995). The polls: A review. *Public Opinion Quarterly*, 59, 98-132.

Walter, D. (1987). Telephone polls and the General Election. *Journal of the Market Research Society*, 29 (4), 413-418.

Warshaw, D. (1980). Predicting purchase and other behaviours from general and contextually specific intentions. *Journal of Marketing Research*, 17, 26-33.

Webb, N. (1988). Evolution and change in political opinion polling. *European Research*, February, 42-45.

Weimann, G. (1990). The obsession to forecast: Pre-election polls in the Israeli press. *Public Opinion Quarterly*, 54, 396-408.

Weir, B. (1975). The distortion of voter recall. *American Journal of Political Science*, 19, 53-62.

Wildman, R. (1977). Effects of anonymity and social setting on survey responses. *Public Opinion Quarterly*, 41, 74-79.

Williams, D., Weber, S., Haaland, G., Mueller, R. &Craig, R. (1976). Voter decision-making in a primary election: An evaluation of three models of choice. *American Journal of Political Science*, 20, 37-49.

Worcester, R. (1984). The polls: Britain at the polls 1945-1983. *Public Opinion Quarterly*, 48, 824-833.

Worcester, R. (1986). The history of public opinion polls. Joint ESOMAR/WAPOR session on social and opinion research. Strausbourg, France, pp. 419-430.

Worcester, R. (1987). Political opinion polling in great Britain: Past, present and future. *European Research*, August, 143-151.

Worcester, R. (1992a). Opinion polls in British general elections. Paper presented to Market Research Society conference, 1992, pp. 51-61.

Worcester, R. (1992b). The performance of the political opinion polls in the 1992 British General Election. *Marketing and Research Today*, 20, 256-263.

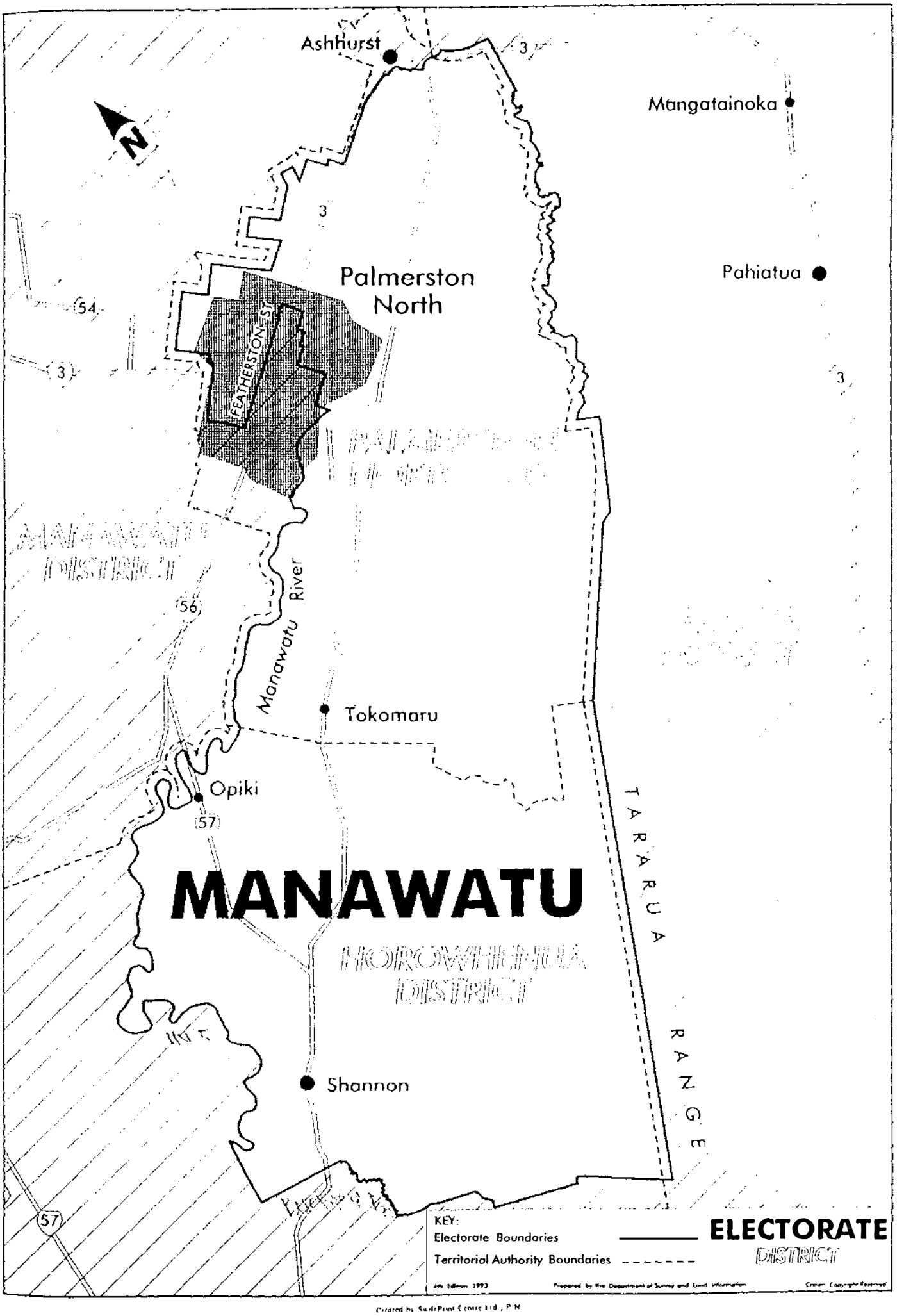
Worcester, R. & Burns, T. (1975). A statistical examination of the relative precision of verbal scales. *Journal of the Market Research Society*, 17 (3), 181-197.

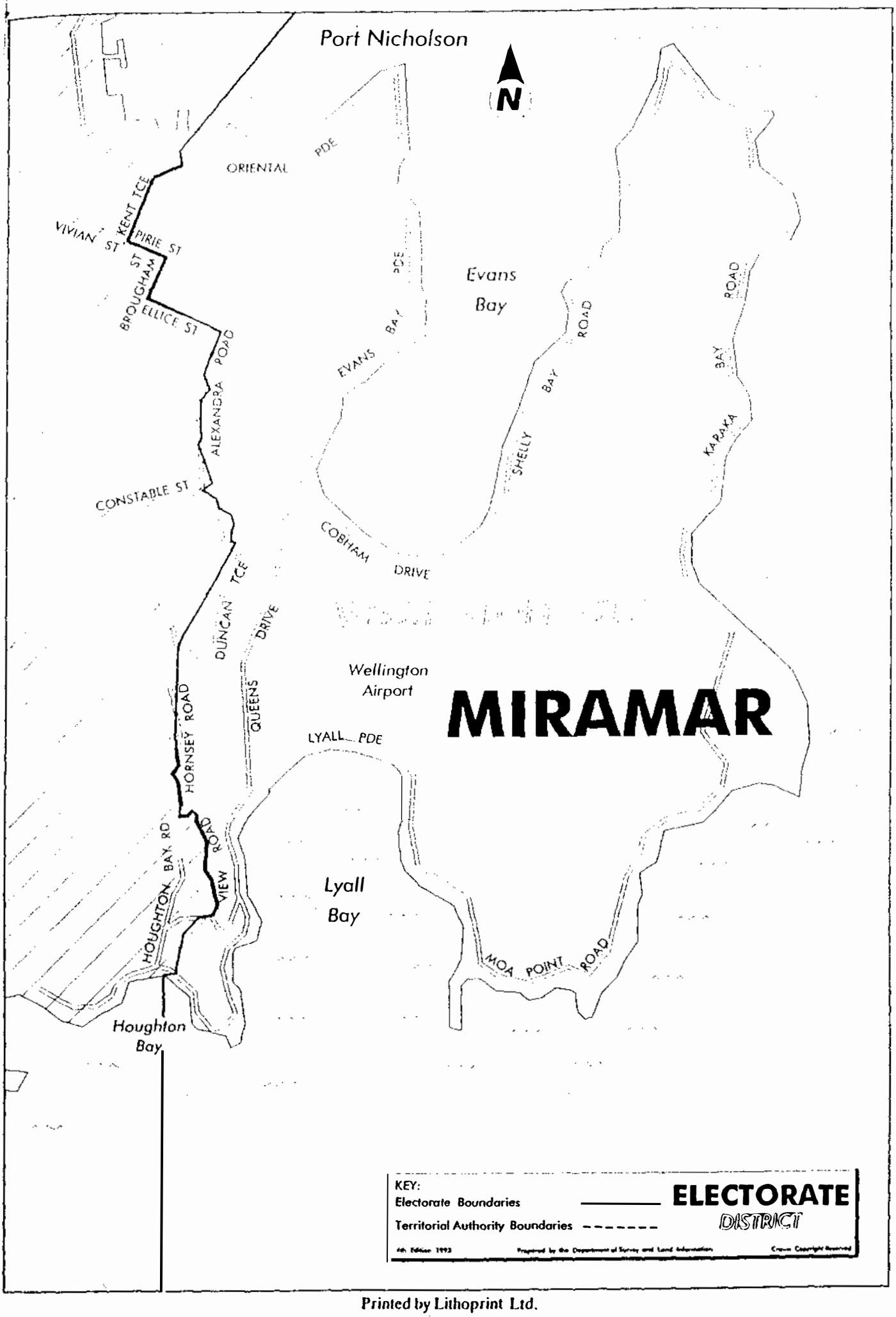
Yalch, R. (1976). Pre-election interview effects on voter turnout. *Public Opinion Quarterly*, 40, 331-336.

APPENDICES

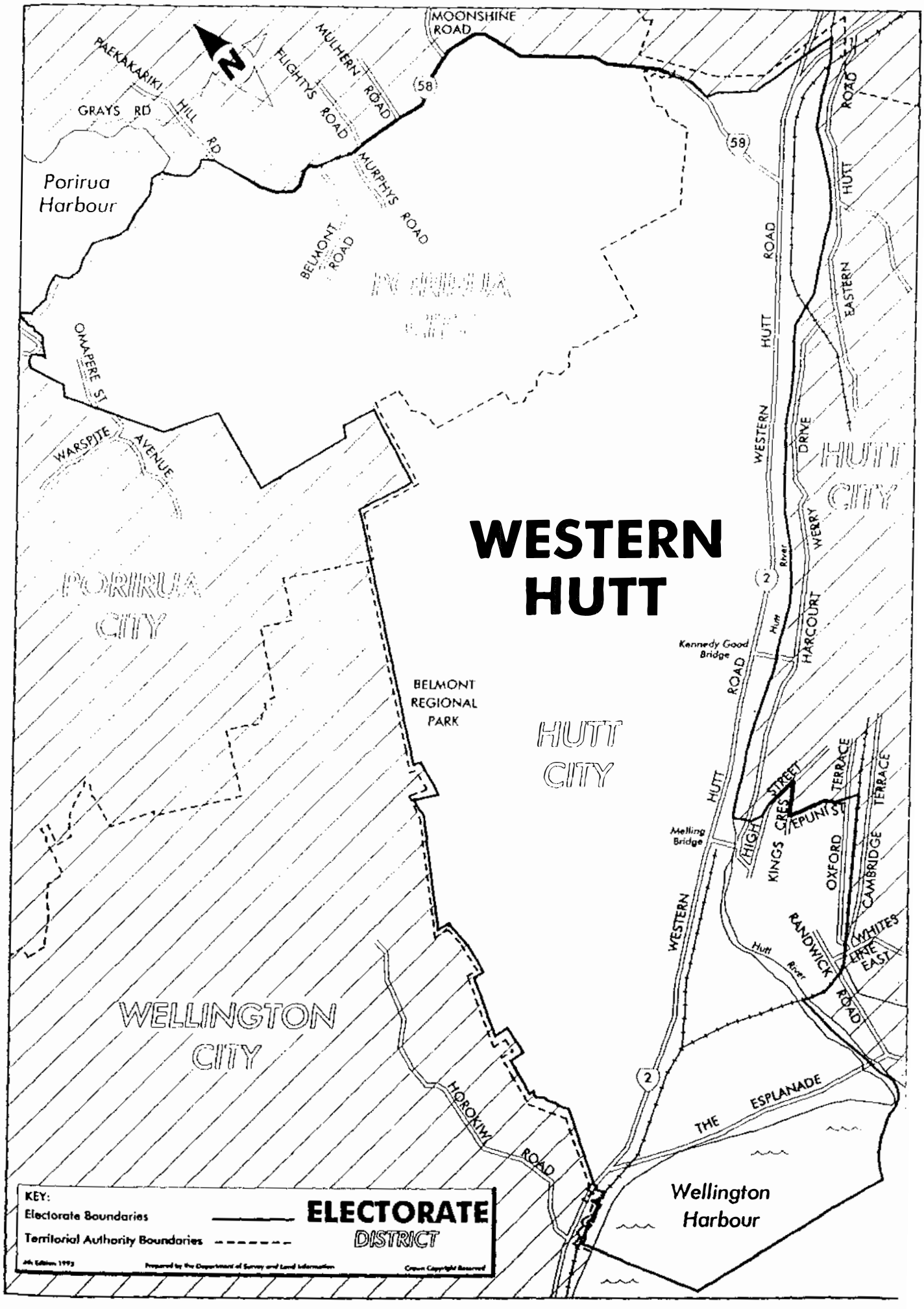
APPENDIX A:

DETAILS OF THE ELECTORATES


Ashhurst
Mangatainoka
Pahiatua
Palmerston North
FEATHERSTON ST
Manawatu River
Tokomaru
Opiki
MANAWATU
TARARUA RANGE
Shannon
KEY:
Electorate Boundaries
Territorial Authority Boundaries
ELECTORATE
DISTRICT
Prepared by the Department of Survey and Land Information
Crown Copyright Reserved



Printed by Lithoprint Ltd.


WESTERN HUTT
MOONSHINE ROAD
PAEKAKARIKI HILL RD
GRAYS RD
MULHERN ROAD
FLIGHTYS ROAD
MURPHYS ROAD
BELMONT ROAD
58
Porirua Harbour
PORIRUA CITY
OMAPERE ST
WARSPITE AVENUE
HUTT ROAD
WESTERN HUTT ROAD
EASTERN HUTT ROAD
HUTT CITY
WERRY DRIVE
Hutt River
2
HARCOURT
Kennedy Good Bridge
BELMONT REGIONAL PARK
HUTT CITY
HIGH STREET
KINGS CRES
EPUNI ST
TERRACE
OXFORD TERRACE
CAMBRIDGE TERRACE
Melling Bridge
WESTERN HUTT ROAD
RANDWICK ROAD
WHITES LINE EAST
WELLINGTON CITY
HOROKIWI ROAD
THE ESPLANADE
Wellington Harbour
KEY:
Electorate Boundaries
Territorial Authority Boundaries
ELECTORATE
DISTRICT
Prepared by the Department of Survey and Land Information
Crown Copyright Reserved

APPENDIX B:

INTERVIEWERS' INSTRUCTIONS

Massey University
DEPARTMENT OF MARKETING

Voting Research
INTERVIEWING NOTES*

Background

This interviewing is being conducted as part of my PhD research, which aims to test and compare different methods of measuring people's voting behaviour. For this reason, there are several versions of the questionnaire, although these share many common characteristics, they do differ in important ways.

The purpose of this training is to ensure you are confident about interviewing, and that you are familiar with, and understand, the questionnaires.

Interviewing

Although interviewing may seem as though it's nothing more than asking people some simple questions, many people see interviewers as strangers asking them answers to questions in which they may have no special interest, for a purpose they find obscure and on behalf of an organisation they may not know existed. Looking at survey research from this point of view makes it clear that interviewers play a vital role in obtaining and maintaining respondents' co-operation and, ultimately, the quality of the data will also depend largely on the skill and conscientiousness of the interviewers who collect it.

When to Interview

1. Try to plan your interviews for Wednesday and Thursday evening (from 5.00pm to 7.30pm); Friday and Saturday. Some people object to being interviewed on Sunday; if this appears to be the case, apologise for disturbing them and ask if you could conduct the interview on Monday instead.

2. Try to complete your evening interviewing by 7.30pm so that you will not inconvenience respondents by calling at a late hour. However, you should not interview after dark unless you feel comfortable doing so.

3. Make all your call backs after 5.00pm unless you have an appointment at another time which would better suit your respondent.

4. Plan your work ahead. Start your interviews as soon as you can.

---

* These notes are based heavily on Hoinville & Jowell's (1978) advice.

Where to Interview

1. You will have several starting addresses set out in your interviewing kit. Go to the first of these. (They have been arranged so as to minimise the amount of travelling you have to do.)

2. Stand in front of the starting address (with your back to it) then turn right and start your interviews, beginning with the first occupied dwelling next to the starting address.

3. Proceed around the block, making right turns and attempt additional interviews at every adjacent dwelling until you have completed six interviews around that starting address. Within each cluster, you must ensure that three interviews are conducted with a female and three with a male.

4. An interview must be attempted at each occupied dwelling indicated by the selection system. However, if the dwelling selected is unoccupied, is a commercial property or an institution (e.g. a school or a hospital), you should continue around the block until you reach the next occupied dwelling and attempt an interview there.

5. Where two or more dwellings occupy the same section (e.g. townhouse or flats) you should treat each flat or unit as a separate dwelling. (In other words, if the street numbers you are following are getting **bigger**, flat 2 would **follow** flat 1; whereas if the numbers you are following are getting **smaller**, flat 2 would come **before** flat 1.)

6. You must follow a route of all right hand turns and this "right hand rule" should be followed regardless of where you are interviewing.

7. If you work an entire block and you find it does not contain enough dwellings for you to attempt six interviews, proceed to the dwelling on the opposite side of the street which is nearest the dwelling where you attempted your first interview. Treat the address across the street as the next occupied dwelling and continue around that block using the "right hand rule".

8. If there are no dwellings on the street or road opposite the dwelling where you attempted your first interview, walk directly across the street from where you attempted your first interview and circle that block in a clockwise direction (following the "right hand rule") until you come to an occupied dwelling. Attempt your next interview there.

9. If you are still not able to attempt the necessary six interviews after circling the adjacent block described in point 8 above, take the next block found, following a clockwise direction.

10. If the designated respondent is not available on your first attempt, you should make two call backs (return visits) before you abandon that household. Each questionnaire has a section following the introduction where you can record details of when to call back. Try to make a specific time to interview someone who is unavailable when you first call.

11. If you are refused an interview, or if you have made three calls at the same dwelling (your original call plus two call backs) and have not obtained an interview with the appropriate person, you should abandon that household. Likewise, if you need to interview a male and you call at a household where there are no males, (and vice versa) you should abandon that address.

12. If you are, for the above reasons, unable to interview the appropriate person, you must record

the outcome details on the questionnaire (see below for these) and you must substitute that person by conducting another interview at the end of that cluster of interviews. You must record the details of the unsuccessful contact at the top of the questionnaire. That is, write the address and "refused" or "no males" or whatever is appropriate and continue to use that questionnaire until you have a successful interview. You have a pack of replacement questionnaires; if necessary use one of these to conduct the successful interview (use replacements where you feel the details will not be obvious because you have had to contact a number of people before obtaining a respondent). The replacements do not have ID numbers; I will allocate these later.

Recording Addresses and Calls

1. Details of each interview you attempt should be recorded on the questionnaire.

2. At the first occupied dwelling (next to the starting address), before you knock or ring the door bell, record the address, including the street, street number and flat or unit number.

3. For each attempted interview, you should record the outcome in the following way:

   1 = successful interview at first attempt
   2 = successful interview at second attempt
   3 = successful interview at third attempt
   4 = refusal
   5 = unable to make contact with respondent after three attempts
   6 = no respondent of required gender
   7 = other (savage dog prevented access; locked gate; people would not answer the door; no-one spoke English etc.)

Selecting Respondents

1. At each dwelling explain to the person who answers the door, that you would like to interview someone (specify the gender of the person you need to interview) currently living in the household who is 18 years or older. Explain that this includes the person you are talking to (if they are of the required gender) and any borders or lodgers, but it does not include people who are away at school or university, away on long trips, or otherwise not living at home.

2. If there is more than one person of the specified gender available, ask to speak to the man or woman whose **birthday is due next.** This is the person who you should interview in this household.

3. If the selected person is not available, make a time and day when it would be convenient to call back and interview her/him. Record this on the questionnaire. Ask for the selected person's name and record this so you will remember who to ask for when you call back.

4. If the person who answers the door does not know whose birthday is due next (this is possible in a flat where occupants are not member of the same family), interview the person who answered the door (if they are of the gender you require), otherwise interview a person of the required gender who is at home you call. However, **do not** resort to this until you have made a genuine effort to establish the selected respondent.

5. If more than one person in a dwelling qualifies to be interviewed as a result of the "next birthday" test (that is, two people of the required gender have their birthdays on the same day) choose the one to be interviewed by tossing a coin.

6. Only **one** person in a household should be interviewed and only the **selected** person should be interviewed; no substitutions should be made. You should make up to three calls to a selected dwelling in an effort to locate and interview a respondent.

Your Approach to the Interview

1. First impressions are crucial. Introducing yourself and persuading the respondent to give the interview is probably one of the most critical and difficult parts of the interview. Your job is more than just making calls and conducting interviews; it is selling people on being interviewed when there might be some resistance.

