

**PRIOR TAX WITHHOLDINGS, DECISION FRAMES
AND MOVEMENTS IN TAXPAYERS' RISK
PREFERENCES WITH RESPECT TO NON-
COMPLIANCE**

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Master of Commerce in Accountancy
in the
University of Canterbury
by
Warwick W. Anderson

University of Canterbury

1996

A Replication and Extension of R. Dusenbury's Study,
"The Effect of Prepayment Position on Individual Taxpayers'
Preferences for Risky Tax-Filing Options."

(R. Dusenbury's paper was published in the Journal of the American Taxation Association,
Volume 16, Number 1, Spring 1994.)

“Doubt is not a pleasant condition,
but certainty is an absurd one.”

Voltaire (Francois Marie Arouet 1694 – 1778)
in a letter to Frederick the Great, April 6 1767.

TABLE OF CONTENTS

ABSTRACT	xi
1. INTRODUCTION	1
1.1 Statement of the Research Questions	1
1.2 Relevance of the Research	5
1.3 Definition of Terms	6
1.4 Overview of the Study	9
2. PROSPECT THEORY	11
2.1 Introduction	11
2.2 The Economic Rationality Problem	11
2.3 Expected Utility Theory	14
2.3.1 The Axioms, Principles and Assumptions underlying Expected Utility Theory.	16
2.3.2 Why Expected Utility Theory is Inadequate.	21
2.4 Prospect Theory	28
2.4.1 Part One: The Editing Phase	28
2.4.2 Part Two: The Evaluation Phase	32
2.4.3 The Equations of Prospect Theory	34
2.4.4 Empirical Corroboration	35
2.5 Cumulative Prospect Theory	37
2.5.1 The Cumulative Prospect Theory Value Function	37
2.5.2 Decision Weights in Cumulative Prospect Theory	39
2.6 The Framing of Decisions	41
2.6.1 Framing Effects and Language-Induced Perceptions of Loss and Gain	41
2.6.2 Psychological Accounts	44
2.6.3 Research involving Decision Frames outside the area of Tax	46
3. TAX RESEARCH INVOLVING PROSPECT THEORY AND DECISION FRAMES	51
3.1 Introduction	51
3.2 Origins of Decision Frame Research in Compliance	52
3.3 Framing Effect Research in Compliance Prior to Dusenbury	56
3.3.1 Decision Frames in the Context of Tax	57
3.3.2 Prospect Theory Applied	59
3.3.3 Gain/Loss Perceptions	62
3.3.4 Rates of Voluntary Compliance	65
3.3.5 Laboratory Analysis	66
3.3.6 Improvement in the Realism and Assumptions of Scenarios	69
3.3.7 International and Cross-Cultural Extensions	71
3.3.8 Other 1990s Papers	73
3.4 Framing Effects and Professional Tax Preparers	76
4. HYPOTHESES AND RESEARCH DESIGN	79
4.1 Introduction	79
4.2 Dusenbury (1994): His Research Objectives and Results	80
4.2.1 Dusenbury's Objectives and Research Design	80
4.2.2 Dusenbury's Subjects	86
4.2.3 The Experimental Procedure	86
4.2.4 Monetary Incentive	87
4.2.5 Questionnaire	88
4.2.6 Dusenbury's Results	89
4.2.7 Links with Earlier Research and Dusenbury's Unique Contribution	90

4.3 Why a Replication?	90
4.3.1 Definition	90
4.3.2 Purpose	91
4.4 The Replication Hypotheses	92
4.4.1 Hypotheses	92
4.4.2 Research Design Modifications	92
4.5 Extension Hypotheses	96
4.5.1 Cash Flow Extension	97
4.5.2 Summary Syndrome Extension	98
4.5.3 Risk Profile Stability Extension	102
4.5.4 Value Function Extension	104
5. SESSION PROCEDURES	109
5.1 Introduction	109
5.2 Session format	109
5.2.1 Introduction	109
5.2.2 Session Kit and Seating	110
5.2.3 Procedure	111
5.2.4 Cash Float Considerations	112
5.3 Peripheral Documentation	114
5.3.1 Introduction	114
5.3.2 Statement of Purpose	115
5.3.3 Ethics Committee Rubric	116
5.3.4 Debriefing Document	116
5.3.5 Tally Sheet	117
5.4 The Monetary Incentive	119
5.4.1 Introduction	119
5.4.2 Monetary Incentives in Prior Studies	120
5.4.3 Setting the Monetary Incentive in the Current Study	123
6. THE EXPERIMENTAL INSTRUMENT	125
6.1 Introduction	125
6.2 Lessons from Pretesting	126
6.3 Simplification of Tax Law in the Experiment	127
6.4 Order of Presentation	128
6.5 Training of Participants	129
6.5.1 Training Script	129
6.5.2 Sample Case	130
6.6 Scripts used in the Conceptual Replication	130
6.6.1 Refund Tax Case Y	130
6.6.2 High Pay Tax Case X	132
6.6.3 Low Pay Tax Case Z	133
6.6.4 Medical Insurance Case	133
6.6.5 Gamble Case	133
6.7 Additional Ratings Questions	134
6.8 Extra Scripts	135
6.8.1 Introduction	135
6.8.2 High Pay Tax Case W	135
6.9 Questionnaire	136
6.9.1 Introduction	136
6.9.2 Category One: Demographic Questions	137
6.9.3 Category Two: Solicitation of Attitudes	140
6.9.4 Category Three: Perceptions of Taxpayer and IRD Behaviour	145
7. THE SAMPLE	149
7.1 Introduction	149
7.2 Recruitment Strategy	149

7.2.1 Introduction	149
7.2.2 Why the Sample was Non-Random	149
7.2.3 Recruitment Technique	150
7.3 Organisations	151
7.3.1 Introduction	151
7.3.2 The Church Subsample	151
7.3.3 The Kindergarten Subsample	152
7.3.4 The School Subsample	152
7.3.5 The Lincoln University Subsample	153
7.4 Modifications to the Sample of Participants	153
7.5 Sample Size with Respect to the Hypotheses	154
7.5.1 Introduction	154
7.5.2 Sample Size and the Replication Hypotheses	154
7.5.3 Sample Size and the Extension Hypotheses	156
8. CONCEPTUAL REPLICATION RESULTS	161
8.1 Introduction	161
8.2 Initial Inspection of Summary Information	162
8.3 Testing for Validation	167
8.3.1 Introduction	167
8.3.2 Categorical Data Modelling Procedure	168
8.3.3 The General Linear Model Procedure	168
8.4 Tests for Main and Interaction Effects on the Model	170
8.5 Hypothesis H_{1A}	174
8.5.1 Introduction	174
8.5.2 The Full Sample and H_{1A}	175
8.5.3 The Low Cash Float Subsample and H_{1A}	175
8.5.4 The Various High Cash Float Subsamples and H_{1A}	176
8.5.5 Investigation Incorporating Compliance Attitude Variables	178
8.5.6 H_{1A} : The Complete Picture	182
8.6 Hypothesis H_{2A}	183
8.6.1 The Full Sample and H_{2A}	183
8.6.2 The Low Cash Float Subsample and H_{2A}	187
8.6.3 The Combined High Cash Float Subsample and H_{2A}	189
8.6.4 H_{2A} : The Complete Picture	190
8.7 Hypothesis H_{3A}	191
8.7.1 The Full Sample and H_{3A}	191
8.7.2 The Low Cash Float Subsample and H_{3A}	193
8.7.3 The Various High Cash Float Subsamples and H_{3A}	193
8.7.4 Investigation Incorporating the Attitude to Gambling Variable	194
8.7.5 H_{3A} : The Complete Picture	197
9. CHARACTERISTICS NOTED FROM THE QUESTIONNAIRE	199
9.1 Introduction	199
9.2 Nature of the Sample	200
9.2.1 Introduction	200
9.2.2 Comparison with Dusenbury's Requirements	200
9.2.3 Further Salient Features of the New Zealand Sample	203
9.2.4 Representational Faithfulness of the Sample	206
9.3 The Impact of the Questionnaire Variables	210
9.3.1 Introduction	210
9.3.2 Category One: Demographic Variables	211
9.3.3 Category Two: Attitude Variables	223
9.3.4 Category Three: Perception Variables relating to Taxpayer and IRD Behaviour	230
10. RESULTS OF EXTENSION HYPOTHESES	233
10.1 Introduction	233

10.2 Cash Flow Extension	233
10.3 Summary Syndrome Extension	237
10.3.1 Partitioning by Levels of Declared Influence.	238
10.3.2 Declared Influence Incorporated as a Categorical Variable	240
10.4 Risk Profile Stability Extension	244
10.4.1 Results Relating to H_{6A}	245
10.4.2 Results Relating to H_{7A}	247
10.5 Value Function Extension	249
11. CONCLUSIONS, LIMITATIONS AND EXTENSIONS	255
11.1 Introduction	255
11.2 Conclusions	255
11.3 Limitations	267
11.4 Extensions	270
11.5 Implications for Tax Administration	273
12. REFERENCES	275
ACKNOWLEDGEMENTS	285
APPENDIX A: DUSENBURY'S INSTRUMENT	289
A.1 Dusenbury's Training Script	289
A.2 Dusenbury's Five Scenarios	293
A.3 Additional Rating of Judgment Situations	298
APPENDIX B: INSTRUMENT USED IN THIS STUDY	299
B.1 Preliminary Document	299
B.2 Training Script	300
B.3 The Experimental Instrument	306
B.4 Questionnaire	313
B.5 Tally Sheet	317
B.6 Debriefing Document.	318
APPENDIX C: HOW THE GLM PROCEDURE WORKS	319
APPENDIX D: DATA GROOMING	321
D.1 The Basic Model GLM Procedure	321
D.2 The Repeated Measures GLM Procedure	321
APPENDIX E: VALIDATION STATISTICS	323
E.1 MANOVA Statistics	323
E.2 Basic Model GLM Procedure Results	324
E.3 Repeated Measures Results for Individual Cases	326
APPENDIX F: FURTHER HYPOTHESIS TEST RESULTS	327
F.1 Results For H_{1A}	327
F.2 Results for H_{2A}	328
F.3 Results for H_{3A}	331
APPENDIX G: SPECIMEN TESTS OF THE DEMOGRAPHIC VARIABLES	335