2. Most resistance is due to two reasons:

   Misunderstanding: People may think that this is not really a survey or they may suspect that it's a sales pitch.

   Can't be bothered: People are too busy, or they see it as an invasion of their privacy.

   You need to be able to overcome these objections by convincing the respondent that:

   You are calling for a legitimate and reputable organisation.
   You are engaged in important and worthwhile research.
   The respondent's participation is vital to the success of the project.

   At the start, you need to explain who you are, why you are there and who you are working for (who is conducting the survey and why). If respondents are not sure about these details, they may see you as someone trying to sell them something, or, worse, as someone trying to steal something from them. You have in your interview kit an identification badge which clearly links you to the University, please make sure you wear it while you are interviewing, You also have a letter of authenticity, written by the Head of the Marketing Department, which establishes you as a *bona fide* interviewer engaged in research conducted within the Department. Each questionnaire has a standard introduction which covers these points. Please make sure that you are familiar with it. However, please make sure that you always state your introduction in an interesting and confident manner so that it does not sound like a tired and dreary recital.

3. Your voice, words and your appearance must convey your credibility. Your manner must help people feel at ease, and must also convey confidence and competence. In other words, if you seem hesitant about obtaining an interview, chances are that respondents will hesitate in giving you one, On the other hand, the overbearing exuberance of an encyclopedia salesperson may not be the answer either!! Ideally, try to be neutral in your appearance, clothing and speech.

4. Approach people as if they are friendly and interested. Assume that they are willing to give you an interview. However, do ensure that you **do not** suggest to respondents that they are under an obligation to participate in the survey. That is, implying that they somehow should,

or must, participate is unethical and must not happen. On the other hand, please do not deter respondents. In practice, most people are happy to be interviewed, and, as I have mentioned, their decision to participate will depend heavily on how you have approached them. Make sure that your nametag is clearly visible and be prepared to answer questions about Massey, the Marketing Department, or the survey itself (you can give respondents my work number, 356 9099 ext 8525 if they would like to talk to me about the study).

5. As mentioned above, some people may want reassurance about particular points before they agree to participate. Some people think that surveys are an invasion of their privacy and you may need to explain that the aim of the survey is worthwhile, that their answers will be confidential, and that no-one will be able to trace the details they provide back to them. The explanatory letter you have deals with these issues as well.

6. Some respondents may also wonder how they came to be selected for the study, and how we obtained their address. The answer to this question is the electoral roll. I selected starting addresses randomly from this and their address is one of six that will be randomly interviewed around that starting point. The birthday test, which you will have mentioned to the first person you spoke to at the household, attempts to randomly identify a respondent within a household.

7. Some respondents may also query the purpose of the survey. You should be prepared to discuss this with them, but, at the same time, please do not say anything that could prejudice respondents' replies, or that would give them the idea that the topic was not something they would like to discuss. If you appear ignorant of the survey's purpose, or uninterested in its aims, respondents will be less willing to participate.

8. During the interview, some people may ask why a particular question has been included. I will try to cover the purpose of each question as we discuss the questionnaires later in the day, but if you have any queries, please let me know.

9. Some people will state that they are too busy to be interviewed. It is true that you cannot always call at the most convenient time, so ask if you can make a definite time at which they would be able to be interviewed (making a definite time is much better than some vague time, like later this afternoon). If someone asks you how long the interview will take, tell her or him that it will take between 5 to 10 minutes, do not attempt to deceive them.

10. Above all, talk to people, not at them. If they believe you are really interested in them, they are more likely to participate.

11. If you find that you are getting repeated refusals, something is probably wrong with your interviewing approach, and you should contact me immediately to discuss the problem.

The Interview and Your Role in It

1. As an interviewer, it is important that you are aware of the vital role you are playing in this survey. You are the link between me (the researcher) and the respondents. The quality of the final results depends on your ability to elicit the information needed from respondents.

2. You should think of yourself as a **communicator**. The questionnaire is your tool for communicating to respondents what I would like to know. But, unless you use it correctly, your interviews will not be productive.

3. Remember that respondents have not been trained for their role, and that they may need some help. However, at the same time, remember that it is vitally important that you do not bias or pre-empt their responses. Help respondents to realise that they are not taking part in some sort of test, and that there are no right or wrong answers; it's just their own opinions that we are really interested in. Make sure that you maintain some eye contact with respondents (but don't try to stare them out!!). It is often important to look at respondents after you have asked them a question, so that you can pick up signs of confusion or worry.

4. It is very important to maintain respondents' interest and goodwill throughout the interview. Luckily, most people are still quite willing to discuss their political views. You must ensure that the pace of the interview suits respondents. While some people answer quickly and want to get on with the interview as fast as possible, others need more time to think and should be allowed time to consider their answers. We will practise pacing interviews later in the day.

5. You must record respondents' answers in such a way that they make sense to those who have to enter and analyse the results. This means you have to report completely, filling in something for every question, (unless it was a skip question). And you **MUST** write legibly.

nterviewing Style

1. When you ask respondents questions, it is vitally important that you use the **exact words** printed in the questionnaire as even minor and apparently trivial changes can affect respondents' perceptions of what they are being asked, and so may affect their responses. If a respondent appears to have misunderstood a question, first repeat it slowly. **DO NOT TRY TO EXPLAIN IT IN DIFFERENT WORDS**, as these may bias her or his response. If a respondent asks you what a question means, try to find out what s/he believes it means; if the respondent's interpretation is correct, tell her or him; if it is wrong, try to guide her or him to the correct meaning, without deviating from the original wording. If this sounds very difficult, don't worry, we'll have some practice before the end of the day.

2. Sometimes, it is difficult to keep respondents to the point (this can be a particular problem with political surveys). If a respondent starts to answer one question by giving information that is needed in a later question, you will need to stop her or him, and explain that that point occurs later. Even if the respondent appears to be making a full response to the later question, do not record it as intervening questions may affect her or his reply, or may remind her or him of other aspects of the topic. It is **VITAL** that you **DO NOT** alter the sequence of questions (since the versions differ according to this), so when you come to the question a respondent appears to have answered earlier, you can preface it by saying "I'd just like to check". If a respondent wishes to go back and change an answer to a question, please note this in writing beside the question.

3. Although you can add some link phrases (such as "Now, I'll move on to the next question") to move from one question to the next, it is vital that these have no bearing on the answers respondents give and so cannot bias the results.

4. It is important to realise that many things can bias respondents' answers. I have tried to design questionnaires which I hope will be easy and straightforward for you to administer and for respondents to answer, however, interviewers can have a profound effect on **how** respondents answer questions, and you should do your best to minimise any pre-empting that may bias respondents' answers.

Interviewing Technique

1. In normal conversation, people often bias their answers according to what they believe another person expects or wants to hear. While this may be a good idea in social relationships, it is a real danger in survey research. You **MUST** avoid giving respondents any cues as to what you expect them to say, or any feedback on what they have said, until the interview has concluded. Even your answers to simple questions such as what you do or where you live may lead respondents to modify their answers.

2. If respondents are unsure what to say, they may ask you what you think. Again, try to turn the question around so that you encourage the respondents to talk about themselves and their views.

3. You must be sure to ask all questions in a neutral and straightforward way. A note of surprise or an over-sympathetic reaction may easily affect a respondent's subsequent answers. You must make a strong effort to show a uniform, polite and consistent interest in respondents' answers. Even apparently very trivial comments such as "good" can seriously bias subsequent answers.

4. If a respondent seems flippant and appears not to be taking the interview seriously, the best thing is to break off interviewing and remind the person that the survey is serious and to ask again for their co-operation. If it is clear that you are serious, and that you take your job seriously, only a very determined and troublesome respondent will persevere with flippancy. However, if you feel a respondent has not been giving you helpful or truthful answers, please make a note of this on the questionnaire (it might be tactful if you do this discretely!).

5. There may be times when respondents appear to have given you inconsistent answers. This may be because they have misunderstood the question, or you may have misunderstood their answer. If you notice a clear contradiction in a matter of fact, tactfully draw the respondent's attention to it. However, if the inconsistency is to do with a matter of opinion, then leave it, as many people do hold inconsistent views (especially with respect to politics).

6. Sometimes it may be necessary to probe respondents' reactions in order to help them give an answer. Probing can be both verbal and non-verbal. For example, an expectant glance or a "mmmm" can probe just as direct questions such as "tell me more about that" can. All the questions in this study are pre-coded, but even then probing may be necessary. For example, suppose you had just asked someone how interested they were in this year's election. The answer may be "I couldn't really say". This answer doesn't usually mean that the respondent has no views, but rather that s/he is hesitating between the categories offered. In such cases, you could ask "Which comes closest to your views?", and repeat the question. Many of the questions in this survey contain a "don't know" category, but only use this as a last resort.

We will discuss this in more detail later.

7. On the questionnaire the questions which you are to ask are printed in capitals and lower case, like these notes. However, the instructions to you are printed in capitals ***LIKE THIS***. In some questions all you need to do is read the question, you do not need to read the reply options. However, other questions do require you to read the options, and some questions involve the use of a showcard. The instructions beside each question make this clear and we will discuss the questionnaires in detail later in the day.

8. For most questions all you have to do is circle the appropriate number; however, sometimes you may have to write an answer. Please ensure you use a blue ballpoint pen (**NOT BLACK or a fountain pen**). If you make a mistake in circling an answer, put two diagonal lines (//) through the incorrect answer and then circle the correct response.

9. All questions have "Don't know"; "None" etc categories to use where appropriate. However, if you come to a question where respondents simply cannot use one of the options available, please write their response in and I will deal with it when I check the questionnaires.

10. Make sure that you record the answers as the interview proceeds. If you try to remember the answers and complete the questionnaire afterwards, you nearly always introduce error through incorrect or incomplete recollection of the actual answers.

11. If the interview presents any special difficulties (for example, if the respondent has a hearing or sight disability and the usual method of questioning has to be changed, or if there are any language difficulties) please make sure you note this at the end of the questionnaire.

12. Some questions contain skip instructions which look like this ***GO TO Q4***. These are instructions about which question you should go to next. If there is no skip instruction beside the answer given by a respondent, you should continue to the next question.

13. Your set of materials also contains a set of showcards. You are to use these cards whenever there is an instruction on the questionnaire telling you to ***SHOWCARD -***. (Each card is identified by a letter or a title.) At that time in the interview you are to show that particular card to the respondent as an aid in answering the question.

14. Occasionally, respondents ask if they may have access to the results of the survey, or if they may retain a copy of the questionnaire. If this happens, give the respondent my name and Massey phone number (356 9099 ext 8525) and ask them to get in touch directly with me. (Alternatively, tell them that you will ensure that their name is passed onto me so that I can get in direct contact with them.)

15. Before leaving, check quickly through the questionnaire to ensure that all relevant questions have been asked and the questions correctly recorded. Give respondents a last chance to ask questions and thank them for their participation and help so that they are left thinking that the interview has been worthwhile.

16. I have discussed the way in which you are to select households, but you must always ensure that your own safety is paramount. That is, if you come to a household, that appears in any way dangerous, or to a respondent who alarms you for whatever reason, **DO NOT PROCEED WITH THE INTERVIEW**.

17. The interviews you conduct will be audited. If there is **ANY** evidence that the procedures outlined in this document have not been complied with, **PAYMENT WILL BE FORFEITED**.

THE QUESTIONNAIRES

There are two basic versions of the questionnaire, and these are split further to make a total of eight versions. I've set out the characteristics of these below.

Traditional Questionnaire Version A

This begins with a question asking people their interest in this year's election. There is a showcard that you must give them, then, when respondents have indicated which statement best describes their interest, code the number in the column. If someone says that no statement really describes his or her interest, ask her or him which one is closest, or seems nearest to how she or he feels.

Over the page there is a set of questions which ask people their views on some key issues. Questions 2A to 2D all use a table printed beneath question 2D. Each row in this table corresponds to a question and, once respondents have told you which party they think has the best policies relating to a specific issue, circle the number in that row which corresponds with the party they have mentioned.

Question 3 asks respondents about their party identification. Read the list of parties starting at the asterisked party **but do not read the options in the shaded box**, and circle the number corresponding to the party they feel most affinity with. Try to get respondents to make a choice, but if they say they cannot choose, or that they do not feel an affinity with any party, then code either "don't know" or "none" in the shaded box. If you code either "don't know" or "none" then turn the page and move to question 5.

If respondents have identified a party, then move to question 4 and ask them how strongly (Alliance/ Labour/ National/ NZ First/ or other party, depending on who they stated in question 3) they feel. Read the options to respondents and code the one that best applies. There should be no problem with one of the options not applying to respondents.

Question 5 asks whether respondents belong to a political party. If they answer yes, move to question 5B. If they say "no", refuse to tell you, or say that they "don't know", move to question 6A. Question 5B asks those respondents who indicated that they do belong to a political party to specify that party.

Question 6A asks if respondents are enrolled to vote; that is, whether they have registered on the electoral roll. If they answer yes, ask them in which electorate they have registered and record the answer in the space provided.

Questions 6B to 6C all relate to the table printed beneath question 6C. Question 6B asks how respondents voted in the 1990 General Election. If they voted for a party not listed in the table, code other and write the name of the party in the space proved in question 6B. Question 6C asks respondents which party they would have voted for, if this year's General Election had been held yesterday. Code their response to this question in row Q6C. It is important in all these questions that you **DO NOT** prompt respondents. That is, do not read out the names of the parties to them, but let them tell you unaided.

Question 8 uses a showcard with the Juster Scale. This is a probability scale and is used to enable respondents to project the likelihood or chance that they will do something. There is a detailed introduction to this question, make sure that you read it carefully and understand it. Give respondents Showcard B and allow them a little time to read this. Then ask respondents to use the scale to tell you how likely it is that they will eat out at a restaurant during the next month. This is an introductory question designed to give respondents practice in using the scale. Do not code the response to this question. Next ask respondents to use the scale to tell you how likely it is that they will vote in this

year's General Election. They may say probable, in which case you should answer "What number is that?" ; then, once they have told you, repeat the Juster number as a check before coding it in the column provided.

The remainder of this questionnaire collects demographic details. First, ask respondents in which year they were born and write down the last two digits (i.e., if someone said 1958, you would write down 58). Question 10 asks respondents about their education level. Give respondents Showcard C and ask them to tell you which statement best describes their highest level of education. Finally, note their gender (you should only need to ask this in really exceptional circumstances!!)

TRADITIONAL Version 2

This questionnaire is the same as version 1, except the questions are asked in a different order. You now begin with the question asking respondents if they are enrolled and, if so, in which electorate. You then move onto the questions about the last election and this election (see the previous section for details on these). Over the page are the questions relating to respondents' interest in the election and their views on key election issues. At the bottom is the question on party identification and on page 3 the strength of this identification. Then follow the questions on party membership and the Juster Scale question investigating respondents' probability of voting. The demographic questions are the same and appear in the same order.

TRADITIONAL Version 3

This questionnaire is the same as version 1 and the questions appear in the same order as version 1 except instead of asking respondents directly who they plan to vote for in this year's General Election, you will instead ask them to indicate this on a ballot paper which they will place into a Secret Ballot Box. The idea of this is that respondents will associate a higher level of secrecy with their responses and so those who may otherwise declare themselves to be undecided because they do not wish to tell an interviewer their planned voting behaviour may be willing to express this if offered a "secret" way of doing so. It is **VITAL** that you note the number of the ballot paper you give a respondent on the questionnaire that you used to record the rest of her or his answers. Use the space provided by question 6C.

TRADITIONAL Version 4

This version is the same as version 2, but it also uses the secret ballot method instead of the direct question.

JUSTER Version 1

The Juster questionnaires make greater use of the Juster Scale than the traditional versions do. Version A begins with the same interest question used in the traditional questionnaires and uses exactly the same format until question 7. Here, instead of asking respondents directly who they plan to vote for and have them tell us or indicate their choice on a ballot paper, you will instead use the Juster Scale Showcard to ask respondents their probability of voting for each of the parties shown in the table below question 7. Again the probability is written down the left hand side of the page and the parties are written along the top. One of the parties will have an asterisk by it; start by asking respondents about this party. Thus in this example, you would start by asking "How likely is it that you will vote for the New Zealand First candidate in this year's General Election?" Then ask "How likely is it that

you will vote for the Alliance candidate in this year's General election?"; then "How likely is it that you will vote for the Labour candidate in this year's General election?"; then "How likely is it that you will vote for the National candidate in this year's General election?"; finally "And how likely is it that you will vote for a candidate from another party in this year's General Election?". In each case check respondents' answer by repeating the numeric probability. A typical conversation might be as follows:

You: How likely is it that you will vote for the Labour Candidate in this year's General Election?

Respondent: Oh, there's a fair possibility.

You: What number is that?

Respondent: Four.

You: You mean there's a four in ten chance that you'll vote for the Labour candidate in this year's General Election?

Respondent: Yes, that's right.

You: Now, how likely is it that you'll vote for the National candidate in this year's General Election?

Respondent: I'd say there's no chance of this happening.

You: What number is that?

Respondent: Zero.

You: So there's a 1 in 100 chance that you'll vote for the National candidate in this year's General election?

Respondent: Yes.

Once you have worked through the parties, always ending with the question about the **other** candidates, ask respondents to use the same card to indicate the chance that they will cast a vote in this year's General Election. Code the answer to this question in the column on the far right labelled "Vote".

The demographic questions are identical to those used in the traditional questionnaires.

JUSTER Version 2

This uses exactly the same questions as the Juster version 1, but reverses the order of them. Please read through the questionnaire carefully and then check the question details given in version 1 above.

You will have an opportunity to role play each of these questionnaire versions during today's session. It is **really important** that you ask questions if you cannot follow any aspect of the questionnaires, as you must be sure you understand the questionnaires before you can hope to interview somebody.

APPENDIX C:

SURVEY QUESTIONNAIRES

SURVEY QUESTIONNAIRES

The survey questionnaires reproduced here are discussed in detail in Chapter 6. Interviewers were instructed to complete one version of each questionnaire in each cluster. The order in which the questionnaires were presented was rotated in each cluster.

ID_____1-4
OUTCOME_____5
SURVEY_____6
VERSION_____7

Massey University
DEPARTMENT OF MARKETING

TRADITIONAL QUESTIONNAIRE A

RECORD ADDRESS ON BACK PAGE OF QUESTIONNAIRE

Hello, my name is _____, and I'm a researcher from Massey University (in Palmerston North). We're conducting a very short survey as part of an important research project about this year's General Election; it will only take 5 to 10 minutes. Could I speak to the male/female in the household who is 18 years or older and whose birthday comes next please. Are you that person?

IF YES, SAY: Thank you. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move onto the next one. We're very interested in your ideas and opinions and, as I said, these will be held in the strictest confidence.

IF NO, ASK IF YOU CAN SPEAK TO THAT PERSON. IF UNAVAILABLE, ARRANGE A SUITABLE TIME FOR A REINTERVIEW.

Name:_____ Time:_____ Date:_____

Address:_____

IF A NEW PERSON COMES TO THE DOOR, SAY:

I'm a researcher from Massey University (in Palmerston North) and we're conducting a short survey about the General Election; it will only take about 5 minutes of your time. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move onto the next one. We're very interested in your ideas and opinions, and, as I said, these will be held in the strictest confidence.

1. First of all, please read this card and tell me which statement best describes your interest in this year's General Election.

SHOWCARD A AND CIRCLE ONE NUMBER BELOW

I am very interested in this year's election 1
I have some interest in this year's election 2
I don't really feel interested in this year's election 3
I am completely uninterested in this year's election 4

There have been several issues affecting New Zealanders discussed in the news recently, I'd like to ask you now which party you think has the best policies for dealing with some of these issues.

2A. First, which party do you think has the best healthcare policies?

CIRCLE ONE NUMBER IN ROW 2A BELOW DO NOT PROMPT

*IF OTHER, SPECIFY HERE*_____

2B. And which party do you think has the best education policies?

CIRCLE ONE NUMBER IN ROW 2B BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

2C. Now, which party do you think has the best policies for dealing with unemployment?

CIRCLE ONE NUMBER IN ROW 2C BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

2D. And which party do you think has the best policies for dealing with crime and violence?

CIRCLE ONE NUMBER IN ROW 2D DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q2A. Healthcare	1	2	3	4	5	6	7
Q2B. Education	1	2	3	4	5	6	7
Q2C. Unemployment	1	2	3	4	5	6	7
Q2D. Crime & Violence	1	2	3	4	5	6	7

3. Now, I'd like to ask you some questions about your feelings towards the political parties. Generally speaking, do you think of yourself as being..... [START READING AT *] or what?

READ AND CIRCLE ONE NUMBER BELOW. DO NOT READ SHADED BOX

Alliance 1
Labour 2
National 3
New Zealand First 4
Other . . . . . 5 --> Specify _____

None.....6
Don't Know.....7 --> GO TO Q5

4. And how strongly _____(*PARTY CODED IN Q3*) do you feel?

READ AND CIRCLE ONE NUMBER BELOW

Very strongly 1
Reasonably strongly 2
Not very strongly 3
Not at all strongly 4

5A. Are you currently a member of a political party?

CIRCLE ONE NUMBER

Yes	1	***---> GO TO Q5B***
No	2	
Refused	3	***---> GO TO Q6A***
Don't Know	4	

5B. Which party are you a member of?

NOTE PARTY HERE _____

6A. Are you enrolled to vote in this year's General Election?

CIRCLE ONE NUMBER ONLY

Yes	..1---->	Which electorate_____
No	..2	
Don't Know	..3	

6B. Now, thinking back to the 1990 General Election, which party did you vote for?

CIRCLE ONE NUMBER IN ROW 6B BELOW DO NOT PROMPT

IF OTHER, SPECIFY PARTY HERE _____

6C. If this year's General Election had been held yesterday, which party would you have voted for?

CIRCLE ONE NUMBER IN ROW 6C BELOW DO NOT PROMPT

IF OTHER, SPECIFY PARTY HERE _____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q6B. 1990 Voting	1	2	3	4	5	6	7
Q6C. Vote Yesterday	1	2	3	4	5	6	7

8. Now, please look at this card. It describes a series of chances, or likelihoods, that you might do something. It works like this. If I asked you how likely it is that you will eat out at a restaurant in the next month, and you were certain, or practically certain you will do this, you would answer 10. On the other hand, if you thought there was no chance, or almost no chance, of this happening, you would answer 0. If you are uncertain about doing this, you would choose another answer as close to 0 or 10 as you think it should be. Please take a moment to read this scale.

SHOWCARD B

Now, using this scale, please tell me how likely it is that you will eat out at a restaurant during the next month?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT"? DO NOT CIRCLE THIS IS A TEST QUESTION ONLY.

Still using this card, please tell me how likely you are to vote in this year's General election.

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT"? CIRCLE ONE NUMBER BELOW. REPEAT JUSTER NUMBER TO RESPONDENT AS A CHECK.

	Vote
Certain, practically certain (99/100)	10
Almost sure (9/10)	9
Very probable (8/10)	8
Probable (7/10)	7
Good possibility (6/10)	6
Fairly good possibility (5/10)	5
Fair possibility (4/10)	4
Some possibility (3/10)	3
Slight possibility (2/10)	2
Very slight possibility (1/10)	1
No chance, almost no chance (1/100)	0

Now could I collect a few details about you so we can make sure our sample is representative. Please remember that everything you tell me will be treated as strictly confidential.

9. First, in what year were you born? 19____

IF REFUSED, MAKE A GUESS AT AGE AND ENTER THIS WITH A QUESTION MARK

10. Which of these best describes your **highest** level of education?

SHOW CARD C CIRCLE ONE NUMBER ONLY

No secondary schooling	.1
Less than 3 years secondary	.2
School Certificate	3
6th Form Certificate/UE/Bursary	4
Technical or trade certificate	5
Teacher's Certificate or Diploma	.6
University Degree	7
Other	.8
Don't Know	9
Refused	10

11. Gender *(FROM OBSERVATION ONLY)*

Male 1
Female 2

Just so my supervisor can audit my work, could you tell me your name and telephone number please.

Name:_____ Phone:_____

Address:_____

Thank you very much for your help, I appreciate your assistance very much.

_____

I certify that this is a true and accurate record of the interview conducted by me, in full accordance with my instructions.

INTERVIEWER'S SIGNATURE:_____ DATE:_____

TIME:_____ DURATION:_____

SUPERVISOR'S CHECK:_____

ID_____1-4
OUTCOME_____5
SURVEY_____6
VERSION_____7

Massey University
DEPARTMENT OF MARKETING

TRADITIONAL QUESTIONNAIRE B

RECORD ADDRESS ON BACK PAGE OF THE QUESTIONNAIRE

Hello, my name is _____, and I'm a researcher from Massey University in Palmerston North We're conducting a very short survey as part of an important research project about this year's General Election; it will only take 5 to 10 minutes. Could I speak to the male/female in the household who is 18 years or older and whose birthday comes next please? Are you that person?

IF YES, SAY: Thank you. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move on to the next one. We're very interested in your ideas and opinions and, as I said, these will be held in the strictest confidence.

IF NO, ASK IF YOU CAN SPEAK TO THAT PERSON. IF UNAVAILABLE, ARRANGE A SUITABLE TIME FOR A REINTERVIEW.

Name:_____ Time:_____ Date:_____

Address:_____

IF A NEW PERSON COMES TO THE DOOR, SAY: I'm a researcher from Massey University (in Palmerston North) and we're conducting a short survey about the General election; it will only take about 5 minutes of your time. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move on to the next one. We're very interested in your ideas and opinions, and, as I said, these will be held in the strictest confidence.

1. Are you enrolled to vote in this year's General Election?

   CIRCLE ONE NUMBER ONLY

   Yes . . . . . 1 ---> Which electorate _____
   No 2
   Don't Know 3

2A Now, thinking back to the 1990 General Election, which party did you vote for?

   CIRCLE ONE NUMBER IN ROW 2A BELOW. DO NOT PROMPT

   IF OTHER, SPECIFY PARTY HERE _____

2B. If this year's General Election had been held yesterday, which party would you have voted for?

   CIRCLE ONE NUMBER IN ROW 2B BELOW. DO NOT PROMPT

   IF OTHER, SPECIFY PARTY HERE _____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q2A 1990 Voting	1	2	3	4	5	6	7
Q2B Vote Yesterday	1	2	3	4	5	6	7

3. Now please read this card and tell me which statement best describes your interest in this year's General Election.

SHOWCARD A AND CIRCLE ONE NUMBER BELOW

I am very interested in this year's election 1
I have some interest in this year's election 2
I don't really feel interested in this year's election 3
I am completely uninterested in this year's election 4

There have been several issues affecting New Zealanders discussed in the news recently, I'd like to ask you now which party you think has the best policies for dealing with some of these issues.

4A. First, which party do you think has the best healthcare policies?

CIRCLE ONE NUMBER IN ROW 4A BELOW. DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

4B. And which party do you think has the best education policies?

CIRCLE ONE NUMBER IN ROW 4B BELOW. DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

4C. Now, which party do you think has the best policies for dealing with unemployment?

CIRCLE ONE NUMBER IN ROW 4C BELOW. DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

4D. And which party do you think has the best policies for dealing with crime and violence?

CIRCLE ONE NUMBER IN ROW 4D BELOW. DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q4A Healthcare	1	2	3	4	5	6	7
Q4B Education	1	2	3	4	5	6	7
Q4C Unemployment	1	2	3	4	5	6	7
Q4D Crime & Violence	1	2	3	4	5	6	7

5. Now, I'd like to ask you some questions about your feelings towards the political parties. Generally speaking, do you think of yourself as being ... *[START READING AT *]* or what?

READ AND CIRCLE ONE NUMBER BELOW DO NOT READ SHADED BOX

Alliance 1
Labour 2
National 3
New Zealand First 4
Other . . . . . 5 Specify _____

None.....6)
Don't Know.....7) --> GO TO Q7

6. And how strongly _____(*PARTY CODED IN Q5*) do you feel?

READ AND CIRCLE ONE NUMBER BELOW

Very strongly	1
Reasonably strongly	2
Not very strongly	3
Not at all strongly	4

7. Are you currently a member of a political party?

CIRCLE ONE NUMBER

Yes	1	--->	*GO TO Q8*
No	2	)	
Refused	3	)-->	*GO TO Q9*
Don't Know	4	)	

8. Which party are you a member of?

*NOTE PARTY HERE*_____

9. Now, please look at this card. It describes a series of chances, or likelihoods, that you might do something. It works like this. If I asked you how likely it is that you will eat out at a restaurant in the next month and you were certain, or practically certain you will do this, you would answer 10. On the other hand, if you thought there was no chance, or almost no chance of this happening, you would answer 0. If you are uncertain about doing this, you would choose another answer as close to 0 or 10 as you think it should be. Please take a moment to read this scale.

SHOWCARD B

Now, using this scale, please tell me how likely it is that you will eat out at a restaurant during the next month?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT?" DO NOT CIRCLE. THIS IS A TEST QUESTION ONLY.

Still using this card, please tell me how likely you are to vote in this year's General election.

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT?" CIRCLE ONE NUMBER BELOW. REPEAT JUSTER NUMBER TO RESPONDENT AS CHECK.

	Vote
Certain, practically certain (99/100)	10
Almost sure (9/10)	9
Very probable (8/10)	8
Probable (7/10)	7
Good possibility (6/10)	6
Fairly good possibility (5/10)	5
Fair possibility (4/10)	4
Some possibility (3/10)	3
Slight possibility (2/10)	2
Very slight possibility (1/10)	1
No chance, almost no chance (1/100)	0

Now could I collect a few details about you so we can make sure our sample is representative. Please remember that everything you tell me will be treated as strictly confidential.

10. First, in what year were you born?

19____

IF REFUSED, MAKE A GUESS AT AGE AND ENTER THIS WITH A QUESTION MARK

11. Which of these best describes your **highest** level of education?

SHOW CARD C CIRCLE ONE NUMBER ONLY

No secondary schooling	.1
Less than 3 years secondary	.2
School Certificate	.3
6th Form Certificate/UE/Bursary	.4
Technical or trade certificate	5
Teacher's Certificate or Diploma	6
University Degree	7
Other	8
Don't Know	9
Refused	10

12. Gender (***FROM OBSERVATION ONLY)***

Male	.1
Female	.2

Just so my supervisor can audit my work, could you tell me your name and telephone number please.

Name:_____ Phone:_____

Address:_____

Thank you very much for your help, I appreciate your assistance very much.

_____

I certify that this is a true and accurate record of the interview conducted by me, in full accordance with my instructions.

INTERVIEWER'S SIGNATURE:_____ DATE:_____

TIME:_____ DURATION:_____

SUPERVISOR'S CHECK:_____

ID_____1-4
OUTCOME_____5
SURVEY_____6
VERSION_____7

Massey University
DEPARTMENT OF MARKETING

TRADITIONAL QUESTIONNAIRE C

RECORD ADDRESS ON BACK PAGE OF THE QUESTIONNAIRE

Hello, my name is _____, and I'm a researcher from Massey University (in Palmerston North). We're conducting a very short survey as part of an important research project about this year's General Election; it will only take 5 to 10 minutes. Could I speak to the male/female in the household who is 18 years or older and whose birthday comes next please? Are you that person?

IF YES, SAY: Thank you. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move on to the next one. We're very interested in your ideas and opinions and, as I said, these will be held in the strictest confidence.

IF NO, ASK IF YOU CAN SPEAK TO THAT PERSON. IF UNAVAILABLE, ARRANGE A SUITABLE TIME FOR A REINTERVIEW.

Name:_____ Time:_____ Date:_____

Address:_____

IF A NEW PERSON COMES TO THE DOOR, SAY:
I'm a researcher from Massey University, (in Palmerston North) and we're conducting a short survey about the General Election; it will only take about 5 minutes of your time. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move onto the next one. We're very interested in your ideas and opinions, and, as I said, these will be held in the strictest confidence.

1. First of all, please read this card and tell me which statement best describes your interest in this year's General Election.

SHOWCARD A AND CIRCLE ONE NUMBER BELOW

I am very interested in this year's election 1
I have some interest in this year's election 2
I don't really feel interested in this year's election 3
I am completely uninterested in this year's election 4

There have been several issues affecting New Zealanders discussed in the news recently, I'd like to ask you now which party you think has the best policies for dealing with some of these issues.

2A. First, which party do you think has the best healthcare policies?

CIRCLE ONE NUMBER IN ROW 2A BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

2B. And which party do you think has the best education policies?

CIRCLE ONE NUMBER IN ROW 2B BELOW DO NOT PROMPT

*IF OTHER, SPECIFY HERE*_____

2C. Now, which party do you think has the best policies for dealing with unemployment?

CIRCLE ONE NUMBER IN ROW 2C BELOW DO NOT PROMPT

*IF OTHER, SPECIFY HERE*_____

2D. And which party do you think has the best policies for dealing with crime and violence?

CIRCLE ONE NUMBER IN ROW 2D BELOW DO NOT PROMPT

*IF OTHER, SPECIFY HERE*_____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q2A. Healthcare	1	2	3	4	5	6	7
Q2B. Education	1	2	3	4	5	6	7
Q2C. Unemployment	1	2	3	4	5	6	7
Q2D. Crime & Violence	1	2	3	4	5	6	7

3. Now, I'd like to ask you some questions about your feelings towards the political parties. Generally speaking, do you think of yourself as being..... [*START READING AT* *] or what?

READ AND CIRCLE ONE NUMBER BELOW. DO NOT READ SHADED BOX

Alliance 1
Labour 2
National 3
New Zealand First 4
Other . . . . . 5 --> Specify _____

None.....6
Don't Know.....7 --> GO TO Q5

4. And how strongly _____(*PARTY CODED IN Q3*) do you feel?

READ AND CIRCLE ONE NUMBER BELOW

Very strongly 1
Reasonably strongly 2
Not very strongly 3
Not at all strongly 4

5. Are you currently a member of a political party?

CIRCLE ONE NUMBER

Yes	1	***---> GO TO Q5B***
No	2	
Refused	3	***---> GO TO Q6A***
Don't Know	4	

5B. Which party are you a member of?

*NOTE PARTY HERE*_____

6A. Are you enrolled to vote in this year's General Election?

CIRCLE ONE NUMBER ONLY

Yes . . . . . 1 ---> Which electorate _____
No 2
Don't Know 3

6B. Now, thinking back to the 1990 General Election, which party did you vote for?

CIRCLE ONE NUMBER IN ROW 6B BELOW DO NOT PROMPT

IF OTHER, SPECIFY PARTY HERE _____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q6B. 1990 Voting	1	2	3	4	5	6	7

6C. Please take this ballot paper and place a tick beside the party you would have voted for if this year's General Election had been held yesterday. Once you have done this, please fold your ballot paper and place it inside this box.

GIVE RESPONDENT A BALLOT PAPER.
NOTE THE NUMBER ON THE BALLOT PAPER HERE _____
HOLD THE BALLOT BOX FOR RESPONDENT TO PLACE HER/HIS PAPER IN.

8. Now, please look at this card. It describes a series of chances, or likelihoods, that you might do something. It works like this. If I asked you how likely it is that you will eat out at a restaurant in the next month and you were certain, or practically certain, you will do this, you would answer 10. On the other hand, if you thought there was no chance, or almost no chance, of this happening, you would answer 0. If you are uncertain about doing this, you would choose another answer as close to 0 or 10 as you think it should be. Please take a moment to read this scale.

SHOWCARD B.

Now, using this scale, please tell me how likely it is that you will eat out at a restaurant during the next month?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT"? DO NOT CIRCLE. THIS IS A TEST QUESTION ONLY.

Still using this card, please tell me how likely you are to vote in this year's General election.

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT"? CIRCLE ONE NUMBER BELOW. REPEAT JUSTER NUMBER TO RESPONDENT AS CHECK.

	Vote
Certain, practically certain (99/100)	10
Almost sure (9/10)	9
Very probable (8/10)	8
Probable (7/10)	7
Good possibility (6/10)	6
Fairly good possibility (5/10)	5
Fair possibility (4/10)	4
Some possibility (3/10)	3
Slight possibility (2/10)	2
Very slight possibility (1/10)	1
No chance, almost no chance (1/100)	0

Now could I collect a few details about you so we can make sure our sample is representative. Please remember that everything you tell me will be treated as strictly confidential.

9. First, in what year were you born? 19____

IF REFUSED, MAKE A GUESS AT AGE AND ENTER THIS WITH A QUESTION MARK

10. Which of these best describes your **highest** level of education?

SHOW CARD C CIRCLE ONE NUMBER ONLY

No secondary schooling	.1
Less than 3 years secondary	.2
School Certificate	3
6th Form Certificate/UE/Bursary	4
Technical or trade certificate	5
Teacher's Certificate or Diploma	.6
University Degree	7
Other	.8
Don't Know	9
Refused	10

11. Gender (*FROM OBSERVATION ONLY)*

Male 1
Female 2

Just so my supervisor can audit my work, could you tell me your name and telephone number please.

Name:_____ Phone:_____

Address:_____

Thank you very much for your help, I appreciate your assistance very much.

_____

I certify that this is a true and accurate record of the interview conducted by me, in full accordance with my instructions.

INTERVIEWER'S SIGNATURE:_____ DATE:_____

TIME:_____ DURATION:_____

SUPERVISOR'S CHECK:_____

ID_____1-4
OUTCOME_____5
SURVEY_____6
VERSION_____7

Massey University
DEPARTMENT OF MARKETING

TRADITIONAL QUESTIONNAIRE D

RECORD ADDRESS ON BACK PAGE OF QUESTIONNAIRE

Hello, my name is _____, and I'm a researcher from Massey University (in Palmerston North). We're conducting a very short survey as part of an important research project about this year's General Election; it will only take 5 to 10 minutes. Could I speak to the male/female in the household who is 18 years or older and whose birthday comes next please? Are you that person?

IF YES, SAY: Thank you. Everything you tell me will be strictly confidential and, if I come to any question you'd rather not answer, please tell me and I'll move on to the next one. We're very interested in your ideas and opinions and, as I said, these will be held in the strictest confidence.

IF NO, ASK IF YOU CAN SPEAK TO THAT PERSON. IF UNAVAILABLE, ARRANGE A SUITABLE TIME FOR A REINTERVIEW.

Name:_____ Time:_____ Date:_____

Address:_____

IF A NEW PERSON COMES TO THE DOOR, SAY: I'm a researcher from Massey University, (in Palmerston North) and we're conducting a short survey about the General Election; it will only take about 5 minutes of your time. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move on to the next one. We're very interested in your ideas and opinions, and, as I said, these will be held in the strictest confidence.

1. Are you enrolled to vote in this year's General Election?

   CIRCLE ONE NUMBER ONLY

   Yes . . . . . 1 ---> Which electorate _____
   No 2
   Don't Know . . . 3

2. Now, thinking back to the 1990 General Election, which party did you vote for?

   CIRCLE ONE NUMBER BELOW DO NOT PROMPT

   Alliance 1
   Labour 2
   National 3
   New Zealand First 4
   Other . . . . . 5 Specify_____
   None 6
   Don't Know 7

3. Please take this ballot paper and place a tick beside the party you would have voted for if this year's General Election had been held yesterday. Once you have done this, please fold your ballot paper and place it inside this box.

   GIVE RESPONDENT A BALLOT PAPER.
   NOTE THE NUMBER ON THE BALLOT PAPER HERE _____
   HOLD THE BALLOT BOX FOR RESPONDENT TO PLACE HER/HIS PAPER IN.

4. Now please read this card and tell me which statement best describes your interest in this year's General Election.

SHOWCARD A AND CIRCLE ONE NUMBER

I am very interested in this year's election 1
I have some interest in this year's election 2
I don't really feel interested in this year's election 3
I am completely uninterested in this year's election 4

There have been several issues affecting New Zealanders discussed in the news recently, I'd like to ask you now which party you think has the best policies for dealing with some of these issues.

5A. First, which party do you think has the best healthcare policies?

CIRCLE ONE NUMBER IN ROW 5A BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

5B. And which party do you think has the best education policies?

CIRCLE ONE NUMBER IN ROW 5B BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

5C. Now, which party do you think has the best policies for dealing with unemployment?

CIRCLE ONE NUMBER IN ROW 5C BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

5D. And, which party do you think has the best policies for dealing with crime and violence?

CIRCLE ONE NUMBER IN ROW 5D BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE_____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q5A Healthcare	1	2	3	4	5	6	7
Q5B Education	1	2	3	4	5	6	7
Q5C Unemployment	1	2	3	4	5	6	7
Q5D Crime & Violence	1	2	3	4	5	6	7

6. Now, I'd like to ask you some questions about your feelings towards the political parties. Generally speaking, do you think of yourself as being ... *[START READING AT *]* or what?

READ AND CIRCLE ONE NUMBER BELOW DO NOT READ SHADED BOX

Alliance 1
Labour 2
National 3
New Zealand First 4
Other . . . . . 5 Specify _____

None.....6
Don't Know.....7 --> GO TO Q8

7. And how strongly _____(*PARTY CODED IN Q6*) do you feel?

READ AND CIRCLE ONE NUMBER BELOW

Very strongly	1
Reasonably strongly	2
Not very strongly	3
Not at all strongly	4

8. Are you currently a member of a political party?

CIRCLE ONE NUMBER

Yes	1	--->	*GO TO Q9*
No	2	)	
Refused	3	)-->	*GO TO Q10*
Don't Know	4	)	

9. Which party are you a member of?

*NOTE PARTY HERE*_____

10. Now, please look at this card. It describes a series of chances, or likelihoods, that you might do something. It works like this. If I asked you how likely it is that you will eat out at a restaurant in the next month, and you were certain, or practically certain, you will do this, you would answer 10. On the other hand, if you thought there was no chance, or almost no chance, of this happening, you would answer 0. If you are uncertain about doing this, you would choose another answer as close to 0 or 10 as you think it should be. Please take a moment to read this scale.

SHOWCARD B

Now, using this scale, please tell me how likely it is that you will eat out at a restaurant during the next month?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT?" DO NOT CIRCLE. THIS IS A TEST QUESTION ONLY.

Still using this card, please tell me how likely you are to vote in this year's General election?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT?" CIRCLE ONE NUMBER BELOW. REPEAT JUSTER NUMBER TO RESPONDENTS AS A CHECK.

	Vote
Certain, practically certain (99/100)	10
Almost sure (9/10)	9
Very probable (8/10)	8
Probable (7/10)	7
Good possibility (6/10)	6
Fairly good possibility (5/10)	5
Fair possibility (4/10)	4
Some possibility (3/10)	3
Slight possibility (2/10)	2
Very slight possibility (1/10)	1
No chance, almost no chance (1/100)	0

Now could I collect a few details about you so we can make sure our sample is representative. Please remember that everything you tell me will be treated as strictly confidential.