TABLE OF TABLES

TABLE 4.1: Summary Table Provided for High Pay Tax Case X.	81
TABLE 4.2: Summary Table Provided for Refund Tax Case Y.	82
TABLE 4.3: Construction of Tax Cases in Dollars.	83
TABLE 4.4: Differences between High Pay Tax Cases X and W.	103
TABLE 5.1: A Sample Tally Sheet.	118
TABLE 5.2: (Tear-off Slip).	119
TABLE 6.1: Demographic Questions in Numerical Order.	138
TABLE 6.2: Questions Soliciting Participants' Attitudes.	140
TABLE 6.3: Questions Soliciting Participants' Tax-Related Perceptions.	145
TABLE 7.1: Sample Sizes with Respect to the Replication Hypotheses.	155
TABLE 7.2: Summary Syndrome Extension Sample Sizes for Refund Tax Case Y.	157
TABLE 7.3: Summary Syndrome Extension Sample Sizes for High Pay Tax Case X.	158
TABLE 7.4: Sample Size for Risk Profile Stability Extension.	158
TABLE 7.5: Sample Size for Value Function Extension.	159
TABLE 8.1: Subjects' Choices: Descriptive Statistics.	163
TABLE 8.2: Subjects' Choices Partitioned into Low and High Cash Float Subsamples.	164
TABLE 8.3: Subjects' Choices: High Cash Float Subsample Partitioned into \$3,500 and \$4,200 Subsamples.	166
TABLE 8.4: Tests for Main and Interaction Effects on the Five Replication Scenarios. Full Sample (N = 132).	172
TABLE 8.5: Tests for Main and Interaction Effects on the Five Replication Scenarios. Low Cash Float Subsample (N = 66).	173
TABLE 8.6: Refund Tax Case Y and High Pay Tax Case X. Full Sample (N = 132).	175
TABLE 8.7: Refund Tax Case Y and High Pay Tax Case X. Low Cash Float Subsample (N = 66).	176
TABLE 8.8: Refund Tax Case Y and High Pay Tax Case X. Combined High Cash Float Subsample (N = 66).	177
TABLE 8.9: Refund Tax Case Y and High Pay Tax Case X. \$3,500 Cash Float Subsample (N = 36).	178
TABLE 8.10: Refund Tax Case Y and High Pay Tax Case X. \$4,200 Cash Float Subsample (N = 30).	179
TABLE 8.11: Responses to Compliance Questions with Respect to Refund Tax Case Y and High Pay Tax Case X. Full Sample (N = 132).	180
TABLE 8.12: Responses to Compliance Questions with Respect to Refund Tax Case Y and High Pay Tax Case X. \$3,500 Subsample (N = 36).	181
TABLE 8.13: Responses to Compliance Questions with Respect to Refund Tax Case Y and High Pay Tax Case X. \$4,200 Subsample (N = 30).	182

TABLE 8.14: High Pay Tax Case X versus Medical Insurance Case. Full Sample (N = 132).	184
TABLE 8.15: High Pay Tax case X versus the Medical Insurance Case Partitioned by Q4 (Attitude to Health Insurance). Full Sample (N = 132).	186
TABLE 8.16: Means obtained when the Full Sample was partitioned by Attitude to Personal Health Insurance (Q4).	187
TABLE 8.17: High Pay Tax Case X versus Medical Insurance Case. Low Cash Float Subsample (N = 66).	188
TABLE 8.18: Means obtained when the Low Cash Float Subsample was partitioned by Attitude to Personal Health Insurance (Q4).	189
TABLE 8.19: High Pay Tax Case X versus Medical Insurance Case Effect of Partitioning by Q4 (Attitude to Health Insurance). Combined High Cash Float Subsample (N = 66).	190
TABLE 8.20: Low Pay Tax Case Z versus Gamble Case. Full Sample (N = 132).	192
TABLE 8.21: Low Pay Tax Case Z versus the Gamble Case Low Cash Float Subsample (N = 66).	192
TABLE 8.22: Low Pay Tax Case Z versus the Gamble Case. Combined High Cash Float Subsample (N = 66).	193
TABLE 8.23: Low Pay Tax Case Z versus the Gamble Case. \$3,500 Cash Float Subsample (N = 36).	194
TABLE 8.24: Low Pay Tax Case Z versus Gamble. \$4,200 Cash Float Subsample (N = 30).	195
TABLE 8.25: Low Pay Tax Case Z versus the Gamble Case with Respect to Attitudes to Gambling. \$4,200 Cash Float Subsample (N = 30).	196
TABLE 8.26: Low Pay Tax Case Z versus the Gamble Case with Respect to Disclosed Attitudes to Compliance. \$4,200 Cash float subsample (N = 30).	198
TABLE 9.1: Breakdown by Age and Recruitment Origin (N = 132).	201
TABLE 9.2: Breakdown by Tax Filing Experience and by Recruitment Origin (N = 132).	203
TABLE 9.3: Breakdown by Frequency of Payment-Due Tax Filings (N = 132).	204
TABLE 9.4: Breakdown by Nature of Income Earning Experience (N=132).	205
TABLE 9.5: Income Earning Experience: Sample versus Census. Full Sample (N = 132).	207
TABLE 9.6: Income Estimates (Full Sample N = 131).	209
TABLE 9.7: The Three Tax Cases and the Eleven Demographic Variables. Full Sample (N = 130).	212
TABLE 9.8:: The Three Tax Cases and the Eleven Demographic Variables. Full Sample (N = 130).	213
TABLE 9.9 : F-statistics Associated with CASE on Subsamples Determined by Age (for H_{1A}).	217
TABLE 9.10: Refund Tax Case Y and High Pay Tax Case X: Effect of Partitioning by Income Earning Experience. Full Sample (N = 113).	221

TABLE 9.11: Effect of Source of Income on Individual Case Responses. Full Sample (N = 113).	222
TABLE 9.12: The Three Tax Cases and Q4 and Q5. Full Sample (N = 131).	224
TABLE 9.13: The Three Tax Cases and Q11. Full Sample (N = 132).	226
TABLE 9.14: The Three Tax Cases and Q6, Q7 and Q8. Full Sample (N = 132).	227
TABLE 9.15: Means of Tax Cases X, Y and Z: Subsamples Determined by Strongly Disapproving Responses to Q6, Q7 and Q11.	229
TABLE 9.16: Three Tax Cases and Q10 and Q16. Full Sample (N = 128).	230
TABLE 10.1: Cash Flow Effect on Responses to High Pay Tax Case X (C2) Low Cash Float and \$3,500 Cash Float. (NPAR1WAY PROCEDURE).	235
TABLE 10.2: Summary of NPAR1WAY Procedure Results for all other Comparisons.	236
TABLE 10.3: Kolmogorov-Smirnov Two-Sample Test Results (Asymptotic).	237
TABLE 10.4: Refund Tax Case Y and High Pay Tax Case X: CASE Statistics by Partition.	240
TABLE 10.5: Summary Table Syndrome Impact on Choice of Risk Levels: Refund Tax Case Y and High Pay Tax Case X Considered Separately. Full Model (N = 129).	241
TABLE 10.6: Summary Syndrome Impact on Choice of Risk Levels: Refund Tax Case Y and High Pay Tax Case X Considered Jointly. Full Sample (N = 129).	243
TABLE 10.7: Stability of Risk Profiles in High Pay Tax Cases X and W. All Available Participants (N = 83).	246
TABLE 10.8: Stability of Risk Preference Shifts in Refund Tax Case Y and High Pay Tax Case W. All Available Participants (N = 83).	248
TABLE 10.9: Smoothness Data for Tax Cases X, Y and Z, the Medical Case and the Gamble Case. Full Sample (N = 131).	250
TABLE 10.10: Smoothness Data for Tax Cases X, Y and Z, the Medical Case and the Gamble Case. Full Sample (N = 131).	251
TABLE 10.11: Tax Cases X, Y, Z, the Medical Insurance Case and the Gamble: Smoothness of Value Function. Full Sample (N = 131).	252
TABLE 10.12: Tax Cases X, Y, Z, the Medical Insurance Case and the Gamble: Smoothness of Value Function.	253
TABLE E.1: MANOVA Output for All Five Cases. Full Sample (N = 132).	323
TABLE E.2: Tests for Main and Interaction Effects on the Five Replication Scenarios. Full Sample (N = 132).	324
TABLE E.3: Tests for Main and Interaction Effects on the Five Replication Scenarios. Low Cash Float Subsample (N = 66).	325
TABLE E.4: Univariate Output of the Repeated Measures General Linear Model Procedure: Responses to Individual Cases. Low Cash Float Subsample (N = 66).	326
TABLE F.1: Refund Tax Case Y versus High Pay Tax Case X: Basic Model GLM Analysis of Variance Results.	328

TABLE F.2: High Pay Tax Case X versus Medical Insurance Case: General Linear Models Procedure Repeated Measures Analysis of Variance Univariate Tests of Hypotheses for Within Subject Effects.	329
TABLE F.3: High Pay Tax Case X versus Medical Insurance Case: Basic Model GLM Analysis of Variance Results.	330
TABLE F.4: High Pay Tax case X versus the Medical Insurance Case Partitioned by Q4 (Attitude to Health Insurance). Full Sample (N = 132).	331
TABLE F.5: Low Pay Tax Case Z and the Gamble: Basic Model GLM Analysis of Variance Results.	332
TABLE F.6: Low Pay Tax case Z versus the Gamble Case: General Linear Models Procedure Repeated Measures Analysis of Variance: Tests of Hypotheses for Between Subjects Effects. \$4,200 Cash Float Subsample (N = 30).	333
TABLE F.7: Responses to the Gamble Case versus Responses to Question 11: Basic General Linear Model Analysis of Variance.	333
TABLE F.8: Responses to Low Pay Tax Case Z versus Responses to Questions 6, 7 and 8: Basic General Linear Model Analysis of Variance.	334
TABLE G.1: Tax Cases X, Y and Z, and Gender, Age and Educational Level. Full Sample (N = 132).	339
TABLE G.2: Tax Cases X, Y and Z, and Gender, Age and Educational Level. Low Cash Float Subsample (N = 66).	340
TABLE G.3: Refund Tax Case Y and High Pay Tax Case X. All Males Subsample (N = 49).	341
TABLE G.4: Refund Tax Case Y and High Pay Tax Case X. All Females Subsample (N = 83).	341
TABLE G.5: Refund Tax Case Y and High Pay Tax Case X. Male Low Cash Float Subsample (N = 19).	342
TABLE G.6: Refund Tax Case Y and High Pay Tax Case X. Female Low Cash Float Subsample (N = 47).	342

TABLE OF FIGURES

Figure 2.1: The Utility Function of Money	13
Figure 2.2: The Prospect Theory Weighting Function	32
Figure 2.3: The Prospect Theory Value Function	34
Figure 2.4: Cumulative Prospect Theory: Weighting Functions for Gains and Losses	40

“We believe it will be impossible to gain substantial confidence in the factors causing different levels of compliance by using only ...archival data bases. This is because researchers can only observe these data in a non-controlled setting. This ensures confounding of variables which seemingly have sound theoretical support in explaining compliance behaviour. It appears that the experimental approach is needed in conjunction with the empirical, archival approach to eventually explain taxpayer compliance behaviour.”