14. First, in what year were you born?

19____

IF REFUSED, MAKE A GUESS AT AGE AND ENTER THIS WITH A QUESTION MARK

15. Which of these best describes your **highest** level of education?

SHOW CARD C CIRCLE ONE NUMBER ONLY

No secondary schooling	.1
Less than 3 years secondary	.2
School Certificate	.3
6th Form Certificate/UE/Bursary	.4
Technical or trade certificate	5
Teacher's Certificate or Diploma	6
University Degree	7
Other	8
Don't Know	9
Refused	10

16. Gender (*FROM OBSERVATION ONLY)*

Male	I
Female	.2

Just so my supervisor can audit my work, could you tell me your name and telephone number please.

Name:_____ Phone:_____

Address:_____

Thank you very much for your help, I appreciate your assistance very much.

_____

I certify that this is a true and accurate record of the interview conducted by me, in full accordance with my instructions.

INTERVIEWER'S SIGNATURE:_____ DATE:_____

TIME:_____ DURATION:_____

SUPERVISOR'S CHECK:_____

ID _____1-4
OUTCOME _____5
VERSION_____6
SURVEY_____7

Massey University
DEPARTMENT OF MARKETING

JUSTER QUESTIONNAIRE A

RECORD ADDRESS ON BACK PAGE OF QUESTIONNAIRE

Hello, my name is _____, and I'm a researcher from Massey University (in Palmerston North). We're conducting a very short survey as part of an important research project about this year's General Election; it will only take 5 to 10 minutes. Could I speak to the male/female in the household who is 18 years or older and whose birthday comes next please. Are you that person?

IF YES, SAY: Thank you. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move on to the next one. We're very interested in your ideas and opinions, and, as I said, these will be held in the strictest confidence.

IF NO, ASK IF YOU CAN SPEAK TO THAT PERSON. IF UNAVAILABLE, ARRANGE A SUITABLE TIME FOR A REINTERVIEW.

Name:_____ Time:_____ Date:_____

Address:_____

IF DIFFERENT PERSON COMES TO THE DOOR, SAY:
I'm a researcher from Massey University (in Palmerston North) and we're conducting a short survey about the General Election; it will only take about 5 minutes of your time. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move onto the next one. We're very interested in your ideas and opinions, and, as I said, these will be held in the strictest confidence.

1. First of all, please read this card and tell me which statement best describes your interest in this year's General Election.

SHOWCARD A AND CIRCLE ONE NUMBER BELOW

I am very interested in this year's election 1
I have some interest in this year's election 2
I don't really feel interested in this year's election 3
I am completely uninterested in this year's election 4

There have been several issues affecting New Zealanders discussed in the news recently, I'd like to ask you now which party you think has the best policies for dealing with some of these issues.

2A. First, which party do you think has the best healthcare policies?

CIRCLE ONE NUMBER IN ROW 2A BELOW DO NOT PROMPT

*IF OTHER, SPECIFY HERE*_____

2B. And which party do you think has the best education policies?

CIRCLE ONE NUMBER IN ROW 2B BELOW ***DO NOT PROMPT***

*IF OTHER, SPECIFY HERE*_____

2C. Now, which party do you think has the best policies for dealing with unemployment?

CIRCLE ONE NUMBER IN ROW 2C BELOW ***DO NOT PROMPT***

*IF OTHER, SPECIFY HERE*_____

2D. And which party do you think has the best policies for dealing with crime and violence?

CIRCLE ONE NUMBER IN ROW 2D BELOW ***DO NOT PROMPT***

*IF OTHER, SPECIFY HERE*_____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q2A Healthcare	1	2	3	4	5	6	7
Q2B Education	1	2	3	4	5	6	7
Q2C Unemployment	1	2	3	4	5	6	7
Q2D Crime & Violence	1	2	3	4	5	6	7

3. Now, I'd like to ask you some questions about your feelings towards the political parties. Generally speaking, do you think of yourself as being..... [*START READING AT* *] or what?

READ AND CIRCLE ONE NUMBER BELOW DO NOT READ SHADED BOX

Alliance 1
Labour 2
National 3
New Zealand First 4
Other . . . . . 5 --> Specify _____

None.....6)
Don't Know.....7) --> GO TO Q5

4. And how strongly _____(*PARTY CODED IN Q3*) do you feel?

READ AND CIRCLE ONE NUMBER BELOW

Very strongly 1
Reasonably strongly . . . 2
Not very strongly 3
Not at all strongly 4

5. Are you currently a member of a political party?

CIRCLE ONE NUMBER

Yes	1	--->	*GO TO Q5B*
No	2	)	
Refused	3	)--->	*GO TO Q6A*
Don't Know	4	)	

5B. Which party are you a member of?

*NOTE PARTY HERE*_____

6A. Are you enrolled to vote in this year's General Election?

CIRCLE ONE NUMBER ONLY

Yes	1	---> In which electorate? _____=
No	2	
Don't Know	3	

6B. Now, thinking back to the 1990 General Election, which party did you vote for?

CIRCLE ONE NUMBER IN ROW 6B BELOW DO NOT PROMPT

IF OTHER, NOTE PARTY HERE _____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q6B 1990 Voting	1	2	3	4	5 . . .	6	7

7. Now, please look at this card. It describes a series of chances, or likelihoods, that you might do something. It works like this. If I asked you how likely it is that you will eat out at a restaurant in the next month, and you were certain, or practically certain you will do this, you would answer 10. On the other hand, if you thought there was no chance, or almost no chance, of this happening, you would answer 0. If you are uncertain about doing this, you would choose another answer as close to 0 or 10 as you think it should be. Please take a moment to read this scale.

SHOWCARD B

Now, using this scale, please tell me how likely it is that you will eat out at a restaurant during the next month?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT"? DO NOT CIRCLE. THIS IS A TEST QUESTION ONLY.

Still using this card, if this year's General election had been held yesterday, please tell me how likely it is that you would have voted for the ***(START AT ASTERISKED PARTY)*** candidate?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT"? CIRCLE APPROPRIATE NUMBER IN ASTERISKED COLUMN AND REPEAT QUESTION FOR REMAINING PARTIES WORKING TO THE RIGHT OF THE ASTERISK. ALWAYS ASK OTHER LAST. REPEAT JUSTER NUMBER TO RESPONDENT AS A CHECK.

	Alliance	Labour	National	New Zealand First	Other	Vote
Certain/practically certain (99/100)	10	10	10	10	10	10
Almost sure (9/10)	9	9	9	9	9	9
Very probable (8/10)	8	8	8	8	8	8
Probable (7/10)	7	7	7	7	7	7
Good possibility (6/10)	6	6	6	6	6	6
Fairly good possibility (5/10)	5	5	5	5	5	5
Fair possibility (4/10)	4	4	4	4	4	4
Some possibility (3/10)	3	3	3	3	3	3
Slight possibility (2/10)	2	2	2	2	2	2
Very slight possibility (1/10)	1	1	1	1	1	1
No chance, almost no chance (1/100)	0	0	0	0	0	0

8. Now, still using this card, please tell me how likely you are to vote in this year's General Election.

CIRCLE IN VOTE COLUMN ABOVE FOLLOW SAME INSTRUCTIONS AS FOR Q7.

Now could I collect a few details about you so we can make sure our sample is representative. Please remember that everything you tell me will be treated as strictly confidential.

10. First, in what year were you born? 19____

IF REFUSED, MAKE A GUESS AT AGE AND ENTER THIS WITH A QUESTION MARK

11. Which of these best describes your **highest** level of education?

SHOWCARD C. ***CIRCLE ONE NUMBER ONLY***

No secondary schooling 1
Less than 3 years secondary 2
School Certificate 3
6th Form Certificate/UE/Bursary 4
Technical or trade certificate 5
Teacher's Certificate or Diploma 6
University Degree 7
Other 8
Don't Know 9
Refused 10

15. Gender (***FROM OBSERVATION ONLY)***

Male 1
Female 2

Just so my supervisor can CHECK my work, could you tell me your name and telephone number please.

Name:_____ Phone:_____

Address:_____

Thank you very much for your help, I appreciate your assistance very much.

_____

I certify that this is a true and accurate record of the interview conducted by me, in full accordance with my instructions.

INTERVIEWER'S SIGNATURE:_____ DATE: _____

TIME:_____ DURATION:_____

SUPERVISOR'S CHECK:_____

ID_____1-4
OUTCOME_____5
SURVEY_____6
VERSION_____7

Massey University
DEPARTMENT OF MARKETING

JUSTER QUESTIONNAIRE B

RECORD ADDRESS ON BACK PAGE OF QUESTIONNAIRE

Hello, my name is _____, and I'm a researcher from Massey University (in Palmerston North). We're conducting a very short survey as part of an important research project about this year's General Election; it will only take 5 to 10 minutes. Could I speak to the male/female in the household who is 18 years or older and whose birthday comes next please. Are you that person?

IF YES, SAY: Thank you. Everything you say will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move on to the next one. We're very interested in your ideas and opinions, and as I said, these will be held in the strictest confidence.

IF NO, ASK IF YOU CAN SPEAK TO THAT PERSON. IF UNAVAILABLE, ARRANGE A SUITABLE TIME FOR A REINTERVIEW.

Name:_____ Time:_____ Date:_____

Address:_____

IF DIFFERENT PERSON COMES TO THE DOOR, SAY:
I'm a researcher from Massey University, (in Palmerston North) and we're conducting a short survey about the General Election; it will only take about 5 minutes of your time. Everything you tell me will be strictly confidential and, if we come to any question you'd rather not answer, please tell me and I'll move onto the next one. We're very interested in your ideas and opinions, and, as I said, these will be held in the strictest confidence.

1. Are you enrolled to vote in this year's General Election?

CIRCLE ONE NUMBER ONLY

Yes . . . . . 1 ---> Which electorate _____
No 2
Don't Know . . 3

2. Thinking back to the 1990 General Election, which party did you vote for?

CIRCLE ONE NUMBER BELOW DO NOT PROMPT

Alliance	1	
Labour	2	
National	3	
New Zealand First	4	
Other	5	Specify_____
None	6	
Don't Know	7	

3. Now, please look at this card It describes a series of chances, or likelihoods, that you might do something. It works like this. If I asked you how likely it is that you will eat out at a restaurant in the next month, and you were certain, or practically certain you will do this, you would answer 10. On the other hand, if you thought there was no chance, or almost no chance, of this happening, you would answer 0. If you are uncertain about doing this, you would choose another answer as close to 0 or 10 as you think it should be. Please take a moment to read this scale.

SHOWCARD B

Now, using this scale, please tell me how likely it is that you will eat out at a restaurant during the next month?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT?" DO NOT CIRCLE. THIS IS A TEST QUESTION ONLY.

Still using this card, if this year's General Election had been held yesterday, please tell me how likely it is that you would have voted for the *(START AT ASTERISKED PARTY)* candidate?

IF RESPONDENT ANSWERS WITH A WORD DESCRIPTION, ASK "WHAT NUMBER IS THAT?" CIRCLE APPROPRIATE NUMBER IN ASTERISKED COLUMN AND REPEAT QUESTION FOR REMAINING PARTIES WORKING TO THE RIGHT OF THE ASTERISK. ALWAYS ASK OTHER LAST. REPEAT JUSTER NUMBER TO RESPONDENT AS A CHECK.

	Alliance	Labour	National	New Zealand First	Other	Vote
Certain/practically certain (99/100)	10	10	10	10	10	10
Almost sure (9/10)	9	9	9	9	9	9
Very probable (8/10)	8	8	8	8	8	8
Probable (7/10)	7	7	7	7	7	7
Good possibility (6/10)	6	6	6	6	6	6
Fairly good possibility (5/10)	5	5	5	5	5	5
Fair possibility (4/10)	4	4	4	4	4	4
Some possibility (3/10)	3	3	3	3	3	3
Slight possibility (2/10)	2	2	2	2	2	2
Very slight possibility (1/10)	1	1	1	1	1	1
No chance, almost no chance (1/100)	0	0	0	0	0	0

3A. Now, still using this card, please tell me how likely you are to vote in this year's General election?

CIRCLE IN VOTE COLUMN ABOVE FOLLOW SAME INSTRUCTIONS AS FOR Q2

4. Now, please read this card and tell me which statement best describes your interest in this year's General Election?

SHOWCARD A AND CIRCLE ONE NUMBER

I am very interested in this year's election 1
I have some interest in this year's election 2
I don't really feel interested in this year's election 3
I am completely uninterested in this year's election 4

There have been several issues affecting New Zealanders discussed in the news recently, I'd like to ask you now which party you think has the best policies for dealing with some of these issues.

5A. First, which party do you think has the best healthcare policies?

CIRCLE ONE NUMBER IN ROW 5A BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

5B. And which party do you think has the best education policies?

CIRCLE ONE NUMBER IN ROW 5B BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

5C. Now, which party do you think has the best policies for dealing with unemployment?

CIRCLE ONE NUMBER IN ROW 5C BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

5D. And which party do you think has the best policies for dealing with crime and violence?

CIRCLE ONE NUMBER IN ROW 5D BELOW DO NOT PROMPT

IF OTHER, SPECIFY HERE _____

	Alliance	Labour	National	New Zealand First	Other	None	Don't Know
Q5A Healthcare	1	2	3	4	5	6	7
Q5B Education	1	2	3	4	5	6	7
Q5C Unemployment	1	2	3	4	5	6	7
Q5D Crime & Violence	1	2	3	4	5	6	7

6. Now, I'd like to ask you some questions about your feelings towards the political parties. Generally speaking, do you think of yourself as being..... [*START READING AT **] or what?

READ AND CIRCLE ONE NUMBER BELOW. DO NOT READ SHADED BOX

Alliance 1
Labour 2
National 3
New Zealand First 4
Other . . . . . 5 --> Specify _____

None.....6
Don't Know.....7 --> GO TO Q8

7. And how strongly _____(*PARTY CODED IN Q6*) do you feel?

READ AND CIRCLE ONE NUMBER BELOW

Very strongly 1
Reasonably strongly 2
Not very strongly 3
Not at all strongly 4

8. Are you currently a member of a political party?

CIRCLE ONE NUMBER

Yes . . . . . 1 ---> *GO TO Q9*
No . . . . . 2)
Refused . . . . . 3)---> *GO TO Q10*
Don't Know . . . . . 4)

9. Which party are you a member of?

*NOTE PARTY HERE*_____

Now could I collect a few details about you so I can make sure my sample is representative. Please remember that everything you tell me will be treated as strictly confidential.

10. First, in what year were you born? 19____

IF REFUSED, MAKE A GUESS AT AGE AND ENTER THIS WITH A QUESTION MARK

11. Which of these best describes your highest level of education?

SHOWCARD C CIRCLE ONE NUMBER ONLY

No secondary schooling 1
Less than 3 years secondary 2
School Certificate 3
6th Form Certificate/UE/Bursary 4
Technical or trade certificate 5
Teacher's Certificate or Diploma 6
University Degree 7
Other 8
Don't Know 9
Refused 10

12. Gender (***FROM OBSERVATION ONLY***)

Male 1
Female 2

Just so my supervisor can audit my work, could you tell me your name and telephone number please.

Name:_____ Phone:_____

Address:_____

Thank you very much for your help, I appreciate your assistance very much.

_____

I certify that this is a true and accurate record of the interview conducted by me, in full accordance with my instructions.

INTERVIEWER'S SIGNATURE:_____ DATE:_____

TIME:_____ DURATION:_____

SUPERVISOR'S CHECK:_____

APPENDIX D:

VOTING RESULTS FROM THE

MANAWATU, MIRAMAR AND WESTERN HUTT ELECTORATES

PART II: GENERAL STATISTICS

(A) Electoral District	Votes Cast: Valid (B) Ordinary Votes	Valid (C) Special Votes	Valid (D) Total (B + C)	Informal (E) Ordinary Votes	Informal (F) Special Votes	(G) Total (E + F)	(H) Special Votes Disallowed	(I) Total (D+G+H)	(J) Electors on Master Roll	(K) Electoral Population	(L) Percentage of (a) Votes Cast to Electors on Master Roll (I:J)	(L) Percentage of (b) Valid Votes Recorded for Candidate Elected	(L) Percentage of (c) Informal Votes Recorded (G:D + G)
Albany ..	19639	1993	21632	78	6	84	533	22249	25497	32900	87.26	37.75	0.39
Auckland Central	18213	2498	20709	113	22	135	1001	21845	24414	34378	89.46	39.14	0.65
Avon ..	17862	1542	19404	73	13	86	395	19885	23514	34013	84.67	50.69	0.44
Awarua ..	16776	1356	18134	86	13	99	199	18432	22180	32180	83.10	43.02	0.54
Birkenhead ..	19494	1770	21264	78	10	88	516	21868	24730	34812	88.43	38.99	0.41
Christchurch Central	15731	2094	17825	131	20	151	613	18589	22502	33866	82.61	66.21	0.84
Christchurch North	19905	1678	21583	115	13	128	284	21995	26045	33280	87.82	53.77	0.59
Clutha ..	18194	1577	19771	71	11	82	197	20050	23612	33237	86.28	47.67	0.41
Dunedin North ..	19091	1859	20950	158	26	184	344	21478	24495	35107	87.68	43.53	0.87
Dunedin West ..	19320	1547	20867	97	7	104	281	21252	24606	32921	86.71	47.56	0.50
East Coast Bays	20060	1781	21841	104	18	122	340	22303	25179	33282	88.68	46.74	0.56
Eastern Bay of Plenty	16332	1091	17423	65	9	74	317	17814	21314	33655	83.68	37.58	0.42
Eastern Hutt ..	16799	1409	18208	104	14	118	400	18726	21901	31986	85.50	51.33	0.64
Eden ..	18358	2135	20493	156	19	175	665	21334	24164	34071	88.22	49.39	0.86
Fer North ..	16059	1417	17476	69	13	82	478	18036	21681	34291	83.26	39.92	0.47
Fendalton ..	16646	2156	20602	159	29	188	375	21365	24122	33260	88.57	51.76	0.90
Franklin ..	18737	1216	17953	88	10	98	270	18321	22392	32272	81.82	43.02	0.54
Gisborne ..	18011	1184	19195	71	3	74	366	19635	23029	34989	85.26	42.80	0.38
Glenfield ..	19296	1773	21069	83	11	94	485	21648	24771	34453	87.39	38.51	0.44
Hamilton East ..	18631	1975	20506	80	12	92	569	21157	24104	34755	87.77	37.44	0.45
Hamilton West ..	16872	1789	18661	76	13	89	437	19187	22742	33438	84.37	38.14	0.47
Hastings ..	18656	1332	19988	129	10	139	403	20530	23188	33044	88.54	46.03	0.69
Hauraki ..	16864	1862	18526	75	6	81	313	18920	22291	32508	84.88	40.85	0.44
Hawkes Bay ..	17595	1858	19453	120	23	143	447	20043	23165	33889	86.52	48.69	0.73
Henderson ..	14626	1399	16025	59	4	63	650	16738	19850	32585	84.32	39.82	0.39
Heretaunga ..	18046	1347	19393	53	11	64	255	19712	22477	33017	87.70	41.15	0.33
Hobson ..	16322	1601	17923	100	8	108	352	18383	22006	33273	83.54	36.12	0.60
Horowhenua ..	18708	1337	20045	82	7	89	344	20478	23267	31852	88.01	41.84	0.44
Howick ..	20578	1752	22330	115	18	133	361	22824	26286	33873	86.83	47.81	0.59
Invercargill ..	18008	1407	19415	104	17	121	221	19757	23033	32575	85.78	43.66	0.62
Island Bay ..	17979	2642	20621	184	22	206	765	21592	24417	34021	88.43	49.03	0.99
Kaimai ..	16151	1653	19804	61	14	75	351	20230	24069	32025	84.08	33.18	0.38
Kaipara ..	16613	1999	18612	72	1	73	287	18972	22544	32318	84.16	40.09	0.39
Kapiti ..	21950	1844	23794	75	12	87	326	24207	26745	34730	90.51	38.84	0.36
King Country ..	15379	1299	16678	107	21	128	310	17116	21288	32894	80.40	50.34	0.76
Lyttelton ..	21485	2196	23681	111	16	127	343	24151	28997	34782	89.46	40.81	0.53
Manawatu ..	17924	1606	19732	74	12	86	355	20173	23553	32768	85.86	33.39	0.43
Mangere ..	12959	1279	14238	77	17	94	805	15137	18237	34264	63.00	56.61	0.66
Manurewa ..	15438	1362	16800	61	12	63	619	17482	21484	35072	81.37	44.50	0.37

PART II: GENERAL STATISTICS —*continued*

(A) Electoral District	Votes Cast: Valid (B) Ordinary Votes	Valid (C) Special Votes	Valid (D) Total (B + C)	Informal (E) Ordinary Votes	Informal (F) Special Votes	(G) Total (E + F)	(H) Special Votes Disallowed	(I) Total (D+G+H)	(J) Electors on Master Roll	(K) Electoral Population	(L) Percentage of (a) Votes Cast to Electors on Master Roll (I:J)	(L) Percentage of (b) Valid Votes Recorded for Candidate Elected	(L) Percentage of (c) Informal Votes Recorded (G:D + G)
Marlborough ..	20316	1453	21769	156	14	172	146	22089	25984	34819	85.01	41.04	0.78
Matakana ..	17485	1441	16926	100	13	113	305	19344	22666	31905	86.34	33.47	0.59
Matamata ..	16371	1147	17518	94	9	103	244	17865	22317	33082	80.06	50.92	0.58
Miramar ..	19560	2739	22299	128	22	150	527	22976	25972	34973	88.46	46.42	0.67
Mt Albert ..	17196	2121	19317	105	26	131	813	20261	23146	33461	87.54	49.42	0.67
Napier ..	18136	1469	19606	119	10	129	449	20163	23482	33694	85.95	50.61	0.65
Nelson ..	20599	1411	22010	160	23	183	239	22432	26486	34982	84.69	39.89	0.82
New Lynn ..	16287	1660	17947	45	12	57	743	18747	21376	32135	87.69	38.88	0.32
New Plymouth ..	19695	1704	21399	90	12	102	248	21749	25085	34462	86.70	48.98	0.47
North Shore ..	20400	1943	22343	125	11	136	488	22967	25778	34170	89.10	45.61	0.61
Onehunga ..	18061	2145	20206	87	12	99	632	20937	23929	33217	87.60	35.56	0.49
Onslow ..	19683	2172	21855	82	14	96	342	22293	25126	34069	88.72	41.62	0.44
Otago ..	16164	1990	18154	97	22	119	364	18637	21927	32065	85.00	45.47	0.65
Otara ..	11825	1357	13182	152	68	220	884	14286	17411	31861	82.05	61.30	1.64
Pahiatua ..	16517	1164	17681	102	10	112	236	18029	21819	32274	82.63	49.22	0.63
Pakuranga ..	18493	1542	20035	90	9	99	362	20496	23638	33166	85.98	48.01	0.49
Palmerston North	17800	1795	19595	96	14	110	340	20045	23504	33819	85.28	46.18	0.56
Panmure ..	13874	1523	15397	79	24	103	983	16463	19289	35027	86.45	52.14	0.66
Papakura ..	15905	1404	17309	69	5	74	451	17834	21231	32623	84.00	36.80	0.43
Papatoetoe ..	16729	1473	18202	77	8	85	689	18976	22095	34728	86.88	53.57	0.48
Pencarrow ..	16926	1418	18344	77	7	84	433	18861	21693	32312	86.95	44.19	0.46
Porirua ..	14569	1375	15944	99	23	122	529	16596	19592	34161	84.70	57.47	0.76
Raglan ..	16367	2004	18391	100	14	114	412	18917	22318	33170	84.76	40.75	0.62
Rakaia ..	20133	1244	21377	141	19	160	174	21711	25341	34094	85.68	48.38	0.74
Rangiora ..	19116	1366	20480	121	11	132	204	20816	24085	32778	86.43	44.71	0.64
Rangitikei ..	17326	1628	18954	145	10	155	286	19395	22917	33187	84.63	44.03	0.81
Remuera ..	18641	2350	20991	189	19	208	666	21865	24636	33935	88.75	59.95	0.98
Roskill ..	17879	1471	19350	73	9	82	583	20015	22789	32637	87.83	39.81	0.42
Rotorua ..	16897	1343	17240	84	14	96	526	17864	21376	32693	83.57	34.18	0.57
Selwyn ..	20988	1961	22949	83	14	97	387	23433	27162	34698	86.27	41.23	0.42
St Albans ..	19705	1880	21585	121	20	141	461	22187	25251	33998	87.83	46.43	0.65
St Kilda ..	19625	1660	21285	107	11	118	380	21783	24821	32926	87.76	47.07	0.55
Sydenham ..	19807	1765	21591	70	8	76	377	22046	25069	33927	87.94	67.74	0.36
Tamaki ..	18880	1993	20873	125	10	135	537	21545	24099	32077	89.40	65.40	0.64
Taranaki ..	16150	1332	17482	72	21	93	215	17790	21906	32251	81.21	50.97	0.53
Tarawera ..	16897	1477	18374	74	9	83	408	18865	22610	34326	83.44	42.85	0.45
Tasman ..	20271	1269	21540	114	12	126	210	21876	24962	35207	87.64	51.15	0.58
Tauranga ..	21267	1576	22843	117	9	126	286	23255	27082	34230	85.87	55.33	0.55
Te Atatu ..	17084	1645	18729	72	9	81	617	19327	22588	32010	85.56	38.78	0.43
Timaru ..	20316	1302	21617	86	6	92	179	21886	24902	33324	87.90	46.97	0.42

PART II: GENERAL STATISTICS —*continued*

(A) Electoral District	Votes Cast: Valid (B) Ordinary Votes	Valid (C) Special Votes	Valid (D) Total (B + C)	Informal (E) Ordinary Votes	Informal (F) Special Votes	(G) Total (E + F)	(H) Special Votes Disallowed	(I) Total (D+G+H)	(J) Electors on Master Roll	(K) Electoral Population	(L) Percentage of (a) Votes Cast to Electors on Master Roll (I:J)	(L) Percentage of (b) Valid Votes Recorded for Candidate Elected	(L) Percentage of (c) Informal Votes Recorded (G:D + G)
Titirangi ..	17119	1696	18815	73	9	82	494	19391	22167	32374	87.52	34.07	0.43
Tongariro ..	14972	1298	16270	113	12	125	392	16787	20311	34006	82.65	47.47	0.76
Waikaremoana ..	15460	1304	16784	134	13	147	376	17307	20937	32420	82.68	51.11	0.87
Waikato ..	16657	1624	17281	75	6	81	426	17788	21345	32702	83.34	39.72	0.47
Waipa ..	17308	1282	18590	99	0	99	276	18965	23479	33483	80.77	44.20	0.53
Wairarapa ..	20218	1422	21640	92	11	103	182	21925	24963	34753	87.83	41.61	0.47
Waitakere ..	17789	1951	19740	85	14	99	670	20409	23573	34940	86.68	41.96	0.50
Waitaki ..	19133	1521	20654	106	11	117	160	20931	23660	32339	88.39	41.31	0.56
Waitotara ..	17378	1295	18674	110	4	114	259	19047	23022	32838	82.73	45.50	0.61
Wallace ..	16225	1353	17578	98	6	104	158	17840	21977	31927	81.18	55.93	0.59
Wanganui ..	19014	1190	20204	84	8	92	274	20570	23612	34764	87.12	36.62	0.45
Wellington—Karori	20626	3347	23973	113	25	138	452	24563	26786	35049	91.70	43.95	0.57
West Coast ..	18247	1035	19282	81	4	85	149	19516	22469	32180	86.86	46.26	0.44
Western Hutt ..	16696	1913	20609	93	17	110	406	21128	23843	32603	88.80	43.88	0.53
Whangarei ..	17238	1646	18884	100	8	108	333	19325	23002	32727	84.01	34.46	0.57
Yaldhurst ..	21094	1720	22814	117	22	139	235	23188	26568	34116	87.26	46.72	0.61
General Totals ..	**1702360**	**156944**	**1859304**	**9402**	**1286**	**10688**	**38962**	**1908954**	**2220079**	**3178486**	**85.99**		**0.57**
Eastern Maori	11464	3194	14658	215	86	301	1411	16370	23779	45041	68.84	83.52	2.01
Northern Maori	10563	6187	16750	68	43	111	1286	18147	26041	46394	69.69	33.76	0.66
Southern Maori	8433	8187	16620	61	118	179	936	17735	25964	43897	68.31	57.95	1.07
Western Maori	9589	5875	15464	30	55	85	1337	16886	25801	47125	65.45	60.74	0.55
Maori Totals ..	**40049**	**23443**	**63492**	**374**	**302**	**676**	**4970**	**69138**	**101585**	**181457**	**68.06**		**1.06**
Combined Totals	**1742409**	**180387**	**1922796**	**9776**	**1588**	**11364**	**43932**	**1978092**	**2321664**	**3359943**	**85.20**		**0.59**

APPENDIX E:

CONTEXT EFFECTS:

SUPPLEMENTARY TABLES

Table E.1: Contextual Effect on Voting Intention Estimates - Electorate Level Results

Manawatu - Round 1 Interviews								
	Contextual (N = 102)				Non-Contextual (N = 112)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	11.8	16.0	27.5	-11.0	8.4	15.5	27.5	-12.0
Labour	33.3	45.3	33.4	+11.9	19.6	36.2	33.4	+2.8
National	22.5	30.7	32.6	-1.9	22.4	41.4	32.6	+8.8
NZ First	4.9	6.5	4.6	+1.9	2.8	5.2	4.6	+0.6
Other	1.0	1.3	1.9	-0.6	0.9	1.7	1.9	-0.2
None	4.9	-	-	-	8.4	-	-	-
Undecided	21.6	-	-	-	34.6	-	-	-
Refused	0.0	-	-	-	2.8	-	-	-
TOTAL	100.0	100.0	100.0	5.5[1]	100.0	100.0	100.0	4.9[1]

Miramar - Round 1 Interviews								
	Contextual (N = 116)				Non-Contextual (N = 112)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	3.5	4.7	11.6	-6.9	8.0	12.3	11.6	+0.7
Labour	36.5	49.4	46.4	+3.0	30.4	46.6	46.4	+0.2
National	27.0	36.5	34.8	-1.7	21.4	32.9	34.8	-1.9
NZ First	6.1	8.2	4.3	+3.9	5.4	8.2	4.3	+3.9
Other	0.9	1.2	2.9	-1.7	0.0	0.0	2.9	-2.9
None	10.4	-	-	-	8.0	-	-	-
Undecided	13.9	-	-	-	22.3	-	-	-
Refused	1.7	-	-	-	4.5	-	-	-
TOTAL	100.0	100.0	100.0	3.4[1]	100.0	100.0	100.0	1.9[1]

Table E.1: continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 134)				Non-Contextual (N = 132)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	6.0	7.3	13.6	-6.3	8.3	12.5	13.6	-1.1
Labour	32.8	40.0	37.9	+2.1	26.5	39.8	37.9	+1.9
National	32.8	40.0	41.6	-1.6	29.5	44.3	41.6	+2.7
NZ First	8.2	10.0	3.8	+6.2	2.3	3.4	3.8	-0.4
Other	2.2	2.7	3.1	-0.4	0.0	0.0	3.1	-3.1
None	5.2	-	-	-	11.4	-	-	-
Undecided	12.7	-	-	-	19.7	-	-	-
Refused	0.0	-	-	-	2.3	-	-	-
TOTAL	100.0	100.0	100.0	3.3[1]	100.0	100.0	100.0	1.8[1]

Manawatu - Round 2 Interviews								
	Contextual (N = 128)				Non-Contextual (N = 126)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	10.9	13.6	27.5	-13.9	15.9	22.2	27.5	-5.3
Labour	34.4	42.7	33.4	+9.3	22.2	31.1	33.4	-2.3
National	28.9	35.9	32.6	+3.3	28.6	40.0	32.6	+7.4
NZ First	5.5	6.8	4.6	+2.2	4.8	6.7	4.6	+2.1
Other	0.8	1.0	1.9	-0.9	0.0	0.0	1.9	-1.9
None	10.9	-	-	-	4.0	-	-	-
Undecided	8.6	-	-	-	23.8	-	-	-
Refused	0.0	-	-	-	0.8	-	-	-
TOTAL	100.0	100.0	100.0	5.5[2]	100.0	100.0	100.0	3.8[1]

Table E.1: continued

Miramar - Round 2 Interviews								
	Contextual (N = 111)				Non-Contextual (N = 110)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	10.0	12.2	11.6	+0.6	8.3	12.2	13.6	+0.6
Labour	26.4	32.2	46.4	-14.2	36.7	54.1	46.4	+7.7
National	38.2	46.7	34.8	+11.9	18.3	27.0	34.8	-7.8
NZ First	5.5	6.7	4.3	+2.4	4.6	6.9	4.3	+2.6
Other	1.8	2.2	2.9	-0.7	0.0	0.0	2.9	-2.9
None	5.5	-	-	-	6.4	-	-	-
Undecided	11.8	-	-	-	22.9	-	-	-
Refused	0.9	-	-	-	2.8	-	-	-
TOTAL	100.0	100.0	100.0	6.2[1]	100.0	100.0	100.0	4.3[1]

Western Hutt - Round 2 Interviews								
	Contextual (N = 124)				Non-Contextual (N = 129)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	8.9	10.6	13.6	-3.0	11.6	15.3	13.6	+1.7
Labour	26.6	31.7	37.9	-6.2	27.1	35.4	37.9	-2.5
National	40.3	48.1	41.6	+6.5	31.0	40.4	41.6	-1.2
NZ First	6.5	7.7	3.8	+3.9	6.2	8.1	3.8	+4.3
Other	1.6	1.9	3.1	-1.2	0.8	1.0	3.1	-2.1
None	3.2	-	-	-	7.0	-	-	-
Undecided	11.3	-	-	-	15.5	-	-	-
Refused	1.6	-	-	-	0.8	-	-	-
TOTAL	100.0	100.0	100.0	2.9[1]	100.0	100.0	100.0	2.4[1]

1. Mean absolute error

Table E2: Contextual Effects on Voting Intention Estimates: Predicted by Self-Reported Behaviour

Manawatu Round 1						
Contextual (N=45)				Non-Contextual (N=45)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	18.8	20.5	-1.7	16.7	14.6	+2.1
Labour	43.7	46.1	-2.4	37.5	43.9	-6.4
National	31.3	28.3	+3.0	45.8	39.1	+6.7
Other	6.2	5.1	+1.1	0.0	2.4	-2.4
TOTAL	100.0	100.0	2.1[1]	100.0	100.0	4.4[1]

Miramar Round 1						
Contextual (N=38)				Non-Contextual (N=35)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	6.9	12.1	-5.2	8.3	12.1	-3.8
Labour	48.3	45.5	+2.8	25.0	54.5	-29.5
National	37.9	36.4	+1.5	54.2	30.3	+23.9
Other	6.9	6.0	+0.9	12.5	3.1	+9.4
TOTAL	100.0	100.0	2.6[1]	100.0	100.0	16.7[1 2]

Western Hutt - Round 1						
Contextual (N=73)				Non-Contextual (N=74)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	9.2	14.7	-5.5	16.0	14.9	+1.1
Labour	41.5	39.7	+1.8	34.0	35.8	-1.8
National	38.5	36.8	+1.7	46.0	43.3	+2.7
Other	10.8	8.8	+2.0	4.0	6.0	-2.0
TOTAL	100.0	100.0	2.8[1]	100.0	100.0	1.9[1]

Table E.2 continued

Manawatu Round 2						
	Contextual (N=56)			Non-Contextual (N=65)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	17.4	25.4	-8.0	20.8	29.2	-8.4
Labour	37.0	38.2	-1.2	22.9	26.2	-3.3
National	34.8	29.1	+5.7	47.9	40.0	+7.9
Other	10.8	7.3	+3.5	8.3	4.6	+3.7
TOTAL	100.0	100.0	4.6[1]	100.0	100.0	5.8[1]

Miramar Round 2						
	Contextual (N=45)			Non-Contextual (N=42)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	5.1	15.6	-10.5	7.4	12.2	-4.8
Labour	38.5	33.3	+5.2	66.7	61.0	+5.7
National	48.7	46.7	+2.0	22.2	24.4	-2.2
Other	7.7	4.4	+3.3	3.7	2.4	+1.3
TOTAL	100.0	100.0	5.3[1]	100.0	100.0	3.5[1]

Western Hutt - Round 2						
	Contextual (N=72)			Non-Contextual (N=66)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	10.8	9.0	+1.8	17.6	16.9	+0.7
Labour	36.9	37.3	-0.4	31.4	37.3	-5.9
National	46.2	41.8	+4.4	41.2	38.9	+2.3
Other	6.1	11.9	-5.8	9.8	6.8	+3.0
TOTAL	100.0	100.0	3.1[1]	100.0	100.0	3.0[1]

1. Mean absolute error.
2. The only result to show evidence of a possible bandwagon effect.

APPENDIX F

CONFIDENTIALITY EFFECTS:

SUPPLEMENTARY TABLES

Table F.1: Effects of Increased Confidentiality on Voting Estimates - Electorate Results

Manawatu - Round 1 Interviews								
	Contextual (N = 101)				Non-Contextual (N = 104)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	11.9	13.6	27.5	-13.9	15.4	18.0	27.5	-9.5
Labour	35.6	40.4	33.4	+7.0	29.8	34.8	33.4	+1.4
National	33.7	38.2	32.6	+5.6	27.9	32.6	32.6	0.0
NZ First	5.9	6.7	4.6	+2.1	11.5	13.5	4.6	+8.9
Other	1.0	1.1	1.9	-0.8	1.0	1.1	1.9	-0.8
None	2.0	-	-	-	6.7	-	-	-
Undecided	5.9	-	-	-	6.7	-	-	-
Refused	4.0	-	-	-	0.0	-	-	-
TOTAL	100.0	100.0	100.0	5.9[1]	100.0	100.0	100.0	4.1[1]

Miramar - Round 1 Interviews								
	Contextual (N = 96)				Non-Contextual (N = 97)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	13.5	14.9	11.6	+3.3	12.4	13.3	11.6	+1.7
Labour	39.6	43.8	46.4	-2.6	37.1	40.0	46.4	-6.4
National	31.3	34.5	34.8	-0.3	32.0	34.4	34.8	-0.4
NZ First	5.2	5.7	4.3	+1.4	9.3	10.0	4.3	+5.7
Other	1.0	1.1	2.9	-1.8	2.1	2.3	2.9	-0.6
None	7.3	-	-	-	3.1	-	-	-
Undecided	1.0	-	-	-	2.1	-	-	-
Refused	1.0	-	-	-	2.1	-	-	-
TOTAL	100.0	100.0	100.0	1.9[1]	100.0	100.0	100.0	3.0[1]

Table F.1: continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 131)				Non-Contextual (N = 128)			
Party	**Raw Estimates**	**Re-Allocated**	**Actual Results**	**Error**	**Raw Estimates**	**Re-Allocated**	**Actual Results**	**Error**
Alliance	13.0	13.3	13.6	-0.3	9.4	10.9	13.6	-2.7
Labour	44.3	45.3	37.9	+7.4	35.2	40.9	37.9	+3.0
National	34.4	35.2	41.6	-6.4	30.5	35.5	41.6	-6.1
NZ First	6.1	6.2	3.8	+2.4	8.6	10.0	3.8	+6.2
Other	0.0	0.0	3.1	-3.1	2.3	2.7	3.1	-0.4
None	0.8	-	-	-	4.7	-	-	-
Undecided	1.5	-	-	-	8.6	-	-	-
Refused	0.0	-	-	-	0.8	-	-	-
TOTAL	100.0	100.0	100.0	3.9[1]	100.0	100.0	100.0	3.7[1]

Manawatu - Round 2 Interviews								
	Contextual (N = 117)				Non-Contextual (N = 119)			
Party	**Raw Estimates**	**Re-Allocated**	**Actual Results**	**Error**	**Raw Estimates**	**Re-Allocated**	**Actual Results**	**Error**
Alliance	12.8	13.9	27.5	-13.6	15.1	16.5	27.5	-11.0
Labour	36.8	39.4	33.4	+6.0	36.1	33.6	33.4	+6.0
National	36.8	39.4	32.6	+6.8	33.6	36.7	32.6	+4.1
NZ First	6.8	7.3	4.6	+2.7	4.2	4.6	4.6	0.0
Other	0.0	0.0	1.9	-1.9	2.5	2.8	1.9	+0.9
None	3.4	-	-	-	3.4	-	-	-
Undecided	2.6	-	-	-	3.4	-	-	-
Refused	0.9	-	-	-	1.7	-	-	-
TOTAL	100.0	100.0	100.0	6.2[1]	100.0	100.0	100.0	4.2[1]

Table F1: continued

Miramar - Round 2 Interviews								
	Contextual (N = 110)				Non-Contextual (N = 110)			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	12.7	13.0	11.6	+1.4	11.8	12.5	11.6	+0.9
Labour	46.4	47.2	46.4	+0.8	34.5	36.5	46.4	-9.9
National	26.4	26.9	34.8	-7.9	35.5	37.5	34.8	+2.7
NZ First	10.9	11.1	4.3	+6.8	10.9	11.6	4.3	+7.3
Other	1.8	1.9	2.9	-1.0	1.8	1.9	2.9	-1.0
None	0.9	-	-	-	3.6	-	-	-
Undecided	0.9	-	-	-	0.0	-	-	-
Refused	0.0	-	-	-	1.8	-	-	-
TOTAL	100.0	100.0	100.0	3.6[1]	100.0	100.0	100.0	4.4[1]

Western Hutt - Round 2 Interviews								
	Contextual (N = 124)				Non-Contextual (N = 125			
Party	Raw Estimates	Re-Allocated	Actual Results	Error	Raw Estimates	Re-Allocated	Actual Results	Error
Alliance	8.9	9.0	13.6	-4.6	15.2	16.4	13.6	+2.8
Labour	39.5	40.1	37.9	+2.2	37.6	40.5	37.9	+2.6
National	37.9	38.5	41.6	-3.1	32.8	35.3	41.6	-6.3
NZ First	9.7	9.8	3.8	+6.0	6.4	6.9	3.8	+3.1
Other	2.4	2.6	3.1	-0.5	0.8	0.9	3.1	-2.2
None	0.8	-	-	-	3.2	-	-	-
Undecided	0.0	-	-	-	3.2	-	-	-
Refused	0.8	-	-	-	0.8	-	-	-
TOTAL	100.0	100.0	100.0	3.3[1]	100.0	100.0	100.0	3.4[1]