Chang, O. H. and Schultz, J. J. Jr., “The Income Tax Withholding Phenomenon: Evidence from TCMP Data”, *Journal of the American Taxation Association*, Vol. 2, No. 1, Fall 1990, p. 92.



Bob realized too late that he should have never taken a number.

ABSTRACT

This study replicates and extends Dusenbury (1994). A sample of 132 New Zealand taxpayer-participants performed a laboratory experiment consisting of four decision problems relating to tax situations and two decision problems relating to other financial contexts. The focus of interest was the power of each tax context's stated prepayment position to induce a decision frame (as predicted by *Prospect Theory*) capable of determining each participant's risk preferences. The use of a *repeated measures* analysis of variance procedure (general linear model) allowed for *within-subject* comparisons so that decision-frame-induced shifts in individual participants' risk preferences across decision problems could be diagnosed. The study found, with respect to the tax problems, that the predictions of *Prospect Theory* and the findings of Dusenbury (1994) were generally supported. However, similar risk preference shifts based on alternative decision frames, independent of prepayment position and all other contextual information, were also detected.

When *within-subject* comparisons were also generated by contrasting of tax and non-tax decision problems, identical apart from context, it was found that participants produced evidence of dissimilar, context-sensitive risk preferences. Risk preferences were also found to be stable for decision problems containing identical option sets and similar, but not identical contexts. Also, a significant behavioural difference was detected between participants with high cash floats and those with low cash floats; but this was not differentiated from an *awareness of cash float status* variable in the study. Finally, while the underlying value functions of the participants conformed to the predictions of *Prospect Theory* in the tax contexts, there were some departures from the predicted form when the decision problem was set in a gambling context.

1. INTRODUCTION

1.1 STATEMENT OF THE RESEARCH QUESTIONS

The purpose of this study is to investigate the nature of the risk profiles exhibited by New Zealand taxpayers at the time of filing their annual tax returns. This will be achieved by means of a replication and extension of the laboratory experiment undertaken in the United States of America by Dusenbury (1994)¹ and therefore involves two goals of a more specific nature. The first of these is to provide independent confirmation of Dusenbury's findings in terms of a conceptual replication undertaken on an unrelated sample, in a different country, in a different year. The second goal is to furnish a cross-cultural comparison between the taxpayer-participants of Dusenbury's North American study and their equivalents in this New Zealand study. The risk characteristics exhibited by the North American participants, when they filed a set of experimental returns, are endorsed by the discovery of similar characteristics in New Zealand taxpayer-participants under similar experimental conditions. With these goals in mind, the formal research problem in this New Zealand study, is therefore substantially the same as the one implied in Dusenbury's research paper's title which was²:

The Effect of Prepayment Position on Individual Taxpayers' Preferences
for Risky Tax-Filing Options.

Let Dusenbury's title be restated as a question. It asks, do taxpayers make changes to the degree of risk to which they are willing to expose themselves when making different types of tax-reporting decisions? In particular, do taxpayers take riskier courses of action when their decisions are made in the context of owing the tax authorities a further sizeable tax payment above and beyond taxes already paid over in some form during the year? Conversely, are taxpayers more risk averse when their decisions are made in the context of a refund owed to them as a result of an overwithholding of taxes by the tax authorities earlier in the year?

¹ Dusenbury, R., (1994), "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options".

² Ibid, p. 1.

Embedded in these two forms of the question there is a deeper one. Do decision frames exist to the extent that they may be held uniquely responsible for a detectable reversal of risk preferences in decision problems which are identical in all respects other than their context? Decision frames were posited to exist by Kahneman and Tversky (1979),³ in a ground-breaking paper which introduced *Prospect Theory* as an alternative descriptive theory of decision-making under risk to the prevailing paradigm based on the normatively pure, but descriptively inadequate *Expected Utility Theory*. One of *Prospect Theory's* central tenets was the prediction that decision makers will reverse their preferences from risk aversion in a context entailing perceptions of a gain, to risk willingness when the presented context entails perceptions of a loss.

Like Dusenbury's (1994) study, the current study addresses the issue of the existence of decision frames in taxpaying contexts with a view to determining what impact, if any, these frames may have on taxpayers' compliance with their monetary obligations to the state.

Both Dusenbury (1994) and the current study expand this issue into three research questions. The primary research question concerns the existence of a decision frame in the context of computing one's annual tax return. This frame is posited to base its neutral reference point on the level of tax already received from the taxpayer before the time the taxpayer calculates his or her annual income, and from this figure, his or her final tax balance refundable or payable. A tax refund is viewable as a gain; and tax payable is viewable as a loss.

The primary research question is stated as follows. In the New Zealand context, what effect do prior tax withholdings such as PAYE or provisional tax have on taxpayers' propensities for risk taking when reporting annual income and determining monies refundable from, or payable to the taxation authority? The question is addressed in this study, as in Dusenbury (1994), in terms of an examination of a pair of decision problems in which the stated levels of prior tax withholdings produce for the taxpayer-participant a choice among various sizes of refund in

³ Kahneman, D. and Tversky, A., (1979), "Prospect Theory: An Analysis of Decision under Risk".

one instance, and in the other, a choice among options which all entail payments of terminal tax, again ranging in size.

This contrast of paired decision problems was the primary focus of Dusenbury's North American study; and, in modified form, the same two tax-reporting simulations are used in the current study. The term simulation is appropriate in this context because both Dusenbury's study *was*, and this New Zealand undertaking *is*, a social science laboratory experiment in which a sample of members of the public were given floats of a phoney currency (hereafter to be known as pseudo-dollars), and required to make compliance decisions based on the given contextual information, their pseudo-dollar holdings and on the nature of their individual propensities for risk. The problem faced by the participants was that only the least rewarding option in each set of proffered alternatives involved no risk of a challenge regarding the accuracy of the reported taxable income; and each of the risky options entailed the possible obligation to pay a greater monetary amount later in the experiment.

Dusenbury (1994) investigated two secondary research questions; each of which was tied to the first by the fact that the same range of tax decision problems were used in addressing them. These questions sought to determine whether choices amongst financial commitments presented in sets of options, identical in terms of risk and return but embedded in markedly different contexts, would induce significantly different choices of preferred risk in the taxpayer-participants. One of these research questions considered potential risk preference differences between responses to a tax decision problem and responses to the identical option set disguised as a medical insurance problem. The other question investigated response differences between a decision problem pair containing identical risk and returns described in one instance as a tax return filing problem, and in the other, a pure gamble. Both of these research questions, and the hypotheses by which they could be evaluated, were adopted for conceptual replication in the current study.

The three research questions so far discussed comprise the replication aspect of this New Zealand-based tax compliance experiment. It is a *conceptual* rather than *pure* replication because the methods used differ from those used by Dusenbury in several significant ways which are explained in detail in Chapter Four. However, to these three initial investigations, a

further four avenues of research interest have been added. These four further research questions comprise the extension aspect of this study.

The first of these extension research questions (research question four) enquires into the impact on taxpayer-participants' changes of risk preference wrought by different levels of the pseudo-money cash float. The quantity of pseudo-dollars controlled by a participant, whether uniform for all participants or varied, constitutes a variable which may have the power to exacerbate or dampen risk preference shifts brought about in an experimental situation by the presentation of taxation contexts containing contrasting loss perceptions and gain perceptions.

The fifth research question investigates a second social science laboratory experiment-related issue. If the taxpayer-participants are required to perform tasks in accordance with a set of instructions which may or may not be perceived as necessary reading by the participants, how does an experimenter know that the results he or she obtains emanate from the causes to which he or she ascribes them? Dusenbury (1994) took a number of steps to ensure that his taxpayer-subjects' responses could be traced to differences in decision frames inherent in his three tax and two non-tax decision problems. However, these may not have been sufficient to remove all doubt about the nature of the underlying decision frames actually acted upon.

In this New Zealand study, the participants were asked, in what were called two *additional ratings* questions, the extent to which they had based their choices on the putatively most salient information item provided in the tax contexts (the prepaid tax position). The preference shifts of the subsample of respondents who stated they had discounted this information item (or even ignored it) could be compared with changes in the risk preferences of the remainder of the participants, who stated that this information item had been of significant use to them. This issue was termed the Summary Syndrome extension on the ground that it examines the responses of the participants who relied solely on the list of alternative risks and returns displayed in tabular form at the end of each decision problem by way of a problem summation.

The sixth research question is of a more general nature. According to *Prospect Theory*, decision makers will switch between risk aversion and risk seeking behaviour depending on the nature of the perceptions they generate from decision frames inherent in a decision problem's

context. The question is, will taxpayer participants maintain consistent risk positions across contexts of a similar nature when these are presented in a sequence containing intervening decision problems with dissimilar contexts?

The seventh research question investigates the nature of the value functions underlying the taxpayer-participants selection of choices in each of the five replication decision problems. In these five cases, the participants were asked to reveal their second, third, fourth and fifth choices so that the smoothness of their value functions could be assessed. This investigation was prompted by the prediction made by *Prospect Theory* that decision makers will have a smooth value function curved in an S-shape which is convex to the origin in the loss domain (denoting risk willingness) and concave to the origin in the gain domain (denoting risk aversion).

1.2 RELEVANCE OF THE RESEARCH

Any study of the impact on taxpayers' risk preferences wrought by the withholding of taxes at source, or by the requirement of periodic instalments in advance, will be of value to governments reliant on taxation as the main source of income. This was true for Dusenbury (1994) since, as early as 1981, withholdings accounted for 90 percent of the annual total tax liability in the United States (\$US261 billion out of a total of \$US291 billion for that year).⁴

It is also true of this study with respect to New Zealand. This is because New Zealand has increased its reliance on the collection of prepaid taxes in recent years to the extent that virtually all taxpayers now come under one of the three arrangements in place for the levying of tax in advance of the end of the financial year. At present, taxpayers who do not have PAYE (Pay As You Earn) deducted from their wage or salary payments before they receive them⁵ are subject, with few exceptions, to Subpart B of Part M of the Income Tax Act 1994, which prescribes income earners' obligations to pay what is known as *provisional tax*, usually at the

⁴ Peckman, J. A., (1983), *Federal Tax Policy*, The Brookings Institute, Washington. This paper was cited by Yaniv, G., (1988), "Withholding and Non-Withheld Tax Evasion", p. 183.