1. Mean absolute error

Table F2: Effects of Increased Confidentiality on Voting Intention Estimates - Electorate Results

Manawatu Round 1						
	Contextual (N=47)			Non-Contextual (N=40)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	14.3	13.6	+0.7	21.7	25.6	-3.9
Labour	38.1	34.1	+4.0	27.0	35.9	-8.9
National	40.5	45.5	-5.0	37.8	30.8	+7.0
Other	7.1	6.8	+0.3	13.5	7.7	+5.8
TOTAL	100.0	100.0	2.5	100.0	100.0	6.4

Miramar Round 1						
	Contextual (N=41)			Non-Contextual (N=45)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	13.2	14.3	-1.1	18.6	14.0	+4.6
Labour	47.4	51.4	-4.0	37.2	55.8	-18.6
National	28.9	28.6	+0.3	30.2	23.3	+6.9
Other	10.5	5.7	+4.8	14.0	6.9	+7.1
TOTAL	100.0	100.0	2.6	100.0	100.0	9.3

Western Hutt - Round 1						
	Contextual (N=67)			Non-Contextual (N=70)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	16.7	20.3	-3.6	9.8	11.0	-1.2
Labour	42.4	34.4	+8.0	47.5	45.3	+2.2
National	34.8	40.6	-5.8	32.9	35.9	-3.0
Other	6.1	4.7	+1.4	9.8	7.8	+2.0
TOTAL	100.0	100.0	4.7	100.0	100.0	2.1

Table F.2 continued

Manawatu Round 2						
	Contextual (N=56)			Non-Contextual (N=58)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	13.2	28.3	-15.1	16.4	25.9	-9.5
Labour	34.0	26.4	+7.6	38.2	37.0	+1.2
National	43.4	41.5	+1.9	38.2	31.5	+6.7
Other	9.4	3.8	+5.6	7.2	5.6	+1.6
TOTAL	100.0	100.0	7.6	100.0	100.0	4.8

Miramar Round 2						
	Contextual (N=49)			Non-Contextual (N=45)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	12.2	8.5	+3.7	11.9	10.0	+1.9
Labour	47.0	59.6	-12.6	45.3	52.5	-7.2
National	30.6	25.5	+5.1	33.3	30.0	+3.3
Other	10.2	6.4	+3.8	9.5	7.5	+2.0
TOTAL	100.0	100.0	6.3	100.0	100.0	3.6

Western Hutt - Round 2						
	Contextual (N=65)			Non-Contextual (N=75)		
Party	**Predicted**	**Self-reported**	**Error**	**Predicted**	**Self-reported**	**Error**
Alliance	12.3	6.6	+5.7	11.6	18.6	-7.0
Labour	33.8	34.4	-0.6	40.6	42.9	-2.3
National	40.0	52.3	-12.4	37.7	32.9	+4.8
Other	13.9	6.6	+7.3	10.1	5.6	+4.5
TOTAL	100.0	100.0	6.5	100.0	100.0	4.7

APPENDIX G

PROBABILITY MEASURES:

SUPPLEMENTARY TABLES

Table G.1: Effect of Probability Measures on Voting Intention Estimates : Electorate Level

Manawatu Round 1						
	Contextual (N=105)			Non-Contextual (N=97)		
Party	**Raw Estimates**	**Actual Results**	**Error**	**Raw Estimates**	**Actual Results**	**Error**
Alliance	17.2	27.5	-10.3	24.2	27.5	-3.3
Labour	39.3	33.4	+5.9	29.0	33.4	-4.4
National	26.6	32.6	-6.0	23.0	32.6	-9.6
NZ First	11.9	4.6	+7.3	19.1	4.6	+14.5
Other	5.0	1.9	+3.1	4.7	1.9	+2.8
TOTAL	100.0	100.0	6.5[1]	100.0	100.0	6.9[1]

Miramar - Round 1						
	Contextual (N=98)			Non-Contextual (N=96)		
Party	**Raw Estimates**	**Actual Results**	**Error**	**Raw Estimates**	**Actual Results**	**Error**
Alliance	15.2	11.6	+3.6	14.1	11.6	+2.5
Labour	40.4	46.4	-6.0	36.6	46.4	-9.8
National	24.3	34.8	-10.5	29.9	34.8	-4.9
NZ First	13.0	4.3	+8.7	11.8	4.3	+7.5
Other	7.1	2.9	+4.2	7.6	2.9	+4.7
TOTAL	100.0	100.0	6.6[1]	100.0	100.0	5.9[1]

Table G.1 continued

Western Hutt - Round 1						
	Contextual (N=126)			Non-Contextual (N=131)		
Party	**Raw Estimates**	**Actual Results**	**Error**	**Raw Estimates**	**Actual Results**	**Error**
Alliance	14.4	13.6	+0.8	15.2	13.6	+1.6
Labour	35.3	37.9	-2.6	36.5	37.9	-1.4
National	31.8	41.6	-9.8	32.1	41.6	-9.5
NZ First	14.8	3.8	+11.0	11.9	3.8	+8.1
Other	3.7	3.1	+0.6	4.3	3.1	+1.2
TOTAL	100.0	100.0	5.0[1]	100.0	100.0	4.4[1]

Manawatu Round 2						
	Contextual (N=124)			Non-Contextual (N=116)		
Party	**Raw Estimates**	**Actual Results**	**Error**	**Raw Estimates**	**Actual Results**	**Error**
Alliance	20.4	27.5	-7.1	17.9	27.5	-9.6
Labour	35.8	33.4	+2.4	35.1	33.4	+1.7
National	26.9	32.6	-5.7	31.8	32.6	-0.8
NZ First	13.5	4.6	+8.9	10.7	4.6	+6.1
Other	3.4	1.9	+1.5	4.5	1.9	+2.6
TOTAL	100.0	100.0	5.1[1]	100.0	100.0	4.2[1]

Table G.1 continued

Miramar - Round 2						
	Contextual (N=112)			Non-Contextual (N=112)		
Party	**Raw Estimates**	**Actual Results**	**Error**	**Raw Estimates**	**Actual Results**	**Error**
Alliance	15.7	11.6	+4.1	13.8	11.6	+2.2
Labour	33.6	46.4	-12.8	42.6	46.4	-3.8
National	30.3	34.8	-4.5	24.9	34.8	-9.9
NZ First	14.6	4.3	+10.3	11.9	4.3	+7.6
Other	5.8	2.9	+2.9	6.8	2.9	+3.9
TOTAL	100.0	100.0	6.9[1]	100.0	100.0	5.5[1]

Western Hutt - Round 2						
	Contextual (N=125)			Non-Contextual (N=126)		
Party	**Raw Estimates**	**Actual Results**	**Error**	**Raw Estimates**	**Actual Results**	**Error**
Alliance	16.5	13.6	+2.9	15.3	13.6	+1.7
Labour	32.5	37.9	-5.4	35.1	37.9	-2.8
National	35.3	41.6	-6.3	32.9	41.6	-8.7
NZ First	10.9	3.8	+7.1	11.8	3.8	+8.0
Other	4.8	3.1	+1.7	4.9	3.1	+1.8
TOTAL	100.0	100.0	4.7[1]	100.0	100.0	4.6[1]

1. Mean absolute errors

Table G.2: Effect of Probability Measures on Voting Intention Estimates : Predicted by Self Reported Behaviour

Manawatu - Round 1						
	Contextual (N=47)			Non-Contextual (N=46)		
Party	**Predicted**	**Self-Reported**	**Error**	**Predicted**	**Self-Reported**	**Error**
Alliance	19.6	24.4	-4.8	24.4	31.0	-6.6
Labour	37.8	40.0	-2.2	26.4	26.2	+0.2
National	28.0	28.9	-0.9	22.8	28.6	-5.8
NZ First	11.1	0.0	+11.1	23.1	11.9	+11.2
Other	3.5	6.7	-3.2	3.3	2.3	+1.0
TOTAL	100.0	100.0	4.4[1]	100.0	100.0	5.0[1]

Miramar - Round 1						
	Contextual (N=39)			Non-Contextual (N=34)		
Party	**Predicted**	**Self-Reported**	**Error**	**Predicted**	**Self-Reported**	**Error**
Alliance	13.1	16.7	-3.6	12.9	17.2	-4.3
Labour	41.3	52.8	-11.5	36.3	48.3	-12.0
National	25.8	25.0	+0.8	34.5	34.5	0.0
NZ First	13.8	5.5	+8.3	10.9	0.0	+10.9
Other	6.0	0.0	+6.0	5.4	0.0	+5.4
TOTAL	100.0	100.0	6.0[1]	100.0	100.0	6.5[1]

Table G.2 continued

Western Hutt - Round 1						
	Contextual (N=75)			Non-Contextual (N=71)		
Party	**Predicted**	**Self-Reported**	**Error**	**Predicted**	**Self-Reported**	**Error**
Alliance	16.6	11.9	+4.7	15.8	12.5	+3.3
Labour	33.3	40.3	-7.0	37.1	42.2	-5.1
National	31.7	44.8	-13.1	32.7	37.5	-4.8
NZ First	14.8	1.5	+13.3	11.0	4.7	+6.3
Other	3.6	1.5	+2.1	3.4	3.1	+0.3
TOTAL	100.0	100.0	8.0[1]	100.0	100.0	4.0[1]

Manawatu - Round 2						
	Contextual (N=64)			Non-Contextual (N=62)		
Party	**Predicted**	**Self-Reported**	**Error**	**Predicted**	**Self-Reported**	**Error**
Alliance	19.6	30.2	-10.6	19.8	25.4	-5.6
Labour	31.3	30.2	+1.1	31.7	28.8	+2.9
National	31.9	31.7	+0.2	35.0	42.4	-7.4
NZ First	13.0	6.3	+6.7	10.3	1.7	+8.6
Other	4.2	1.6	+2.6	3.2	1.7	+1.5
TOTAL	100.0	100.0	4.2[1]	100.0	100.0	5.2[1]

Table G.2 continued

Miramar - Round 2						
	Contextual (N=49)			Non-Contextual (N=46)		
Party	**Predicted**	**Self-Reported**	**Error**	**Predicted**	**Self-Reported**	**Error**
Alliance	17.3	14.9	+2.4	14.7	16.7	-2.0
Labour	32.2	46.8	-14.6	42.6	57.1	-14.5
National	31.0	36.2	-5.2	22.8	26.2	-3.4
NZ First	12.8	2.1	+10.7	12.2	0.0	+12.2
Other	6.7	0.0	+6.7	7.7	0.0	+7.7
TOTAL	100.0	100.0	7.9[1]	100.0	100.0	8.0[1]

Western Hutt - Round 2						
	Contextual (N=68)			Non-Contextual (N=74)		
Party	**Predicted**	**Self-Reported**	**Error**	**Predicted**	**Self-Reported**	**Error**
Alliance	14.5	12.1	+2.4	15.6	15.5	+0.1
Labour	29.5	33.4	-3.9	35.7	32.4	+3.3
National	41.2	50.0	-8.8	33.0	46.5	-13.5
NZ First	11.3	4.5	+6.8	11.4	2.8	+8.6
Other	3.5	0.0	+3.5	4.3	2.8	+1.5
TOTAL	100.0	100.0	5.1[1]	100.0	100.0	5.4[1]

1. Mean absolute errors

APPENDIX H

WEIGHTING PROCEDURES

WEIGHTING PROCEDURES AND CALCULATIONS

Demographic Weighting

As the estimates reported in political opinion polls have usually been weighted to ensure the sample corresponds to the population's age and gender sub-groups, these data were also weighted to ensure no demographic sub-sample was over-or under-represented. Details of the demographic composition of each electorate were obtained from the 1991 Census. and are contained in the table below.

Table H.1 Age and Gender Distributions in the Manawatu, Miramar and Western Hutt Electorates

	Age Group							
	< 5	5-14	15-19	20-29	30-39	40-49	50-59	60+
Manawatu (N=34,302)								
Male	4.3	7.1	5.9	9.9	7.5	5.7	3.7	5.4
Female	4.1	6.8	5.7	9.5	7.6	5.8	3.8	7.2
Miramar (N=36,060)								
Male	3.5	5.7	3.4	10.9	8.5	6.2	4.6	6.6
Female	3.1	5.2	3.6	11.2	8.5	6.0	4.3	8.7
Western Hutt (N=34,155)								
Male	4.3	7.0	4.2	8.6	8.9	7.4	4.3	5.1
Female	4.0	6.9	4.1	8.9	8.8	7.0	4.0	6.5

The weights used for each questionnaire version in each electorate are detailed in Table N.2, below.

Table H.2 Demographic Weighting Factors

Electorate		Age Group 18-29	30-39	40-49	50-59	60+
Manawatu - Round 1						
Version 1	Male	1.60	0.83	0.64	2.60	1.14
	Female	1.13	0.91	0.93	0.92	0.86
Version 2	Male	1.54	0.81	1.08	0.62	1.18
	Female	0.93	0.96	1.26	1.15	0.83
Version 3	Male	2.19	0.56	0.68	1.30	0.71
	Female	1.19	1.08	0.64	2.70	1.28
Version 4	Male	2.25	0.58	1.05	1.08	0.73
	Female	1.15	0.58	4.32	0.93	1.05
Version 5	Male	1.20	0.65	0.65	1.40	1.60
	Female	1.50	1.10	0.95	0.57	1.10
Version 6	Male	1.30	0.94	1.10	2.50	0.50
	Female	1.20	0.74	0.66	1.70	1.60
Miramar - Round 1						
Version 1	Male	1.64	0.96	0.91	0.68	0.74
	Female	0.93	1.40	1.26	0.79	0.97
Version 2	Male	1.03	0.76	0.63	1.31	1.05
	Female	1.30	2.02	0.96	1.53	0.68
Version 3	Male	0.94	0.94	0.84	1.40	0.50
	Female	1.41	1.49	1.35	2.62	1.06
Version 4	Male	1.69	0.70	0.77	1.44	0.58
	Female	1.04	1.06	2.48	5.50	0.67
Version 5	Male	1.50	0.70	0.60	0.97	1.20
	Female	1.76	0.97	1.08	0.67	0.90
Version 6	Male	1.50	1.15	1.50	1.90	1.30
	Female	1.10	0.62	0.74	0.66	0.75
Western Hutt - Round 1						
Version 1	Male	1.60	0.88	1.14	1.70	0.87
	Female	0.90	1.10	1.21	0.93	1.01
Version 2	Male	2.12	0.97	0.80	1.15	0.86
	Female	1.51	0.95	0.82	1.24	0.63
Version 3	Male	1.52	1.16	0.75	0.88	0.66
	Female	1.48	0.80	1.43	0.66	1.18
Version 4	Male	1.41	0.69	0.73	1.97	1.82
	Female	1.11	1.13	0.70	1.02	1.30
Version 5	Male	1.40	1.20	0.93	0.86	1.13
	Female	1.70	0.90	0.83	0.71	0.72
Version 6	Male	1.03	1.24	0.89	0.66	0.91
	Female	1.90	0.79	1.41	0.80	0.90

Table H.2 Continued

Electorate		Age Group 18-29	30-39	40-49	50-59	60+
Manawatu - Round 2						
Version 1	Male	1.59	0.80	0.74	0.83	1.64
	Female	1.18	1.37	0.80	0.86	0.65
Version 2	Male	2.19	1.50	1.69	0.66	0.57
	Female	1.61	0.79	0.74	0.96	1.06
Version 3	Male	1.56	0.62	1.05	0.68	0.56
	Female	1.21	0.89	1.61	2.08	1.07
Version 4	Male	1.47	0.84	1.21	0.88	0.76
	Female	1.41	0.71	1.22	0.81	0.80
Version 5	Male	1.80	0.66	0.77	0.80	0.87
	Female	1.14	1.02	1.12	0.96	1.04
Version 6	Male	1.70	1.03	0.62	1.53	0.69
	Female	1.60	0.88	1.58	0.48	0.83
Miramar - Round 2						
Version 1	Male	0.97	0.67	1.25	0.94	0.93
	Female	2.00	0.76	1.07	1.02	1.02
Version 2	Male	1.15	0.74	0.79	0.54	1.31
	Female	2.22	1.00	0.71	1.00	1.10
Version 3	Male	1.57	1.19	1.23	0.81	0.66
	Female	0.99	1.20	0.71	1.22	0.76
Version 4	Male	1.57	0.92	1.23	0.72	0.77
	Female	1.12	0.70	0.94	1.53	1.01
Version 5	Male	1.60	0.86	0.89	0.83	0.72
	Female	1.80	1.10	1.08	0.77	0.68
Version 6	Male	1.26	1.21	1.11	1.64	0.55
	Female	1.07	0.94	0.87	0.62	1.38
Western Hutt - Round 2						
Version 1	Male	1.19	1.29	0.85	1.27	0.73
	Female	0.97	1.17	0.94	1.40	0.81
Version 2	Male	1.13	1.47	0.74	0.72	0.70
	Female	1.59	0.88	1.06	1.81	0.83
Version 3	Male	1.27	0.91	0.91	0.84	0.88
	Female	1.66	0.90	1.02	0.77	1.02
Version 4	Male	1.37	0.97	1.41	1.09	0.80
	Female	1.66	0.85	0.59	1.00	1.02
Version 5	Male	1.64	0.71	0.99	1.91	0.68
	Female	1.32	0.77	1.03	1.40	1.03
Version 6	Male	0.82	1.21	1.30	1.09	0.75
	Female	1.56	0.86	1.14	1.17	0.76

Weighting by Probability of Voting

Since all respondents indicated their probability of voting, these estimates were used to weight their voting intention. Thus respondents' weighted voting intention was calculated as follows:

Party A support = ((Voting intention * voting probability) /10[1])

[1] Or a multiple of 10.

Adjusting the data in this way meant that the voting intention estimates now summed to the overall probability of voting (for either the whole sample or the individual electorates), thus the estimates were re-calculated to form percentages. Tables H.n to H.n contain both the original weighted intentions and the re-calculated percentages. The cell counts in these tables are lower than the original data because respondents who stated there was a 0 probability that they would cast their vote were dropped during the weighting process.

Weighting by Past Voting Behaviour

The final weighting procedure attempted to balance the samples so they reflected the party support distributions resulting from the 1990 General Election. The 1990 results were as follows:

Table H.3: 1990 General Election Results

Party	Manawatu	Miramar	Western Hutt
	%	%	%
Alliance[1]	11.90[2]	14.59	10.94
Labour	35.30	40.76	42.06
National	50.97	43.49	45.54
Other	1.83	1.16	1.47

1. Although it was not possible for voters to have cast a vote for an Alliance party in 1990, they could have voted for one of the constituent parties that later joined together under the Alliance rubric. Consequently, the Alliance percentages reflect the proportion of voters who supported the New Labour, Green or Democrat parties in the 1990 General Election.

2. Because the candidate who contested the Manawatu electorate as an Alliance candidate was originally elected as a National Party MP, the Manawatu results were adjusted slightly to reflect the personal following that the MP had developed. This adjustment increased the Alliance share of vote to 16.90% and decreased the National vote to 45.97%. While this is clearly an arbitrary re-allocation of support, it does attempt to recognise and address the more complex electoral history associated with the Manawatu electorate.

Since respondents could not have voted for the NZ First Party in any antecedent form during the 1990 General election, this party was allocated a weight of one. Likewise, if respondents stated that they did not vote during the 1990 General Election, they were also allocated a weight of one to ensure that those people too young to vote were not excluded from the weighted results. Those people who refused to state the party for which they vote in 1990, or those who declared they could not remember which party they had voted for were not allocated weights and so were dropped from the final estimates. The sample sizes shown in the Tables H.5 to H.13 vary from the original data because of the weighting factors used. The weights associated with each electorate are detailed in Table H.4 below.

Table H.4 1990 Voting Behaviour Weighting Factors

Electorate	Party			
	Alliance	Labour	National	Other
Manawatu - Round 1				
Version 1	2.61	0.81	0.94	0.38
Version 2	2.03	0.87	0.92	0.44
Version 3	2.46	0.80	0.98	0.36
Version 4	1.75	0.88	0.96	0.36
Version 5	2.09	1.03	1.45	0.96
Version 6	3.84	1.18	1.41	0.86
Miramar - Round 1				
Version 1	2.47	0.84	1.01	0.48
Version 2	1.04	0.77	1.37	0.92
Version 3	2.03	0.91	0.94	0.73
Version 4	1.05	0.95	1.08	0.41
Version 5	3.60	1.00	1.40	0.40
Version 6	2.30	1.20	1.30	1.10
Western Hutt - Round 1				
Version 1	1.56	1.04	0.87	1.00
Version 2	1.88	1.07	0.91	0.31
Version 3	1.45	0.97	0.95	1.50
Version 4	1.58	0.78	1.28	0.43
Version 5	4.50	1.30	1.10	0.90
Version 6	2.90	1.30	1.00	0.60

Table H.4 Continued

	Party			
Electorate	Alliance	Labour	National	Other
Manawatu - Round 2				
Version 1	5.48	0.86	0.80	0.45
Version 2	3.84	0.80	0.85	0.95
Version 3	19.91	0.83	0.77	0.56
Version 4	2.92	0.82	0.85	1.64
Version 5	3.00	0.96	1.20	0.40
Version 6	4.60	1.00	1.20	1.00
Miramar - Round 2				
Version 1	3.32	0.98	0.84	0.50
Version 2	6.63	0.76	1.21	0.14
Version 3	2.61	0.89	0.94	0.50
Version 4	1.02	0.93	1.18	0.23
Version 5	1.80	1.20	1.10	1.20
Version 6	1.50	1.10	1.40	0.60
Western Hutt - Round 2				
Version 1	2.18	1.00	0.88	1.50
Version 2	2.79	0.82	1.12	0.39
Version 3	1.54	1.13	0.88	0.38
Version 4	1.20	0.91	1.15	0.30
Version 5	2.70	1.20	1.00	0.94
Version 6	4.50	1.13	1.19	1.88

Table H.5: Contextual and Non-Contextual Results Weighted by Demographic Variables

Manawatu - Round 1 Interviews								
	Contextual (N = 104)				Contextual (N = 107)			
Party	Initial Estimates	Re- Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	11.1	15.8	27.5	-11.7	8.2	15.0	27.5	-12.5
Labour	35.7	48.7	33.4	+15.3	19.7	35.0	33.4	+1.6
National	21.0	28.9	32.6	-3.7	24.0	43.3	32.6	+10.7
NZ First	4.1	5.3	4.6	+0.7	3.1	5.0	4.6	+0.4
Other	0.8	1.3	1.9	-0.6	0.8	1.7	1.9	-0.2
None	5.1	-	-	-	9.0	-	-	-
Undecided	22.2	-	-	-	32.3	-	-	-
Refused	0.0	-	-	-	3.0	-	-	-
TOTAL	100.0	100.0	100.0	6.4[1]	100.0	100.0	100.0	5.1[1]

Miramar Round 1 Interviews								
	Contextual (N = 115)				Contextual (N = 112)			
Party	Initial Estimates	Re- Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	2.9	3.6	11.6	-8.0	6.6	9.9	11.6	-1.7
Labour	35.0	48.2	46.4	+1.8	30.7	47.9	46.4	+1.5
National	27.9	38.6	34.8	+3.8	20.5	32.4	34.8	-2.4
NZ First	5.9	8.4	4.3	+4.1	6.3	9.8	4.3	+5.5
Other	1.1	1.2	2.9	-1.7	0.0	0.0	2.9	-2.9
None	11.5	-	-		8.9	-	-	-
Undecided	14.4	-	-		22.2	-	-	-
Refused	1.3	-	-		5.0	-	-	-
TOTAL	100.0	100.0	100.0	3.9[1]	100.0	100.0	100.0	2.8[1]

Table H.5 continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 152)				Non-Contextual (N = 135)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	6.4	8.1	13.6	-5.5	6.7	10.6	13.6	-3.0
Labour	31.8	38.7	37.9	+0.8	26.1	41.2	37.9	+3.3
National	32.6	40.3	41.6	-1.3	28.8	45.9	41.6	+4.3
NZ First	8.0	9.7	3.8	+5.9	1.8	2.3	3.8	-1.5
Other	2.7	3.2	3.1	+0.1	0.0	0.0	3.1	-3.1
None	6.0	-	-		10.8	-	-	-
Undecided	12.4	-	-		24.4	-	-	-
Refused	0.0	-	-		1.5	-	-	-
TOTAL	100.0	100.0	100.0	2.7[1]	100.0	100.0	100.0	3.0[1]

Manawatu - Round 2 Interviews								
	Contextual (N = 128)				Non-Contextual (N = 142)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	10.9	13.3	27.5	-14.2	13.1	19.0	27.5	-8.5
Labour	32.5	40.0	33.4	+6.6	21.1	30.0	33.4	-3.4
National	31.7	39.0	32.6	+6.4	31.0	44.0	32.6	+11.4
NZ First	5.8	6.7	4.6	+2.1	4.8	7.0	4.6	+2.4
Other	0.6	1.0	1.9	-0.9	0.0	0.0	1.9	-1.9
None	11.1	-	-	-	4.7	-	-	-
Undecided	7.4	-	-	-	24.2	-	-	-
Refused	0.0	-	-	-	1.1	-	-	-
TOTAL	100.0	100.0	100.0	6.0[1]	100.0	100.0	100.0	5.5[1]

Table H.5 continued

Miramar - Round 2 Interviews								
	Contextual (N = 111)				Contextual (N = 110)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	9.2	11.2	11.6	-0.4	6.5	9.6	11.6	-2.0
Labour	25.8	31.5	46.4	-14.9	36.5	54.8	46.4	+8.4
National	39.3	48.3	34.8	+13.5	18.1	27.4	34.8	-7.4
NZ First	5.2	6.7	4.3	+2.4	5.8	8.2	4.3	+3.9
Other	1.7	2.3	2.9	-0.6	0.0	0.0	2.9	-2.9
None	5.0	-	-		5.1	-	-	-
Undecided	12.9	-	-		25.5	-	-	-
Refused	0.9	-	-		2.6	-	-	-
TOTAL	100.0	100.0	100.0	6.4[1]	100.0	100.0	100.0	4.9[1]

Western Hutt Round 2 Interviews								
	Contextual (N = 127)				Contextual (N = 131)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	9.1	11.4	13.6	-2.2	11.8	14.9	13.6	+1.3
Labour	26.6	32.4	37.9	-5.5	26.6	34.7	37.9	-3.2
National	39.7	47.6	41.6	+6.0	31.7	41.6	41.6	0.0
NZ First	5.7	6.7	3.8	+2.9	5.9	7.8	3.8	+4.0
Other	1.8	1.9	3.1	-1.2	0.6	1.0	3.1	-2.1
None	3.3	-	-	-	6.7	-	-	-
Undecided	12.1	-	-	-	15.6	-	-	-
Refused	1.8	-	-	-	1.2	-	-	-
TOTAL	100.0	100.0	100.0	3.6[1]	100.0	100.0	100.0	2.1[2]

1. Mean absolute error

Table H.6: Contextual and Non-Contextual Results Weighted by Voting Probability

Manawatu - Round 1 Interviews								
	Non-Contextual (N = 75)[1]				Non-Contextual (N = 58)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	16	17.6	27.5	-9.9	13	14.6	27.5	-12.9
Labour	40	44.0	33.4	+10.6	32	36.0	33.4	+2.6
National	29	31.9	32.6	-0.7	38	42.7	32.6	+10.1
NZ First	5	5.4	4.6	+0.8	4	4.5	4.6	-0.1
Other	1	1.1	1.9	-0.8	2	2.2	1.9	+0.3
None		-	-			-	-	-
Undecided		-	-			-	-	-
Refused		-	-			-	-	-
TOTAL	91	100.0	100.0	4.6[2]	89	100.0	100.0	5.2[2]

Miramar - Round 1 Interviews								
	Contextual (N = 85)				Non-Contextual (N = 73)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	5	5.7	11.6	-5.9	12	13.6	11.6	+2.0
Labour	44	50.0	46.4	+3.6	38	43.2	46.4	-3.2
National	32	36.4	34.8	+1.6	30	34.1	34.8	-0.7
NZ First	6	6.8	4.3	+2.5	8	9.1	4.3	+4.8
Other	1	1.1	2.9	-1.8	0	0.0	2.9	-2.9
None		-	-	-			-	-
Undecided		-	-	-			-	-
Refused		-	-	-			-	-
TOTAL	88	100.0	100.0	3.1[2]	88	100.0	100.0	2.7[2]

Table H.6 continued

Western Hutt - Round 1 Interviews								
	Non-Contextual (N = 109)				Non-Contextual (N = 88)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	6	6.5	13.6	-7.1	12	12.6	13.6	-1.0
Labour	36	39.1	37.9	+1.2	38	40.0	37.9	+2.1
National	38	41.3	41.6	-0.3	42	44.2	41.6	+2.6
NZ First	9	9.8	3.8	+6.0	3	3.2	3.8	-0.6
Other	3	3.3	3.1	+0.2	0	0.0	3.1	-3.1
None		-	-			-	-	-
Undecided		-	-			-	-	-
Refused		-	-			-	-	-
TOTAL	92	100.0	100.0	3.0[2]	95	100.0	100.0	1.9[2]

Manawatu - Round 2 Interviews								
	Contextual (N = 103)				Non-Contextual (N = 90)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	12	12.9	27.5	-14.6	21	22.6	27.5	-4.9
Labour	40	43.0	33.4	+9.6	29	31.2	33.4	-2.2
National	34	36.6	32.6	+4.0	37	39.8	32.6	+7.2
NZ First	6	6.4	4.6	+1.6	6	6.4	4.6	+1.8
Other	1	1.1	1.9	-0.8	0	0.0	1.9	-1.9
None		-	-	-			-	-
Undecided		-	-	-			-	-
Refused		-	-	-			-	-
TOTAL	93	100.0	100.0	6.1[2]	93	100.0	100.0	3.6[2]

Table H.6 continued

Miramar - Round 2 Interviews								
Non-Contextual (N = 90)					Non-Contextual (N = 74)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	11	12.1	11.6	+0.5	12	13.0	11.6	+1.4
Labour	30	33.0	46.4	-13.4	47	51.1	46.4	+4.7
National	42	46.1	34.8	+11.3	26	28.3	34.8	-6.5
NZ First	7	7.7	4.3	+3.4	7	7.6	4.3	+3.3
Other	1	1.1	2.9	-1.8	0	0.0	2.9	-2.9
None		-	-			-	-	-
Undecided		-	-			-	-	-
Refused		-	-			-	-	-
TOTAL	91	100.0	100.0	6.1[2]	92	100.0	100.0	3.8[2]

Western Hutt - Round 2 Interviews								
Contextual (N = 104)					Non-Contextual (N = 99)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	10	10.8	13.6	-2.8	14	14.9	13.6	+1.3
Labour	29	31.6	37.9	-6.3	33	35.1	37.9	-2.8
National	44	47.8	41.6	+6.2	39	41.5	41.6	-0.1
NZ First	7	7.6	3.8	+3.8	7	7.4	3.8	+3.6
Other	2	2.2	3.1	-0.9	1	1.1	3.1	-2.0
None		-	-	-			-	-
Undecided		-	-	-			-	-
Refused		-	-	-			-	-
TOTAL	92	100.0	100.0	4.0[2]	94	100.0	100.0	2.0[2]

1. The smaller sub-sample sizes result from the exclusion of all declared non-voters and undecided respondents from the calculations
2. Mean absolute errors

Table H.7: Contextual and Non-Contextual Results Weighted by 1990 Voting Patterns

Manawatu - Round 1 Interviews								
	Contextual (N = 102)				Contextual (N = 107)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	14.1	18.9	27.5	-8.6	8.8	17.6	27.5	-9.9
Labour	29.7	39.2	33.4	+5.8	18.7	35.3	33.4	+1.9
National	27.3	35.1	32.6	+2.5	21.9	41.2	32.6	+8.6
NZ First	4.6	5.4	4.6	+0.8	3.1	5.9	4.6	+1.3
Other	0.8	1.4	1.9	-0.5	0.5	0.0	1.9	-1.9
None	5.7	-	-		8.0	-	-	-
Undecided	17.7	-	-		38.1	-	-	-
Refused	0.0	-	-		0.9	-	-	-
TOTAL	100.0	100.0	100.0	3.6[1]	100.0	100.0	100.0	4.7[1]

Miramar - Round 1 Interviews								
	Contextual (N = 107)				Contextual (N = 105)			
Party	Initial Estimates	Re-allocated	Actual Results	Error	Initial Estimates	Re-allocated	Actual Results	Error
Alliance	3.8	4.8	11.6	-6.8	8.3	12.5	11.6	+0.9
Labour	36.4	47.0	46.4	+0.6	26.9	38.9	46.4	-7.5
National	30.2	38.6	34.8	+3.8	27.8	40.3	34.8	+5.5
NZ First	6.5	8.4	4.3	+4.1	5.7	8.3	4.3	+4.0
Other	0.8	1.2	2.9	-1.7	0.0	0.0	2.9	-2.9
None	9.9	-	-	-	8.1		-	-
Undecided	11.5	-	-	-	22.4		-	-
Refused	0.9	-	-	-	0.7		-	-
TOTAL	92	100.0	100.0	3.4[1]	100.0	100.0	100.0	4.2[1]

Table H.7 continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 130)				Contextual (N = 125)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-allocated	Actual Results	Error
Alliance	6.2	7.6	13.6	-6.0	10.4	14.9	13.6	+1.3
Labour	34.3	41.5	37.9	+3.6	28.0	40.2	37.9	+2.3
National	31.1	37.7	41.6	-3.9	28.5	41.4	41.6	-0.2
NZ First	8.5	10.4	3.8	+6.6	2.4	3.5	3.8	-0.3
Other	2.1	2.8	3.1	-0.3	0.0	0.0	3.1	-3.1
None	4.8	-	-		11.1	-	-	-
Undecided	13.1	-	-		19.5	-	-	-
Refused	-	-	-		0.0	-	-	-
TOTAL	100.0	100.0	100.0	3.5[1]	100.0	100.0	100.0	1.4[1]

Manawatu - Round 2 Interviews								
	Contextual (N = 128)				Contextual (N = 123)			
Party	Initial Estimates	Re-allocated	Actual Results	Error	Initial Estimates	Re-allocated	Actual Results	Error
Alliance	16.2	20.0	27.5	-7.5	18.5	24.5	27.5	-3.0
Labour	36.4	43.8	33.4	+10.4	20.7	26.6	33.4	-6.8
National	24.1	29.5	32.6	-3.1	32.6	42.5	32.6	+9.9
NZ First	5.5	6.7	4.6	+2.1	4.9	6.4	4.6	+1.8
Other	0.4	0.0	1.9	-1.9	0.0	0.0	1.9	-1.9
None	9.8	-	-	-	3.5	-	-	-
Undecided	7.6	-	-	-	19.7	-	-	-
Refused	0.0	-	-	-	0.0	-	-	-
TOTAL	100.0	100.0	100.0	5.0[1]	100.0	100.0	100.0	4.7[1]

Table H.7 continued

Miramar - Round 2 Interviews								
	Contextual (N = 106)				Contextual (N = 122)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	11.6	14.0	11.6	+2.4	6.3	10.3	11.6	-1.3
Labour	26.5	32.6	46.4	-13.8	32.3	51.4	46.4	+5.0
National	36.6	44.2	34.8	+9.4	20.7	32.4	34.8	-2.4
NZ First	5.7	7.0	4.3	+2.7	3.9	5.9	4.3	+1.6
Other	1.8	2.2	2.9	-0.7	0.0	-	2.9	-2.9
None	7.8	-	-	-	10.6	-	-	-
Undecided	10.1	-	-	-	25.5	-	-	-
Refused	0.0	-	-	-	0.7	-	-	-
TOTAL	100.0	100.0	100.0	5.8[1]	100.0	100.0	100.0	2.6[1]

Western Hutt - Round 2 Interviews								
	Contextual (N = 122)				Non-Contextual (N = 122)			
Party	Initial Estimates	Re-allocated	Actual Results	Error	Initial Estimates	Re-allocated	Actual Results	Error
Alliance	11.0	12.6	13.6	-1.0	10.7	13.5	13.6	-0.1
Labour	28.2	33.0	37.9	-4.9	28.2	35.4	37.9	-2.5
National	37.6	44.7	41.6	+3.1	33.5	42.7	41.6	+1.1
NZ First	6.6	7.8	3.8	+4.0	6.0	7.3	3.8	+3.5
Other	1.6	1.9	3.1	-1.2	0.7	1.1	3.1	-2.0
None	4.2	-	-	-	6.3	-	-	-
Undecided	10.8	-	-	-	14.0	-	-	-
Refused	0.0	-	-	-	0.7	-	-	-
TOTAL	100.0	100.0	100.0	2.2[1]	100.0	100.0	100.0	1.8[1]

1. Mean absolute errors

Table H.8: Secret Ballot Results weighted by Demographic Variables

Manawatu - Round 1 Interviews								
	Contextual (N = 101)				Non-Contextual (N = 104)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	11.1	12.3	27.5	-15.2	16.1	18.7	27.5	-8.8
Labour	36.7	41.6	33.4	+8.2	26.1	29.7	33.4	-3.7
National	32.6	37.1	32.6	+4.5	31.3	36.3	32.6	+3.7
NZ First	7.0	7.9	4.6	+3.3	12.7	14.2	4.6	+9.6
Other	0.6	1.1	1.9	-0.8	0.6	1.1	1.9	-0.8
None	2.7	-	-	-	6.5	-	-	
Undecided	5.6	-	-	-	6.8	-	-	
Refused	3.6	-	-	-	0.0	-	-	
TOTAL	100.0	100.0	100.0	6.4[1]	100.0	100.0	100.0	5.3[1]

Miramar - Round 1 Interviews								
	Non-Contextual (N = 102)				Non-Contextual (N = 97)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	14.7	16.7	11.6	+5.1	17.0	18.5	11.6	+6.9
Labour	37.2	42.2	46.4	-4.2	36.6	39.1	46.4	-7.3
National	29.8	33.3	34.8	-1.5	30.0	31.5	34.8	-3.3
NZ First	6.0	6.7	4.3	+2.4	8.3	8.7	4.3	+4.4
Other	0.9	1.1	2.9	-0.8	1.7	2.2	2.9	-0.7
None	8.9	-	-	-	2.5	-	-	
Undecided	1.0	-	-	-	2.5	-	-	
Refused	1.4	-	-	-	1.5	-	-	
TOTAL	100.0	100.0	100.0	3.0[1]	100.0	100.0	100.0	4.5[1]

Table H.8 continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 132)				Non-Contextual (N = 130)			
Party	Initial Estimates	Re-allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	12.2	12.3	13.6	-1.3	9.9	11.6	13.6	-2.0
Labour	46.8	47.7	37.9	+9.8	38.6	44.6	37.9	+6.7
National	34.1	34.6	41.6	-7.0	26.9	31.3	41.6	-10.3
NZ First	5.0	5.4	3.8	+1.6	8.4	9.8	3.8	+6.0
Other	0.0	0.0	3.1	-3.1	2.6	2.7	3.1	-0.4
None	0.5	-	-	-	4.7	-	-	-
Undecided	1.4	-	-	-	8.1	-	-	-
Refused	0.0	-	-	-	0.9	-	-	-
TOTAL	100.0	100.0	100.0	4.6[1]	100.0	100.0	100.0	5.1[1]

Manawatu - Round 2 Interviews								
	Non-Contextual (N = 117)				Non-Contextual (N = 119)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	11.4	11.8	27.5	-15.7	15.0	16.4	27.5	-11.1
Labour	36.6	39.1	33.4	+5.7	34.7	37.3	33.4	+3.9
National	39.2	41.8	32.6	+9.2	36.2	39.1	32.6	+6.5
NZ First	7.1	7.3	4.6	+2.7	3.3	3.6	4.6	-1.0
Other	0.0	0.0	1.9	-1.9	3.1	3.6	1.9	+1.7
None	2.8	-	-	-	3.7	-	-	-
Undecided	2.4	-	-	-	2.6	-	-	-
Refused	0.5		-	-	1.4	-	-	-
TOTAL	100.0	100.0	100.0	7.0[1]	100.0	100.0	100.0	4.8[1]

Table H.8 continued

Mirimar - Round 2 Interviews								
	Contextual (N = 110)				Non-Contextual (N = 110)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	10.9	11.2	11.6	-0.4	11.1	11.5	11.6	-0.1
Labour	47.8	49.5	46.4	+3.1	33.9	35.6	46.4	-10.8
National	27.6	28.0	34.8	-6.8	38.0	40.4	34.8	+5.6
NZ First	10.1	10.3	4.3	+6.0	10.1	10.6	4.3	+6.3
Other	1.3	1.0	2.9	-1.9	1.7	1.9	2.9	-1.0
None	0.9	-	-	-	3.7	-	-	-
Undecided	1.4	-	-	-	0.0	-	-	-
Refused	0.0	-	-	-	1.6	-	-	-
TOTAL	100.0	100.0	100.0	3.6[1]	100.0	100.0	100.0	4.8[1]

Western Hutt - Round 2 Interviews								
	Non-Contextual (N = 126)				Non-Contextual (N = 127)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	8.1	8.1	13.6	-5.5	15.3	16.2	13.6	+2.6
Labour	37.4	37.9	37.9	0.0	37.1	40.2	37.9	+2.3
National	40.7	41.1	41.6	-0.5	33.5	35.9	41.6	-5.7
NZ First	10.1	10.5	3.8	+6.7	6.6	6.8	3.8	+3.0
Other	2.2	2.4	3.1	-0.7	0.6	0.9	3.1	-2.2
None	0.8	-	-	-	3.0	-	-	-
Undecided	0.0	-	-	-	3.1	-	-	-
Refused	0.7	-	-	-	0.8	-	-	-
TOTAL	100.0	100.0	100.0	2.7[1]	100.0	100.0	100.0	3.2[1]