⁵ See Part N of the Income Tax Act 1994.

rate of 105 percent of the preceding year's tax payment on declared taxable income, broken into three instalments.⁶ Furthermore, New Zealand, unlike the United States, has a comprehensive system in place for the collection of tax on interest and dividend income directly from institutions before they disburse such income to its taxpayer recipients.⁷

With respect to New Zealand's Resident Withholding Tax alone, Hasseldine, Kaplan and Fuller (1994)⁸ noted that at the time of its introduction in 1989, approximately \$NZ200 million of tax revenues were thought to have been lost each year as a result of non-declaration of income from interest and dividends. The increase of tax revenues resulting from the imposition of this collection arrangement caused the estimate to be revised upward to \$NZ400 million. In an economy as small as that of New Zealand, these are significant figures.

The upshot of this increased reliance on PAYE, Provisional Tax and Resident Withholding Tax prepayment obligations is that the New Zealand Government has a vested interest in ascertaining that these collections are levied at a rate high enough to induce framing effects in the minds of taxpayers, based on perceptions of gain rather than loss at the time annual tax returns are submitted to the Inland Revenue Department. This study, as did Dusenbury (1994) before it, provides evidence that taxpayers (as represented by the 132 New Zealand taxpayer-participants selected by non-random means⁹) are susceptible to framing effects; and that where a loss context is perceived, the propensity for increased risk of non-compliance is heightened.

1.3 DEFINITION OF TERMS

It is to be noted that neither Dusenbury (1994) nor this New Zealand study enquire directly into the incidence of tax evasion. The focus is, instead, on attitudes towards risk of non-compliance

⁶ This Part contains provisions for estimation of provisional tax in a number of different circumstances.

⁷ See Part N of the Income Tax Act 1994.

⁸ Hasseldine, D. J., Kaplan, S. E. and Fuller, L. R., (1994), "Characteristics of New Zealand Tax Evaders: A Note", p. 80.

⁹ The method of selection is explained in detail in Chapter Seven.

in a tax context involving a history of prior collections. A useful point of commencement would be to determine what is meant by the term *risk*.

Risk denotes a greater than zero chance that at least one alternative event will occur in place of the planned-for event. Usually the term also conveys the connotation that the eventuating alternative occurrence may not be a desired one. Luce and Raiffa (1957)¹⁰ prescribed that *risk* exists “where each action leads to an outcome which is one of a set of possible outcomes, each of which has a known probability.” When *risk* is used in the context of decision making, its antonym is *certainty*, which denotes a zero chance of any alternative event, desirable or undesirable, occurring in the place of the predicted event. Luce and Raiffa said that *certainty* exists “when each action is known to lead without exception to a specific outcome.”¹¹

There is also an apparent synonym for *risk*, which must be distinguished from it. *Uncertainty*, according to Luce and Raiffa, exists “where a decision will lead to an outcome which is one of a set of known possible outcomes; but the probability of these outcomes is not known.”¹²

When the context is narrowed to the field of tax compliance, *risk* may be defined as the known probability that a taxpayer’s stated level of taxable income is deemed by the tax authorities to be non-compliant. Dusenbury (1994) and the current study model their taxpayer-participants’ choice preferences in terms of their willingness to accept known odds that they will be deemed non-compliant. The participants of both studies were required to answer each decision problem by making a unique choice from a set of options containing one prospect with a certain outcome and four prospects with risky outcomes. The likelihood of deemed non-compliance was set for each risky prospect, ranging from a minimum of 15 percent to a maximum of 33 percent.

This use of *risky* prospects, as distinct from *uncertain* prospects, constituted an artificial model of tax authority behaviour, since both the Internal Revenue Service in the United States of

¹⁰ Luce, R. D. and Raiffa, H., (1957), *Games and Decisions*, p. 13.

¹¹ Idem.

¹² Idem.

America and New Zealand's Inland Revenue Department manage the probability of non-compliance detection in a much more complex manner than a five-step scale would suggest. Furthermore the Inland Revenue Department does not disclose its audit rate of annual tax returns. Hence there is necessarily a trade-off in laboratory experiments set up to examine taxpayers' compliance behaviour with respect to *risk* and *uncertainty*. Reactions to *risk* are measurable; but the quantification of preference reversals based on reactions to *uncertainty* would be much more difficult to plan for in a model.

The two theories of decision making under *risk* and *uncertainty* — *Expected Utility Theory* and *Prospect Theory* — to the extent that they deal with *risk* alone, are explained in some detail in Chapter Two.

A working definition of *non-compliance*, as it applies in this study, will also be of use at this juncture. While *tax evasion* involves the deliberate intention of defrauding the taxation authority of its rightful income, *non-compliance* is a more general concept, entailing the failure to meet tax obligations as understood by the taxation authority.¹³ This failure may have resulted from either a deliberate or an honest misapprehension of the tax legislation as interpreted by the tax authorities and the courts. However, the nature of legislation in the field of taxation both in the United States and in New Zealand is so convoluted and opaque, that the boundary dividing *non-compliance* from *compliance* is not readily discernible. While *non-compliance*, by definition, includes tax evasion, honestly-intentioned non-compliance alone is considered in this study; and the moral character of the taxpayer-participants is not questioned.¹⁴

¹³ Non-compliance is defined by Kinsey (1984) as "the failure, intentional or unintentional, of taxpayers to meet their tax obligations". See Kinsey, K. A., (1984), *Survey Data on Tax Compliance: A Compendium and Review*, Taxpayer Compliance Project Working Paper 84 - 1. Chicago Illinois: American Bar Association. This report was cited in the international study, Robben, H. S. J., Webley, P., Weigel, R. H., Warneryd, K., Kinsey, K. A., Hessing, D. J., Alvira Martin, F., Elffers, H., Wahlund, R. Van Langenhove, L. Long, S. B. and Scholz, J. T., (1990), "Decision Frame and Opportunity as Determinants of Tax Cheating", p. 342.

¹⁴ Nevertheless some of the participants may have viewed their actions as intentional evasion. However, the point was made to all participants, before they began working through the decision problems, that all choices were possibly legitimate choices, as the boundary separating legitimate practice from what was unacceptable was not always clear in some instances, until ruled on by the taxation authority and perhaps even by the country's courts of law.

1.4 OVERVIEW OF THE STUDY

The study is partitioned into eleven chapters with this introduction incorporated as Chapter One. Chapter Two provides a grounding in both *Expected Utility Theory* and *Prospect Theory* to the extent required in an application of the latter in the field of tax compliance research. This is then followed by a review of the evolution of tax compliance research involving *Prospect Theory* and, more particularly, framing effects, in Chapter Three. Chapter Four then provides an in-depth analysis of the object of the replication, Dusenbury (1994); and develops this study's hypotheses and research design from that foundation.

In Chapters Five, Six and Seven the fine details of the study's method are covered. Chapter Five contains a description of the procedures necessary for the running of the laboratory experiment's many sessions; Chapter Six explains the experimental instrument; and in the third of these chapters, Chapter Seven, the sample is explained in depth.

Chapters Eight, Nine and Ten contain the study's findings. The results obtained from a statistical analysis of the taxpayer-participants' behaviour with respect to the first three research questions, which are the study's conceptual replication questions, are tabled in Chapter Eight. Then, in Chapter Nine, the nature of the taxpayer-participants is revealed in an analysis of the disclosures made by these people in the study's end-of-session questionnaire. The final of these chapters, Chapter Ten, which contains findings relating to the four research questions constituting the extension aspect of the current study.

The study is then brought to a formal close with Chapter Eleven, which contains the conclusions, some limitations and some brief comments on the significance of the findings. Chapter Eleven also offers a few thoughts about possible future undertakings in this avenue of research mapped initially by Dusenbury (1994) and studied further here.

2. PROSPECT THEORY

2.1 INTRODUCTION

Since it is a relatively new arrival in the area of tax compliance research, an in-depth description of *Prospect Theory* and its related concept, decision frames, is useful at this point. *Prospect Theory* may be viewed as one of a number of new theories of decision-making under risk and under uncertainty which have been developed in response to the shortcomings of the theory which had achieved paradigmatic status in this field of psychological research endeavour in the second half of this century. The paradigmatic theory was *Expected Utility Theory*. At present, both *Expected Utility Theory* and *Prospect Theory* command critical attention; and, in terms of competing research schools, neither has been able to claim undisputed pre-eminence. This is of importance to tax compliance research, as *Expected Utility Theory* underlies *economic deterrence* models of compliance behaviour, while *Prospect Theory* is one of the theories underpinning the use of *fiscal psychology* compliance models.

In order to explain *Prospect Theory* it is necessary to explore the differences between it and *Expected Utility Theory*. The most efficient way of doing this is to provide a potted history of the development of theories of decision under risk. Section 2.2 covers Bernoulli's early work, which was redeveloped as *Expected Utility Theory* by Von Neumann and Morgenstern as explained in Section 2.3. Section 2.3 also provides an in-depth appraisal of the flaws in this theory which *Prospect Theory*, described in Section 2.4, set out to correct. Then in Section 2.5 *Cumulative Prospect Theory* is briefly discussed. The final section, Section 2.6 is devoted to a description of decision frames and their importance to *Prospect Theory*. This section also contains a brief review of studies conducted in this area of research endeavour.

2.2 THE ECONOMIC RATIONALITY PROBLEM

Prospect Theory is one member of a set of competing theories endeavouring to explain how human beings go about the business of making economic decisions, the outcomes of which are not known in advance. Since Daniel Bernoulli published his path-breaking theory of risk measurement in 1738 there has been considerable interest in rendering decision-making under

uncertainty according to a precise logic. The underlying assumption was that there must exist such a logic if for no reason other than that human beings are assumed to be rational decision makers.

Bernoulli set out to explain why people tend to be risk averse in general, and why this risk aversion lessens as their wealth increases. He proposed that a choice among alternatives could be evaluated in terms of the mathematical expectations of the various possible outcomes associated with each option available to the decision maker. This entailed the summation of all possible outcomes each multiplied by the assessed probability of its occurrence, yielding a weighted average or an expected value. This is summarized in the notation¹:

$$\sum v(x)p(x).$$

Here x represents a material outcome or wealth state and $v(x)$ represents the value of this to a rational agent while $p(x)$ denotes the probability of x occurring.

Risk aversion is in evidence whenever a sure outcome is preferred over a gamble that has a higher or equal expected value. Bernoulli noted most people preferred a sure acquisition over the gamble involved with choosing a risky one with a higher expected value. For instance (in modern terms) a sure gain of \$500 would be chosen ahead of an 80% chance of gaining \$700. Bernoulli recognised people did not assess choices by the expected values of their outcomes; but instead that they placed a *subjective value* on each outcome and weighted this *subjective value* by its probability. Another name for this subjective value is *utility*.²

Bernoulli showed that subjective value, or utility, was a concave function of money. He noted that an increase in an individual's wealth, no matter how insignificant, would cause an increase in that person's utility inversely proportional to the level of wealth the individual already enjoyed. In other words, a poor person's utility would rise much more than the rise in utility enjoyed by a multimillionaire if each were to be given a \$100 banknote. Concavity implied

¹ Fishburn, P., (1988), *Nonlinear Preference and Utility Theory*, p. 6. Note that in summary form no subscripts have been included in the equation.

² Kahneman, D. and Tversky, A., (1984), "Choices, Values and Frames", pp. 341-342.

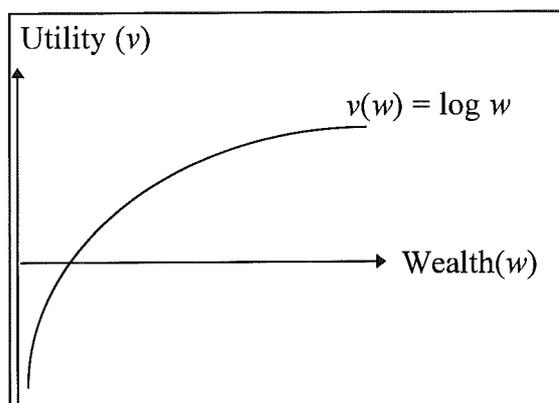
that the difference in utility between \$200 and \$100 is greater than the utility difference existing between \$1,200 and \$1,100.³ Concavity of the utility function also implied that a sure gain of \$800 would be preferred over an 80 percent likelihood of receiving \$1,000 even though the two options had the same expected value. The utility function of money is illustrated in Figure 2.1.

The problem associated with Bernoulli's theory was that subjective value was difficult to quantify. The term *utility* was adopted as the standard nomenclature for it in the second half of the nineteenth century; and it was thought of as a psychological entity, or a measure of moral worth or psychic satisfaction; but around the turn of the century this understanding changed. *Utility* was redefined by Fisher (1892), Pareto (1906) and Slutsky (1915) as simply an individual's preference hierarchy which could not be measured beyond a simple ordering and which did not allow for interpersonal comparisons.⁴

As a consequence the presence of rationality in decision-making under risk continued to be assumed; but it remained unmeasured by any workable model. A number of economists such as Frisch (1926), Lange (1934) and Alt (1936) worked on the measurability issue; but the breakthrough was made when Von Neumann and Morgenstern published their theory of

expected utility in 1944, amplifying it in 1947.⁵ This theory was enshrined as *the* standard explanation and predictor of decision-making under risk for the next forty years, remaining unchallenged by any well formulated alternative till the advent of Kahneman and Tversky's *Prospect Theory* in 1979. At the present time the relationship between these two theories could be thought

Figure 2.1: The Utility Function of Money



³ Idem.

⁴ Fishburn, P., op. cit. n. 1, pp. 3-4.

⁵ Von Neumann, J. and Morgenstern, O. (1947), *Theory of Games and Economic Behavior*.

of in terms of competing Lakatosian research programmes, neither of which has yet vanquished the other.⁶

2.3 EXPECTED UTILITY THEORY

Von Neumann and Morgenstern (1947) founded their theory of decision-making under risk, *Expected Utility Theory*, on Bernoulli's eighteenth century use of mathematics to derive the expected values of risky choices. As a consequence, the mathematical notation of their theory in its most basic form is identical with that of Bernoulli:

$$\sum u(x)p(x).$$

Von Neumann and Morgenstern's interpretation of the twin concepts of *value* and *probability* was, however, radically different.

Value was the concept which, by its intangibility, made Bernoulli's theory problematic. Von Neumann and Morgenstern by-passed this flaw in the nature of *subjective values* by making their definition of utility a monetary one. They argued that all participants in an economy think in monetary terms, and measure their material aims and aspirations in money or in the value of the goods and services it represents. It is clear they saw money as being closely associated with *value* because it is⁷:

...unrestrictedly divisible and substitutable, freely transferable and identical, even in the quantitative sense, with whatever 'satisfaction' or 'utility' is desired by each participant.

⁶ Lakatos, I. (1974), "Falsification and the Methodology of Scientific Research Programmes" in *Criticism and the Growth of Knowledge*, pp. 99 - 196. Imre Lakatos saw advances in science in terms of a struggle for pre-eminence on the part of competing schools of thought which he termed 'research programmes'. A research programme had a hard core of theory which all adherents to the school accepted as inviolate, a protective belt and a positive heuristic. The protective belt contained concepts which could be modified over time in order to shore up allegiance to the programme as a whole; and in Lakatos' own words, the positive heuristic "...consists of a partially articulated set of suggestions or hints on how to change, develop, the 'refutable variants' of the research programme, how to modify, sophisticate, the 'refutable' protective belt." (p. 135.)

⁷ Von Neumann and Morgenstern, op. cit. n. 5, p. 8.

Von Neumann and Morgenstern were thus able to argue that if differences in levels of utility were based on monetary value, then these differences could be numerically measured.

The issue of *probability* in a twentieth century context was easily resolved⁸:

Probability has often been visualised as a subjective concept more or less in the nature of an estimation. ...The simplest procedure is... to insist upon the alternative, perfectly well founded interpretation of probability as frequency in long runs. This gives directly the necessary numerical foothold.

In brief, *Expected Utility Theory* holds that decisions made in uncertain situations are a choice among gambles or ‘prospects’, where a prospect is the set of probabilities (p_1, p_2, \dots, p_n) and the set of outcomes $X = (x_1, x_2, \dots, x_n)$ associated with them⁹:

$$(p_1x_1, p_2x_2, \dots, p_nx_n),$$

Where $0 \leq p_i \leq 1$.

The probabilities of all the possible outcomes of a prospect sum to unity; and since utility is associated with weightings provided by these probabilities, the utility of a risky prospect is formally stated to be linear with respect to the probabilities of outcomes. More generally, there is a system U which is the set of expected utilities, u, v, w, \dots each of which is unique to its own prospect, say Prospect A or Prospects B or C. We can write the expected utilities of these prospects in the form:

$u(X_A) = u(A)$ where X_A is the set of outcome states possible with regard to Prospect A;

$v(X_B) = v(B)$ where X_B is the set of outcome states possible with regard to Prospect B;

$w(X_C) = w(C)$ where X_C is the set of outcome states possible with regard to Prospect C.

⁸ Ibid, p. 19.

⁹ Schisler, D. L., (1992), “An Experimental Examination of Taxpayers’ and Tax Professionals’ Aggressive Tax Preferences — A Prospect Theory Foundation”. p. 7.

If a decision maker prefers Prospect A to Prospect B, this may be expressed in the form, $A \succ B$; and this implies that the utility of A is greater than the utility of B, i.e.:

$$A \succ B \Rightarrow u(A) > v(B)$$

Or, in abbreviated form:

$$u \succ v.$$

The abbreviated terms u , v , and w will be used in the following coverage of Von Neumann and Morgenstern's axioms.

2.3.1 The Axioms, Principles and Assumptions underlying Expected Utility Theory.

In this section Von Neumann and Morgenstern's initial axioms, as laid out in their 1947 work, are briefly enumerated and explained. Where later theorists working in the same research programme have suggested alternatives to these, and the alternatives have been adopted into the programme's hard core, some of these have been mentioned as well. The importance of the axioms is that the principles of the theory are founded upon them and a significant assumption is associated with them. The section is divided into subsections, each devoted to a principle or related set of principles, or to an assumption. The reason for explaining the principles and assumptions of *Expected Utility Theory*, to the extent they are explained here, is that they are either explicitly endorsed or equally explicitly contradicted by *Prospect Theory*.

2.3.1.1 Transitivity

Von Neumann and Morgenstern's first axiom, "(A) $u > v$ is a Complete Ordering of U ," was that for any two u and v , one and only one of the following relations holds¹⁰:

$$(A:a) \quad u = v, \quad u \succ v, \quad u \prec v.$$

¹⁰ Von Neumann and Morgenstern, op. cit. n. 5, p. 26.

The equality symbol denotes indifference between the two utilities since they have the same value; and the \succ sign denotes ‘is preferred to’, while the sign \prec denotes ‘is less attractive than’. This condition states that the system of individual preferences is complete. It is important to note that even an infinitesimal difference in value will cause one prospect to be preferred to another.

An implication of the completeness axiom is that there must be transitivity of preference. Von Neumann and Morgenstern call transitivity of preference “a plausible and generally accepted property”¹¹:

$$(A:b) \quad u \succ v, \quad v \succ w, \text{ implies } u \succ w.$$

An important corollary of this is that $u \succ v$ implies that $u \prec v$ cannot occur.

2.3.1.2 Dominance

Von Neumann and Morgenstern called their second axiom group, “(B) Ordering and Combining”. There were four related axioms in it¹²:

$$(B:a) \quad u \prec v \text{ implies that } u \prec pu + (1 - p)v; \quad 0 \leq p \leq 1.$$

$$(B:b) \quad u \succ v \text{ implies that } u \succ pu + (1 - p)v; \quad 0 \leq p \leq 1.$$

Effectively axiom (B:a) states that if v is preferable to u , then a $(1 - p)$ possibility of obtaining v in place of u is preferred to the sure receipt of u . In other words, if u is assured and v is a possible replacement for it, then a decision maker will opt for the $(1 - p)$ chance of v . This underpins the principle of dominance: a rational decision maker will always maximize his or her utility when it may be done by moving from a given level of wealth to a greater one, no matter how small the difference.

Axiom (B:b) states that the converse case is also true, where u is preferred to v .

¹¹ Ibid, p. 27.

¹² Ibid p. 26.

(B:c) $u \prec w \prec v$ implies the existence of a p such that:
 $pu + (1 - pv) \prec w$; $0 \leq p \leq 1$.

(B:d) $u \succ w \succ v$ implies the existence of a p such that:
 $pu + (1 - pv) \succ w$; $0 \leq p \leq 1$.

Axiom (B:c) states that if v is preferred to w and u , then a small enough chance $(1 - p)$ of v in place of u is less attractive than sure receipt of w . But if, as in (B:d), the preference order is reversed so that u is the most preferred of the three variables, then a small chance $(1 - p)$ of v in place of u is preferred to a sure receipt of w .