1. Mean absolute error

Table H.9: Secret Ballot Results weighted by Voting Probability

Manawatu - Round 1 Interviews								
	Contextual (N = 89)				Non-Contextual (N = 89)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	11	12.8	27.5	-14.7	16	17.8	27.5	-9.7
Labour	34	39.5	33.4	+6.1	31	34.4	33.4	+1.0
National	34	39.5	32.6	+6.9	30	33.3	32.6	+0.7
NZ First	06	7.0	4.6	+2.4	12	13.3	4.6	+8.7
Other	01	1.2	1.9	-0.7	01	1.2	1.9	-0.7
None								
Undecided								
Refused								
TOTAL	86	100.0	100.0	6.2[1]	90	100.0	100.0	4.2[1]

Miramar - Round 1 Interviews								
	Non-Contextual (N = 87)				Non-Contextual (N = 90)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	12	13.5	11.6	+1.9	10	11.9	11.6	+0.3
Labour	41	46.1	46.4	-0.3	35	41.7	46.4	-4.7
National	29	32.6	34.8	-2.2	30	35.7	34.8	+0.9
NZ First	6	6.7	4.3	+2.4	08	9.5	4.3	+5.2
Other	1	1.1	2.9	-1.8	01	1.2	2.9	-1.7
None								
Undecided								
Refused								
TOTAL	89	100.0	100.0	1.7[1]	84	100.0	100.0	2.6[1]

Table H.9 continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 128)				Non-Contextual (N = 110)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	12	13.2	13.6	-0.4	9	10.1	13.6	-3.5
Labour	40	44.0	37.9	+6.1	38	42.7	37.9	+4.8
National	33	36.3	41.6	-5.3	32	36.0	41.6	-5.6
NZ First	6	6.5	3.8	+2.7	7	7.9	3.8	+4.1
Other	0	0.0	3.1	-3.1	3	3.3	3.1	+0.2
None								
Undecided								
Refused								
TOTAL	91	100.0	100.0	3.5[1]	89	100.0	100.0	3.6[1]

Manawatu - Round 2 Interviews								
	Non-Contextual (N = 109)				Non-Contextual (N = 109)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	13	13.7	27.5	-13.8	16	17.0	27.5	-10.5
Labour	39	41.0	33.4	+7.6	36	38.3	33.4	+4.9
National	36	37.9	32.6	+5.3	36	38.3	32.6	+5.7
NZ First	7	7.4	4.6	+2.8	4	4.3	4.6	-0.3
Other	0	0.0	1.9	-1.9	2	2.1	1.9	+0.2
None								
Undecided								
Refused								
TOTAL	95	100.0	100.0	6.3[1]	94	100.0	100.0	4.3[1]

Table H.9 continued

Miramar - Round 2 Interviews								
	Contextual (N = 108)				Non-Contextual (N = 104)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	13	14.2	11.6	+2.6	12	13.3	11.6	+1.7
Labour	42	46.2	46.4	-0.2	33	36.7	46.4	-9.7
National	24	26.4	34.8	-8.4	35	38.9	34.8	+4.1
NZ First	10	11.0	4.3	+6.7	9	10.0	4.3	+5.7
Other	2	2.2	2.9	-0.7	1	1.1	2.9	-1.8
None								
Undecided								
Refused								
TOTAL	91	100.0	100.0	3.7[1]	90	100.0	100.0	4.6[1]

Western Hutt - Round 2 Interviews								
	Non-Contextual (N = 122)				Non-Contextual (N = 116)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	8	8.8	13.6	-4.8	15	16.5	13.6	+2.9
Labour	37	40.6	37.9	+2.7	37	40.6	37.9	+2.7
National	36	39.6	41.6	-2.0	32	35.2	41.6	-6.4
NZ First	8	8.8	3.8	+5.0	6	6.6	3.8	+2.8
Other	2	2.2	3.1	-0.9	1	1.1	3.1	-2.0
None								
Undecided								
Refused								
TOTAL	91	100.0	100.0	3.1[1]	91	100.0	100.0	3.4[1]

1. Mean absolute errors

Table H.10: Secret Ballot Results weighted by 1990 Voting Patterns

Manawatu - Round 1 Interviews								
	Contextual (N = 101)				Non-Contextual (N = 104)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	11.9	13.5	27.5	-14.0	15.4	18.0	27.5	-9.3
Labour	35.6	40.4	33.4	+7.0	29.8	34.8	33.4	+1.4
National	33.7	38.2	32.6	+5.6	27.9	32.6	32.6	0.0
NZ First	5.9	6.7	4.6	+2.1	11.5	13.5	4.6	+8.9
Other	1.0	1.2	1.9	-0.7	1.0	1.1	1.9	-0.8
None	2.0	-	-	-	6.7	-	-	-
Undecided	5.9	-	-	-	7.7	-	-	-
Refused	4.0	-	-	-	-	-	-	-
TOTAL	100.0	100.0	100.0	5.9[1]	100.0	100.0	100.0	4.1[1]

Miramar - Round 1 Interviews								
	Contextual (N = 96)				Non-Contextual (N = 97)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-allocated	Actual Results	Error
Alliance	13.5	14.9	11.6	+3.3	12.4	13.3	11.6	+1.7
Labour	39.6	43.7	46.4	-2.7	37.1	40.0	46.4	-6.4
National	31.3	34.5	34.8	-0.3	32.0	34.4	34.8	-0.4
NZ First	5.2	5.7	4.3	+1.4	9.3	10.0	4.3	+5.7
Other	1.0	1.2	2.9	-1.7	2.1	2.3	2.9	-0.6
None	7.3	-	-	-	3.1	-	-	-
Undecided	1.0	-	-	-	2.1	-	-	-
Refused	1.0	-	-	-	2.1	-	-	-
TOTAL	100.0	100.0	100.0	1.9[1]	100.0	100.0	100.0	3.0[1]

Table H.10 continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 128)				Non-Contextual (N = 121)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	14.7	15.1	13.6	+1.5	9.1	10.8	13.6	-2.8
Labour	43.1	43.7	37.9	+5.8	31.4	37.3	37.9	-0.6
National	33.3	34.1	41.6	-7.5	33.7	40.2	41.6	-1.4
NZ First	6.6	7.1	3.8	+3.3	9.1	10.8	3.8	+7.0
Other	0.0	0.0	3.1	-3.1	1.2	0.9	3.1	-2.2
None	0.8	-		-	5.0	-	-	-
Undecided	1.5	-		-	9.7	-	-	-
Refused	0.0	-		-	0.8	-	-	-
TOTAL	100.0	100.0	100.0	4.2[1]	100.0	100.0	100.0	2.8[1]

Manawatu - Round 2 Interviews								
	Contextual (N = 113)				Non-Contextual (N = 112)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	9.7	10.3	27.5	-17.2	19.5	21.0	27.5	-6.5
Labour	47.5	50.4	33.4	+17.0	32.6	34.3	33.4	+0.9
National	30.2	31.8	32.6	-0.8	34.1	36.2	32.6	+3.6
NZ First	6.7	7.5	4.6	+2.9	5.1	5.7	4.6	+1.1
Other	0.0	0.0	1.9	-1.9	2.4	2.8	1.9	+0.9
None	3.2	-	-	-	3.3	-	-	-
Undecided	2.0	-	-	-	2.3	-	-	-
Refused	0.7	-	-	-	0.7	-	-	-
TOTAL	100.0	100.0	100.0	8.0[1]	100.0	100.0	100.0	2.6[1]

Table H.10 continued

Miramar - Round 2 Interviews								
	Contextual (N = 119)				Non-Contextual (N = 104)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	17.3	17.6	11.6	+6.0	11.1	12.2	11.6	+0.6
Labour	43.9	44.4	46.4	-2.0	36.0	37.8	46.4	-8.6
National	25.1	25.0	34.8	-9.8	37.0	38.8	34.8	+4.0
NZ First	11.0	11.1	4.3	+6.8	10.1	10.2	4.3	+5.9
Other	1.7	1.9	2.9	-1.0	0.9	1.0	2.9	-1.9
None	0.9	-	-	-	3.9	-	-	-
Undecided	0.0	-	-	-	0.0	-	-	-
Refused	0.0	-	-	-	1.0	-	-	-
TOTAL	100.0	100.0	100.0	5.1[1]	100.0	100.0	100.0	4.2[1]

Western Hutt - Round 2 Interviews								
	Contextual (N = 122)				Non-Contextual (N = 115)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	9.1	9.3	13.6	-4.3	12.2	13.1	13.6	-0.5
Labour	41.3	42.0	37.9	+4.1	38.1	41.1	37.9	+3.2
National	35.5	36.2	41.6	-5.4	36.0	38.3	41.6	-3.3
NZ First	9.8	10.0	3.8	+6.2	6.3	6.5	3.8	+2.7
Other	2.7	2.5	3.1	-0.6	0.9	1.0	3.1	-2.1
None	0.8	-		-	2.9	-	-	-
Undecided	0.0	-		-	2.6	-	-	-
Refused	0.7	-		-	1.0	-	-	-
TOTAL	100.0	100.0	100.0	4.1[1]	100.0	100.0	100.0	2.4[1]

1. Mean absolute error

Table H.11: Juster Questionnaire Results Weighted by Demographic Variables

Manawatu - Round 1 Interviews								
	Contextual (N = 105)				Non-Contextual (N = 97)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	20.5	17.3	27.5	-10.2	28.7	25.5	27.5	-2.0
Labour	46.0	38.7	33.4	+5.3	33.4	29.6	33.4	-3.8
National	30.6	25.8	32.6	-6.8	24.4	21.6	32.6	-11.0
NZ First	15.1	12.7	4.6	+8.1	20.9	18.6	4.6	+14.0
Other	6.5	5.5	1.9	+3.6	5.3	4.7	1.9	+2.8
TOTAL	118.7	100.0	100.0	6.8[1]	112.7	100.0	100.0	6.7[1]

Miramar - Round 1 Interviews								
	Contextual (N = 97)				Non-Contextual (N = 95)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	17.1	13.9	11.6	+2.3	17.3	15.0	11.6	+3.4
Labour	50.0	40.8	46.4	-5.6	41.3	35.9	46.4	-10.5
National	29.3	23.9	34.8	-10.9	33.6	29.2	34.8	-5.6
NZ First	16.8	13.7	4.3	+9.4	14.5	12.5	4.3	+8.2
Other	9.4	7.7	2.9	+4.8	8.5	7.4	2.9	+4.5
TOTAL	122.6	100.0	100.0	6.6[1]	115.2	100.0	100.0	6.4[1]

Table H.11 continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 129)				Non-Contextual (N = 131)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	16.3	13.6	13.6	0	15.0	14.5	13.6	+0.9
Labour	42.4	35.4	37.9	-2.5	37.5	36.3	37.9	-1.6
National	37.8	31.6	41.6	-10.0	34.7	33.7	41.6	-7.9
NZ First	18.0	15.0	3.8	+11.2	11.8	11.4	3.8	+7.6
Other	5.3	4.4	3.1	+1.3	4.2	4.1	3.1	+1.0
TOTAL	119.8	100.0	100.0	5.0[1]	103.2	100.0	100.0	3.8[1]

Manawatu - Round 2 Interviews								
	Contextual (N = 124)				Non-Contextual (N = 117)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	23.1	20.5	27.5	-7.0	21.8	17.7	27.5	-9.8
Labour	40.7	36.1	33.4	+2.7	44.2	36.0	33.4	+2.6
National	30.3	26.9	32.6	-5.7	38.2	31.1	32.6	-1.5
NZ First	15.0	13.3	4.6	+8.7	12.5	10.2	4.6	+5.6
Other	3.6	3.2	1.9	+1.3	6.1	5.0	1.9	+3.1
TOTAL	112.7	100.0	100.0	5.1[1]	122.8	100.0	100.0	4.5[1]

Table H.11 continued

Miramar - Round 2 Interviews								
	Contextual (N = 112)				Non-Contextual (N = 112)			
Party	**Initial Estimates**	**Re-Allocated**	**Actual Results**	**Error**	**Initial Estimates**	**Re-Allocated**	**Actual Results**	**Error**
Alliance	18.6	15.0	11.6	+3.4	16.9	14.2	11.6	+2.6
Labour	40.5	32.7	46.4	-13.7	49.5	41.6	46.4	-4.8
National	39.0	31.5	34.8	-3.3	30.0	25.2	34.8	-9.6
NZ First	18.2	14.7	4.3	+10.4	13.6	11.4	4.3	+7.1
Other	7.5	6.1	2.9	+3.2	9.0	7.6	2.9	+4.7
TOTAL	123.8	100.0	100.0	6.8[1]	119.0	100.0	100.0	5.8[1]

Western Hutt - Round 2 Interviews								
	Contextual (N = 128)				Non-Contextual (N = 129)			
Party	**Initial Estimates**	**Re-Allocated**	**Actual Results**	**Error**	**Initial Estimates**	**Re-Allocated**	**Actual Results**	**Error**
Alliance	19.2	16.4	13.6	+2.8	17.5	14.7	13.6	+1.1
Labour	36.6	31.2	37.9	-6.7	42.4	35.7	37.9	-2.2
National	42.1	36.0	41.6	-5.6	37.9	31.9	41.6	-9.7
NZ First	13.4	11.4	3.8	+7.6	14.4	12.1	3.8	+8.3
Other	5.8	5.0	3.1	+1.9	6.7	5.6	3.1	+2.5
TOTAL	117.1	100.0	100.0	4.9[1]	118.9	100.0	100.0	4.8[1]

1. Mean absolute errors

Table H.12: Juster Questionnaire Results Weighted by Probability of Voting

Manawatu - Round 1 Interviews								
	Contextual (N = 99)				Non-Contextual (N = 95)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	14	15.6	27.5	-11.9	21	23.3	27.5	-4.2
Labour	37	41.1	33.4	+7.7	28	31.1	33.4	-2.3
National	26	28.9	32.6	-3.7	22	24.5	32.6	-8.1
NZ First	9	10.0	4.6	+5.4	15	16.7	4.6	+12.1
Other	4	4.4	1.9	+2.5	4	4.4	1.9	+2.5
TOTAL	90	100.0	100.0	6.2[1]	90	100.0	100.0	5.8[1]

Miramar - Round 1 Interviews								
	Contextual (N = 96)				Non-Contextual (N = 92)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	11	12.5	11.6	+0.9	11	12.8	11.6	+1.2
Labour	39	43.8	46.4	-2.6	32	37.2	46.4	-9.2
National	23	25.8	34.8	-9.0	29	33.7	34.8	-1.1
NZ First	10	11.2	4.3	+6.9	9	10.5	4.3	+6.2
Other	6	6.7	2.9	+3.8	5	5.8	2.9	+2.9
TOTAL	89	100.0	100.0	4.6[1]	86	100.0	100.0	4.1[1]

Table H.12 continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 124)				Non-Contextual (N = 127)			
Party	**Initial Estimates**	**Re-Allocated**	**Actual Results**	**Error**	**Initial Estimates**	**Re-Allocated**	**Actual Results**	**Error**
Alliance	12	13.3	13.6	-0.3	12	13.6	13.6	0
Labour	33	36.7	37.9	-1.2	34	38.6	37.9	+0.7
National	31	34.4	41.6	-7.2	30	34.2	41.6	-7.4
NZ First	12	13.3	3.8	+9.5	9	10.2	3.8	+6.4
Other	2	2.3	3.1	-0.8	3	3.4	3.1	+0.3
TOTAL	90	100.0	100.0	3.8[1]	88	100.0	100.0	3.0[1]

Manawatu - Round 2 Interviews								
	Contextual (N = 121)				Non-Contextual (N = 115)			
Party	**Initial Estimates**	**Re-Allocated**	**Actual Results**	**Error**	**Initial Estimates**	**Re-Allocated**	**Actual Results**	**Error**
Alliance	19	20.9	27.5	-6.6	16	17.6	27.5	-9.9
Labour	34	37.3	33.4	+3.9	34	37.3	33.4	+3.9
National	25	27.5	32.6	-5.1	30	33.0	32.6	+0.4
NZ First	11	12.1	4.6	+7.5	8	8.8	4.6	+4.2
Other	2	2.2	1.9	+0.3	3	3.3	1.9	+1.4
TOTAL	91	100.0	100.0	4.7[1]	91	100.0	100.0	4.0[1]

Table H.12 continued

Miramar - Round 2 Interviews								
	Contextual (N = 110)				Non-Contextual (N = 109)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	13	14.4	11.6	+2.8	11	12.0	11.6	+0.4
Labour	32	35.7	46.4	-10.7	41	44.6	46.4	-1.8
National	30	33.3	34.8	-1.5	25	27.2	34.8	-7.6
NZ First	11	12.2	4.3	+7.9	10	10.8	4.3	+6.5
Other	4	4.4	2.9	+1.5	5	5.4	2.9	+2.5
TOTAL	90	100.0	100.0	4.9[1]	92	100.0	100.0	3.8[1]

Western Hutt - Round 2 Interviews								
	Contextual (N = 125)				Non-Contextual (N = 126)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	13	14.7	13.6	+1.1	13	14.4	13.6	+0.8
Labour	29	33.0	37.9	-4.9	32	35.6	37.9	-2.3
National	35	39.8	41.6	-1.8	33	36.7	41.6	-4.9
NZ First	8	9.1	3.8	+5.3	9	10.0	3.8	+6.2
Other	3	3.4	3.1	+0.3	3	3.3	3.1	+0.2
TOTAL	88	100.0	100.0	2.7[1]	90	100.0	100.0	2.9[1]

1. Mean absolute errors

Table H.13: Juster Questionnaire Results Weighted by 1990 Voting Behaviour

Manawatu - Round 1 Interviews								
	Contextual (N = 118)				Non-Contextual (N = 108)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	15	15.3	27.5	-12.2	21	23.6	27.5	-3.9
Labour	43	43.9	33.4	+10.5	28	31.5	33.4	-1.9
National	28	28.6	32.6	-4.0	23	25.8	32.6	-6.8
NZ First	9	9.2	4.6	+4.6	13	14.6	4.6	+10.0
Other	3	3.0	1.9	+1.1	4	4.5	1.9	+2.6
TOTAL	98	100.0	100	6.5[1]	89	100.0	100.0	5.0[1]

Miramar - Round 1 Interviews								
	Contextual (N = 113)				Non-Contextual (N = 111)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	12	13.3	11.6	+1.7	12	14.1	11.6	+2.5
Labour	37	41.1	46.4	-5.3	33	38.9	46.4	-7.5
National	24	26.7	34.8	-8.1	28	32.9	34.8	-1.9
NZ First	10	11.1	4.3	+6.8	8	9.4	4.3	+5.1
Other	7	7.8	2.9	+4.9	4	4.7	2.9	+1.8
TOTAL	90	100.0	100.0	5.4[1]	85	100.0	100.0	3.8[1]

Table H.13 continued

Western Hutt - Round 1 Interviews								
	Contextual (N = 95)				Non-Contextual (N = 107)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	12	13.5	13.6	-0.1	14	15.6	13.6	+2.0
Labour	33	37.1	37.9	-0.8	36	40.0	37.9	+2.1
National	31	34.8	41.6	-6.8	28	31.1	41.6	-10.5
NZ First	11	12.4	3.8	+8.6	9	10.0	3.8	+6.2
Other	2	2.2	3.1	-0.9	3	3.3	3.1	+0.2
TOTAL	89	100.0	100.0	3.4[1]	90	100.0	100.0	4.2[1]

Manawatu - Round 2 Interviews								
	Contextual (N = 135)				Non-Contextual (N = 126)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	21	22.3	27.5	-5.2	21	22.6	27.5	-4.9
Labour	32	34.0	33.4	+0.6	30	32.3	33.4	-1.1
National	28	29.9	32.6	-2.7	31	33.3	32.6	+0.7
NZ First	11	11.7	4.6	+7.1	8	8.6	4.6	+4.0
Other	2	2.1	1.9	+0.2	3	3.2	1.9	+1.3
TOTAL	94	100.0	100.0	3.2[1]	93	100.0	100.0	2.4[1]

Table H.13 continued

Miramar - Round 2 Interviews								
	Contextual (N = 127)				Non-Contextual (N = 130)			
Party	Initial Estimates	Re-Allocated	Actual Results	Error	Initial Estimates	Re-allocated	Actual Results	Error
Alliance	14	15.2	11.6	+3.6	11	12.0	11.6	+0.4
Labour	33	35.9	46.4	-10.5	40	43.5	46.4	-2.9
National	30	32.6	34.8	-2.2	27	29.3	34.8	-5.5
NZ First	11	12.0	4.3	+7.7	9	9.8	4.3	+5.5
Other	4	4.3	2.9	+1.4	5	5.4	2.9	+2.5
TOTAL	92	100.0	100.0	5.1[1]	92	100.0	100.0	3.4[1]

Western Hutt - Round 2 Interviews								
	Contextual (N = 138)				Non-Contextual (N = 142)			
Party	Initial Estimates	Re-allocated	Actual Results	Error	Initial Estimates	Re-Allocated	Actual Results	Error
Alliance	12	14.1	13.6	+0.5	16	17.6	13.6	+4.0
Labour	29	34.1	37.9	-3.8	32	35.2	37.9	-2.7
National	34	40.1	41.6	-1.5	32	35.2	41.6	-6.4
NZ First	7	8.2	3.8	+4.4	8	8.8	3.8	+5.0
Other	3	3.5	3.1	+0.4	3	3.2	3.1	+0.1
TOTAL	85	100.0	100.0	2.1[1]	91	100.0	100.0	3.6[1]