2.3.1.3 Invariance

Von Neumann and Morgenstern called their third grouping, “(C) Algebra of Combining”:

(C:a) $pu + (1 - pv) = (1 - pv) + pu$, $0 \leq p \leq 1$.

This axiom states that the order in which information is presented is irrelevant to a decision maker. Rational behaviour ordains that one assess the utility associated with a choice strictly in accordance with the expected values of the constituent prospects offered in the choice. The axiom represents mathematically the principle of invariance: the order of presentation does not matter.

In order to examine the second axiom in this grouping, let α and β represent p over two different steps in a decision, each containing a chance p that an outcome will occur, and a chance $(1 - p)$ that an alternative outcome will occur. Then let γ represent some p inherent in the prospect in which the two steps are combined¹³:

(C:b) $\alpha(\beta u + (1 - \beta)v) + (1 - \alpha)v = \gamma u + (1 - \gamma)v$,
 Where $\gamma = \alpha\beta$ $0 \leq \gamma, \alpha, \beta \leq 1$.

The principle of invariance is underscored further by axiom (C:b) in which Von Neumann and Morgenstern declare the irrelevance of whether a prospect of a ‘certain chance of u and

¹³ Ibid p. 27.

otherwise v' is to be considered in two steps or as one in which the combined probabilities have been multiplied together. The rational decision maker will combine the probabilities correctly. The authors were very much aware of the significance of this statement and were moved to declare¹⁴:

The preceding analysis made it clear we feel free to make use of a numerical conception of utility. On the other hand, ...we cannot avoid the assumption that all subjects of the economy under consideration are completely informed about the physical characteristics of the situation in which they operate and are able to perform all statistical, mathematical, etc., operations which this knowledge makes possible.

They noted that this assumption had been and would continue to be given extensive attention in academia; but they proposed to ignore it on the ground that their theory explained and predicted behaviour where the information available to individuals was complete.

2.3.1.4 Substitution

Expected Utility Theory was amplified by Luce and Raiffa (1957); and the formulation of the axioms was rearranged by, among others, Jensen (1967).¹⁵ Jensen in particular provided an axiom related to Von Neumann and Morgenstern's axioms, (B:a) and (B:b). He called it the axiom of independence; but it could just as easily be called the axiom of substitution¹⁶:

$$u \succ v \Rightarrow pu + (1 - p)w \succ pv + (1 - p)w.$$

The independence axiom states that if u is preferred to v , then a non-trivial combination of u and w is preferred to a similar combination of v and w . It is closely associated with the concept of additivity and with the cancellation conditions associated with multiple-step choices. In this respect, it is used as a criterion for showing the consistency and coherency of preferences stated in axiom (C:b) in a different way. The argument is as follows.

¹⁴ Ibid, p. 30.

¹⁵ Jensen's version of the Von Neumann-Morgenstern axioms is explained in Fishburn, P., op. cit. n. 1, pp. 9 - 14.

¹⁶ Kahneman, D. and Tversky, A. (1984), op. cit. n. 2, p. 343. Kahneman and Tversky call this axiom the *axiom of substitution*.

- The prospect $pu + (1 - p)w$ can be viewed as a first step choice which produces either an outcome associated with u or an alternative outcome associated with w .
- The prospect, $pv + (1 - p)w$, can be regarded as an alternative first step choice from which one goes on to a second step.
- Both first-step prospects have a $(1 - p)$ probability of producing w . Hence the rational decision maker drops w from consideration.
- The decision maker is now free to choose a first step prospect on the strength of the preference relation existing between u and v .¹⁷

2.3.1.5 Additivity

One of the most important arguments based on the axioms of *Expected Utility Theory* is the principle of additivity: the expected total utility of a prospect is equal to the sum of the expected utilities of every outcome in the set of outcomes. In other words, the utility of the whole is equal to the sum of the utilities of the parts. Kahneman and Tversky (1979) called this the principle of expectation since the core concept here was the pre-eminence of the expected value (as weighted by probabilities) in prospect evaluation¹⁸:

$$\begin{aligned} u(X) &= u(x_1, p_1; x_2, p_2; \dots; x_n, p_n) \\ &= p_1 u(x_1) + p_2 u(x_2) + \dots + p_n u(x_n). \end{aligned}$$

2.3.1.6 Risk Aversion

A decision maker will always prefer a decision outcome (x) obtained with certainty to any risky decision outcome which has the same expected value. The utility function for money, $u(x)$ was proposed by Bernoulli to be the concave function shown in Figure 2.1, Subsection 2.2, above. Kahneman and Tversky (1979), citing Pratt (1964)¹⁹ and Arrow (1971),²⁰ state that in *Expected*

¹⁷ Fishburn, op. cit. n. 1, p. 17.

¹⁸ Kahneman and Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk", p. 263.

¹⁹ Pratt, J. W., (1964), "Risk Aversion in the Small and in the Large."

Utility Theory, risk aversion is equivalent to the concavity of this utility function. The function $u(x)$ is concave when its second differential is less than zero ($u'' < 0$); and Pratt provided a formula by which the degree of risk aversion could be quantified: $-u''(x)/u'(x)$.

2.3.1.7 Asset Integration

Risk aversion, as the natural state of rational decision makers in *Expected Utility Theory*, requires that one assumes that $u(x)$ is a function of total assets rather than of changes which may result from a decision. Consequently $x = 0$ equates with ruin or the loss of all disposable assets.²¹ A prospect will only be acceptable to a decision maker if the utility of adding the prospect to his/her asset holdings is greater than the utility of the asset holdings alone. In other words decision makers think in terms of levels of wealth. The assumption was formulated as a principle and described in the following notation²²:

$(x_1, p_1; \dots; x_n, p_n)$ is acceptable at asset position w
if and only if $u(w + x_1, p_1; \dots; w + x_n, p_n) > u(w)$.

Kahneman and Tversky (1979) referred to this principle as the *tenet of asset integration*.²³

2.3.2 Why Expected Utility Theory is Inadequate.

Although the use of outcomes weighted by their probabilities to determine utility was normatively satisfying, the fact that *Expected Utility Theory* furnished explanations and predictions at variance with what was observed in practice moved a number of researchers to determine the nature of that practice. From this it was intended that a positive theory of decision-making would be developed, which would be *descriptively*, as distinct from *normatively*, satisfying.²⁴ Among the earliest of the critics was Allais (1953)²⁵ who devised a

²⁰ Arrow, K. J., (1971), *Essays in the Theory of Risk Bearing*.

²¹ Pratt, (1964), op. cit. n. 19, p. 123.

²² Kahneman and Tversky (1979), op cit. n. 18, p. 264.

²³ Idem.

²⁴ Mario Bunge defined the distinction between *normative* and *descriptive* in the following terms: "Decision theory, game theory and other modern theories of human behavior are often described as *normative* or prescriptive, rather than *descriptive*, because they do not represent actual behavior but an ideally correct behavior in certain areas — much as logic, which sets

series of scenarios in which participants were found to make decisions that did not maximise utility in accordance with *Expected Utility Theory*. A second major critic was the economist and Nobel Prize winner, Herbert Simon (1955, 1956, 1976, 1978), who maintained that the human brain was not capable of meeting the computational requirements of any but the simplest of calculations. Human thinking, instead, was serial in its organisation and relied on memory of past experience to provide a decision template.²⁶ However, it is the work of Amos Tversky and Daniel Kahneman which is of most immediate relevance to this thesis.

Kahneman and Tversky launched a major attack on *Expected Utility Theory*'s axioms and principles — an attack which simultaneously promulgated their descriptive theory of decision-making behaviour, *Prospect Theory*.²⁷ The following subsections describe Kahneman and Tversky's arguments without necessarily adhering to the terminology used in their 1979 paper since the authors changed some of it in ensuing papers. Instead, an attempt is made to relate their attack to the axioms and principles of *Expected Utility Theory* as formulated by Von Neumann and Morgenstern, and their principle elucidators. The central importance of additivity, alias the principle of expectation, to actual decision-making processes is shown in all of the following subsections to be misconceived. The first principle to be discredited was the corollary to axiom (A:b) — the principle of transitivity of preference.

2.3.2.1 Transitivity of Preference and Certainty

Transitivity of preference, as laid out in the corollary to Von Neumann and Morgenstern's axiom (A:b), was shown to be descriptively incorrect by Kahneman and Tversky (1979) in a number of one- and two-step decision problems given to students and staff at universities in

standards for deductive inference but does not account for actual reasoning, a problem left to the psychology of thinking." Bunge, M., (1967), *Scientific Research I: The Search for System*, pp. 390 - 391.

²⁵ Allais, M., (1953), "Le comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine" pp. 503 - 546.

²⁶ Simon, H., (1976), "From Substantive to Procedural Rationality", pp. 135 - 137. Simon wrote a series of papers on the theme, the first of which, "A Behavioral Model of Rational Choice" appeared in 1955. In his 1956 paper, "Rational Choice and the Structure of the Environment", Simon coined the term *satisficing* to cover the idea of increasing one's utility by use of approximations rather than exact calculations, and by use of rules of thumb.

²⁷ Kahneman and Tversky, (1979), op. cit. n.18, pp. 263 - 291.

Israel, Sweden and in the United States of America. In the exposition of the initial pair of problems, which follows, demonstrating violation of transitivity, the total number of participants is denoted N , and the percentage who chose each option is bracketed. Problem 1 illustrates in particular the power of what Kahneman and Tversky (1979) call the certainty effect.²⁸:

Problem 1 ($N = 72$)

Choose between:

Prospect **A**: \$2,500 with probability .33, \$2,400 with probability .66,
and nothing with probability .01; (18%);

and

Prospect **B**: \$2,400 with certainty; (82%).

A certainty effect exists when decision makers choose prospects promising outcomes with certainty but with lower expected values than those of alternative prospects with risky outcomes. According to the evidence in Problem 1, Prospect **B** was preferred by 82 percent of Kahneman and Tversky's subjects even though it had an expected value of \$2,400 versus Prospect **A**'s slightly higher expected value of \$2,409 in conjunction with an almost negligible (1%) chance of no gain at all. If Prospect **A** is rearranged it becomes clear there was actually a 99 percent probability that at least \$2,400 would be received and the probability that the decision maker would receive an extra \$100 stood at 33 percent; yet Prospect **B** was more widely chosen.

This result also violates Von Neumann and Morgenstern's axiom (B:a). This is evident in that if Prospect **B** (\$2,400) represents u in the axiom, and if v represents the utility of \$2,500; and $u < v$; then Prospect **A** could be stated in the form:

$$p(u) + (1 - (p - .01))v + .01(0);$$

Where $u \succ v$.

²⁸ Ibid, Problems 1 and 2, pp. 265 - 266.

2.3.2.2 Transitivity and Reduction from Certainty

Combining Problem 1 with Problem 2 (below), Kahneman and Tversky also showed that transitivity of preferences specifically did not hold when the probability of receiving a sum of money was reduced from certainty to some lesser probability.

When a 66 percent probability of \$2,400 was subtracted uniformly from both **A** and **B**, and the two prospects were repackaged as Prospects **C** and **D** in Problem 2, the experimental subjects reversed their original preference order:

Problem 2 ($N = 72$)

Choose between:

Prospect **C**: \$2,500 with probability .33
and nothing with probability .67; (83%);

and

Prospect **D**: \$2,400 with probability .34
and nothing with probability .66; (17%).

Here Prospect **C**, with an expected value of \$825, was a much more popular choice than **D** with an expected value of \$816. This reversal of preference, given that **C** is a scaled version of **A**, and **D** is a scaled version of **B**, violates the principle of additivity and the axioms underlying the principle of transitivity.

Kahneman and Tversky further illustrated the phenomenon in the following pair of problems, showing that the axiom of substitution was also violated²⁹:

Problem 3 ($N = 95$)

Choose between

Prospect **E**: an 80% chance of \$4,000 and 20% chance of nothing; (20%);

and

Prospect **F**: the receipt of \$3,000 with certainty; (80%).

Problem 4: ($N = 95$)

Choose between

Prospect **G**: a 20% chance of \$4,000; (65%);

and

Prospect **H**: a 25% chance of \$3,000; (35%).

²⁹ Ibid, Problems 3 and 4, p. 266.

In Problems 3 and 4 the substitution axiom was violated since the reduction from 100 percent to 25 percent in the probability of winning between Prospects **F** and **H** was shown empirically to have a greater effect than the reduction from 80 percent to 20 percent between Prospects **E** and **G**.³⁰

2.3.2.3 Transitivity and the Reflection Effect

Kahneman and Tversky (1979) also showed that transitivity of preference was violated when decision makers switched from perceptions of positive dollar amounts to perceptions of negative dollar amounts. Problem 3 above was matched with its mirror image in which the outcome in dollars represented losses:

Problem 5: ($N = 95$)

Prospect **I**: an 80% chance of -\$4,000 and a 20% chance of nothing; (92%);

and

Prospect **J**: -\$3,000 with certainty; (8%).

The percentage of participants choosing Prospect **I** runs strongly counter to the percentage choosing Prospect **E** in Problem 3. This indicates that the risk averse behaviour predicted by *Expected Utility Theory*, with respect to all decisions governing possible changes in wealth, does occur for outcomes which are viewed as gains, but is conspicuously absent when potential outcomes are viewed as losses. In this context 92 percent of the participants provided evidence of risk seeking behaviour. Transitivity of preference is violated here because Prospect **I** has a larger negative expected value (- \$3,200) than that of Prospect **J** (- \$3,000) according to the principle of additivity.

Kahneman and Tversky (1979) also argued that the reflection effect eliminates risk aversion in the face of uncertainty as an explanation of the certainty effect in the above empirically demonstrated preference for a prospect with a more deeply negative expected value than a

³⁰ Idem.

prospect offering a negative outcome with certainty. They noted that certainty increases aversion for losses as well as the desirability for gains.³¹

2.3.2.4 Transitivity and the Possibility Effect

Kahneman and Tversky (1979) also showed that the magnitude of a probability will affect perceptions of the utility of a prospect and bring about a violation of transitivity. They demonstrated this with a pair of problems. In Problem 6 there was a choice between prospects in which the probabilities of receiving cash were either mid-range or very high. In Problem 7 decision makers were required to choose between probabilities that were very low; but in both problems, one prospect had double the probability of the other:

Problem 6 ($N = 66$)

Prospect **K**: A 45% probability of receiving \$6,000 and a 55% probability of nothing; (14%);
and

Prospect **L**: A 90% probability of receiving \$3,000 and a 10% probability of nothing; (86%).

Problem 7: ($N = 66$)

Prospect **M**: A .001 probability of \$6,000 and a .999 probability of nothing; (73%);
and

Prospect **N**: A .002 probability of \$3,000 and a .998 probability of nothing; (27%).

2.3.2.5 Invariance and the Isolation Effect

The principle of invariance laid out in Von Neumann and Morgenstern's axiom (C:b) — and, by extension, the principle of intransitivity of preference — were shown to be descriptively incorrect when Kahneman and Tversky (1979) obtained the percentage results associated with the prospects attached to the scenario in Problem 8.³² Kahneman and Tversky called the phenomenon they found the *isolation effect* since decision makers apparently *isolated* what they considered to be the salient elements and jettisoned the rest.

³¹ Ibid, p. 269.

³² Ibid, Problem 10, p. 271.

Problem 8 ($N = 141$):

Consider the following two-stage game in which your choice must be made before the first stage outcome is known. In the first stage, there is a probability of .75 to end the game without winning anything, and a probability of .25 to move into the second stage. If you reach the second stage you have a choice between:

Prospect **O**: an 80% chance of \$4,000 and 20% chance of nothing; (22%);
 and
 Prospect **P**: the receipt of \$3,000 with certainty; (78%).

It is clear that, because the outcomes of the first stage were common to both prospects, most of Kahneman and Tversky's subjects ignored it and viewed the problem as a simple choice between an 80 percent chance of \$4,000 and certain receipt of \$3,000. These results conform with the findings associated with the one-step problem with these exact dimensions in Problem 3 above;³³ and in so doing, they violate Von Neumann and Morgenstern's axiom (C:b). Kahneman and Tversky were moved by these findings to state³⁴:

The reversal of preferences due to the dependency among events is particularly significant because it violates the basic supposition of a decision-theoretical analysis, that choices between prospects are determined solely by the probabilities of final states.

The actual probabilities associated with Prospects **O** and **P** in Problem 8, once one incorporates the first stage in accordance with *Expected Utility Theory*, are as laid out in the prospects of Problem 4 above. Specifically, a .25 probability of an 80 percent chance is $.25 \times .8 = .2$; and a .25 probability of certainty is .25.³⁵

2.3.2.6 Coda

Given that the principle of dominance is grounded in the axioms supporting transitivity of preference, Tversky and Kahneman (1986) sum up the general shortcomings of *Expected*

³³ Ibid, Problem 3, p. 266.

³⁴ Ibid, p. 272.

³⁵ Idem, Problem 4.

Utility Theory in terms which provide an echo of the theme of research programmes in conflict³⁶:

Because invariance and dominance are normatively essential and descriptively invalid, a theory of rational decision cannot provide an adequate description of choice behavior.

2.4 PROSPECT THEORY

Kahneman and Tversky (1979) proposed *Prospect Theory* as an explanation of the phenomena their laboratory testing had made apparent. The theory is descriptive where *Expected Utility Theory* is normative. *Prospect Theory* strongly focuses on what human decision makers actually do. According to *Prospect Theory*, we partition the job of deciding among prospects into two discrete, sequential parts — an editing phase and an evaluation phase. These will be explained in turn in the following subsections. An amended form of *Prospect Theory* proposed by Tversky and Kahneman (1992) is described in Section 2.5.

2.4.1 Part One: The Editing Phase

According to both forms of *Prospect Theory*, the editing phase occurs first. In this phase, the decision maker perceives the problem involving risk or uncertainty which is to be solved and formulates it in a format which facilitates evaluation. There are six possible editing operations. Of these the first, coding, is of particular significance to the laboratory experiment undertaken in this study.

2.4.1.1 Coding

Individuals do not perceive their decisions in terms of a final wealth position or holding, but in terms of a loss or a gain. This is a significant departure from *Expected Utility Theory's* principle of asset integration. The coding of a decision problem under risk or uncertainty into a

³⁶ Tversky and Kahneman, (1986), "Rational Choice and the Framing of Decisions", p. s254.

gain context or a loss context is of major importance in the ensuing evaluation process, as risk preferences in gain contexts tend to conform with the risk aversion predicted by *Expected Utility Theory*, while in loss contexts, they become risk seeking in nature.

The coding of a decision problem into gains or losses involves a conscious or subconscious sorting in terms of a neutral reference point. We can think of this point as the origin in two-dimensional space where the vertical axis measures *perceived value* and the horizontal axis measures the *loss-gain continuum*. In monetary terms it is the zero point where no dollars change hands as neither a gain nor a loss will be sustained. It could also be thought of as the status quo. The reference point is not stable across decision problems of different sorts; nor is it stable over time. It shifts with the decision maker's perceptions. Kahneman and Tversky (1979) state³⁷:

...the location of the reference point, and the consequent coding of outcomes as gains or losses can be affected by the (1) formulation of the offered prospects, and (2) by the expectations of the decision maker.

Tversky and Kahneman (1981)³⁸ adopted the term *decision frame*:

...to refer to the decision-maker's conception of the acts, outcomes, and contingencies associated with a particular choice.

In other words, a decision frame is related to the decision problem's context. As the concept of coding in terms of a decision frame is central both to *Prospect Theory* and to this study, it is explored in greater detail in Section 2.6.

³⁷ Kahneman and Tversky (1979), op. cit. n. 18, p. 274.

³⁸ Tversky and Kahneman (1981), "The Framing of Decisions and the Psychology of Choice", p. 453.

2.4.1.2 Combination

This is the act of combining probabilities with identical outcomes in order to simplify a prospect. For instance a prospect offering a 30 percent chance of \$100 and another 30 percent chance of \$100 would be combined as a 60 percent chance of \$100.

2.4.1.3 Segregation

This is the act of separating prospects into a riskless component and a risky component. Consider the following prospect:

Prospect **Q**: An 80% probability of receiving \$300 and a 20% probability of receiving \$500.

Prospect **Q** would be segregated and restated as:

Prospect **Q'**: Sure receipt of \$300 and a 20% probability of receiving \$200 more.

2.4.1.4 Cancellation

Individuals ignore elements of a decision that are common to all of the prospects under consideration. An example of this can be seen in the two-step problem posed by Kahneman and Tversky (1979), which was printed as Problem 8 in Subsection 2.3.2.5 above. In this problem the first step was ignored by 78 percent of Kahneman and Tversky's experimental subjects.

2.4.1.5 Simplification

Decision makers round off probabilities and outcomes to easier-to-deal-with figures. A 52 percent probability is treated like a 50 percent probability. Kahneman and Tversky (1979) observe that a particularly important form of simplification involves the jettisoning of outcomes with extremely low probabilities from consideration.³⁹

³⁹ Kahneman and Tversky (1979), op. cit. n. 18, p. 275.

2.4.2 Part Two: The Evaluation Phase

In this phase the decision maker chooses the prospect with the highest value. The value of prospect is denoted by V (note that it is capitalised). There are two functions — a *weighting function* and a *value function*.

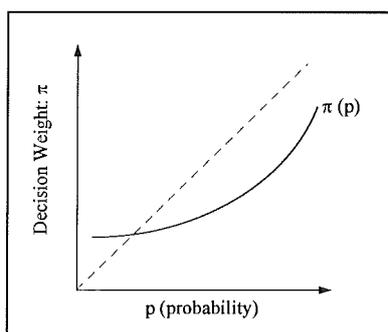
2.4.2.1 Weighting Function

In this function the probability (p) associated with an outcome is adjusted by a weighting factor (π) to provide a decision weight, $\pi(p)$. This decision weight reflects how heavily the decision maker subscribes to the probability that a prospect will deliver the desired outcome. This π factor is entirely subjective. Kahneman and Tversky (1979) point out⁴¹:

In *Prospect Theory*, the value of each outcome is multiplied by a decision weight. Decision weights are inferred from choices between prospects much as subjective probabilities are inferred from preferences... However, decision weights are not probabilities: they do not obey the probability axioms and they should not be interpreted as measures of degree or belief.

Salient here is the fact that decision weights are not required to sum to unity; and $\pi(p) + \pi(1-p)$ is typically less than unity.⁴²

Figure 2.2: The Weighting Function



The weighting function $\pi(p)$ follows the curve in Figure 2.2. The function's non-linearity is in direct contrast with the linearity of decision weighting in *Expected Utility Theory* based on probabilities alone. The 45 degree dashed line in the figure represents the *Expected Utility Theory* case.

Kahneman and Tversky noted that while extremely

⁴¹ Kahneman and Tversky (1979), op. cit. n. 18, p. 280.

⁴² Ibid p. 275.

low probabilities tend to be overweighted, extremely high probabilities are treated as if they were 100 percent certain. Hence the weighting function in the diagram has an upward slope. The phenomenon of the overweighting of low probabilities was revisited in their 1984 paper⁴³:

It enhances the value of long shots and amplifies the aversiveness of a small chance of severe loss. Consequently, people are often risk seeking in dealing with improbable gains and risk averse in dealing with unlikely losses. Thus, the characteristics of decision weights contribute to the attractiveness of both lottery tickets and insurance policies.

The overweighting of low probabilities is known as the subadditivity effect; and the treatment of extremely high probabilities as if they were certain is known as the pseudocertainty effect.

2.4.2.2 Value Function

The value function measures the value v (not capitalised) of an outcome (x), written $v(x)$. This $v(x)$ is a measure of the subjective value the decision maker places on outcome x . The value function has three properties, all of which relate to the perceived decision frame:

1. The value function is defined relative to a reference point. Kahneman and Tversky (1979) noted⁴⁴:

...outcomes are defined relative to a reference point, which serves as the zero point of the value scale. Hence v measures the value of deviations from that reference point, ie., gains and losses.

2. The value function is normally concave to the origin above the reference point, indicating that when facing a choice among prospects involving a gain decision makers are risk averse. Below the reference point the curve becomes convex to the origin, indicating that decision makers become risk takers when faced with a choice among prospects involving losses.

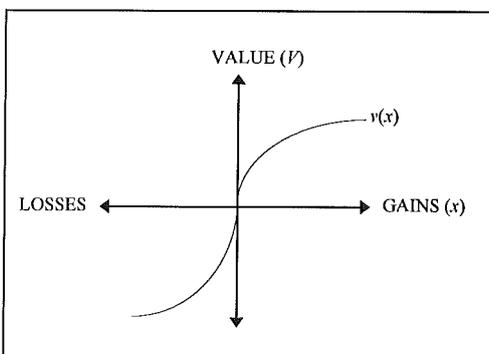
⁴³ Kahneman and Tversky, (1984), op. cit. n. 2, p. 345.

⁴⁴ Kahneman and Tversky, (1979), op cit n. 18, p 275.

3. The value function is usually steeper for losses than for gains, which is an indication that a loss will be regarded as more highly significant than a gain of the same absolute value. Tversky and Kahneman (1986) used the term *loss aversion* to denote this and provided the following amplification⁴⁵:

...[T]he response to losses is more extreme than the response to gains. The common reluctance to accept a fair bet on the toss of a coin suggests that the displeasure of losing a sum of money exceeds the pleasure of winning the same amount.

Figure 2.3: The Value Function



The value function $v(x)$ is graphed in Figure 2.3. It is to be noted in Figure 2.3 that the function shows diminishing marginal perceived value for both large gains and large losses.

The validity of this convex-concave combination was independently confirmed by Fishburn and Kochenberger (1979).⁴⁶ They found it to be the most common utility function

for people assessing changes in wealth position or returns on investment.

2.4.3 The Equations of Prospect Theory

The weighting and value functions are combined in the overall evaluation function, V , which has a basic form covering prospects offering mixed gains and losses and a secondary form which covers prospects offering outcomes which are clearly either in the loss domain of the decision frame, or in the gain domain. The basic form is considered first. The explanation is set up in the format used on decision problems already surveyed in this chapter. The variables have been rendered in bold for ease of reading:

Prospect R: A probability \mathbf{p} of receiving \mathbf{x} and a probability \mathbf{q} of receiving \mathbf{y} .

⁴⁵ Tversky and Kahneman, (1986), op. cit. n. 36, p. s258.

⁴⁶ Fishburn, P. C. and Kochenberger, G. A., (1979), "Two-piece Von Neumann - Morgenstern Utility Functions".

The basic form *Prospect Theory* equation is⁴⁷:

$$V(x, p; y, q) = \pi(p)v(x) + \pi(q)v(y),$$

where $p + q < 1$, and $x \geq 0 \geq y$, or $x \leq 0 \leq y$.

The secondary form applies to the evaluation of strictly positive and strictly negative prospects. Kahneman and Tversky note that prospects of this sort are segregated in the editing phase into riskless and risky components as demonstrated in Prospect **Q'** in subsection 2.4.1.3. If $p + q = 1$ and either $x > y > 0$, or $x < y < 0$, then⁴⁸ :

$$V(x, p; y, q) = v(y) + \pi(p)[v(x) - v(y)] \quad .$$

The salient feature of the secondary equation is that the decision weight modifies only the risky component, $[v(x) - v(y)]$, identified in the segregation operation.

2.4.4 Empirical Corroboration

A large number of studies have been conducted in the years since 1979 into the properties of *Prospect Theory* and its siblings. (Other theories developed to account for the shortcomings of the *normatively* impeccable, but *descriptively* flawed *Expected Utility Theory* included Sugden and Loomes' (1982)⁴⁹ *Regret Theory*, Fishburn's (1982)⁵⁰ *Skew Symmetric Bilinear Utility Theory*, and Machina's (1982)⁵¹ *Local Expected Utility Theory*.) While this body of research is too large and diverse to do proper justice to here, several papers warrant a by-line. While Hershey and Schoemaker (1980)⁵² found little conclusive evidence to support the property of reflection, Budescu and Weiss (1987)⁵³ found strong support for both reflection and the S-shape

⁴⁷ Kahneman and Tversky, (1979), op. cit. n. 18 , p. 276. Note that π has been defined in Subsection 2.4.2.1 of the current chapter.

⁴⁸ Idem.

⁴⁹ Loomes, G. and Sugden, R., (1982), "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty".

⁵⁰ For a full treatment of this, see Peter C. Fishburn's 1988 book, *Nonlinear Preference and Utility Theory*.

⁵¹ An excellent summary of developments in the entire field of research on choice under uncertainty is contained in Mark J. Machina's 1987 paper, "Choice Under Uncertainty: Problems Solved and Unsolved".

⁵² Hershey, J. C. and Schoemaker, P. J. H., (1980), "Prospect Theory's Reflection Hypothesis: A Critical Examination".

⁵³ Budescu, D. V. and Weiss, W., (1987), "Reflection of Transitive and Intransitive Preferences: A Test of Prospect Theory".

of the value function. In particular, Budescu and Weiss found that most of their subjects who appeared to conform with the principle of transitivity had a preference ranking for negative gambles that was an exact mirror image of their ranking of positive gambles comparable in absolute value. However, Schneider and Lopes (1986)⁵⁴ found that *Prospect Theory* was inadequate to describe individuals' risk preferences over a full range of lottery types.

Zahedi (1986)⁵⁵ extended investigation of decision-making under uncertainty to consider the mechanisms of group consensus; and in the same general area, McGuire, Kiesler and Siegel (1987)⁵⁶ found strong support for *Prospect Theory* in that groups involved with face-to-face mediation were risk averse for gains and risk seeking for losses. On the other hand, Kameda and Davis (1990)⁵⁷ found that groups with majorities that did not perceive situations in terms of losses tended to override minority members who did perceive losses. Further *Prospect Theory* paradigm research into group behaviour was done by Whyte (1993).

Kanto, Rosenqvist and Suvas (1992)⁵⁸ found that behaviour at racetracks in Finland supported the possibility effect impounded in *Prospect Theory* in that the Finnish punters showed a systematic long-shot bias in favouring of horses saddled with low odds of winning.

More relevant, however, is a sequence of papers exploring decision-making under risk and/or *Prospect Theory* with specific reference to the coding function phenomenon of decision frames. These papers are touched upon in Subsection 2.6.3.

⁵⁴ Schneider, S. L. and Lopes, L. L., (1986), "Reflection in Preference under Risk: Who and When May Suggest Why".

⁵⁵ Zahedi, F., (1986), "Group Consensus Function Estimation when Preferences Are Uncertain".

⁵⁶ McGuire, T. W., Kiesler, S. and Siegel, J., (1987), "Group and Computer-Mediated Discussion Effects in Risk Decision Making".

⁵⁷ Kameda, T. and Davis, J. H., (1990), "The Function of the Reference Point in Individual and Group Risk Decision Making".

⁵⁸ Kanto, A. J., Rosenqvist, G. and Suvas, A., (1992), "On Utility Function Estimation of Racetrack Bettors".

2.5 CUMULATIVE PROSPECT THEORY

2.5.1 The Cumulative Prospect Theory Value Function

Tversky and Kahneman (1992) abandoned the weighting function set out in their 1979 paper, criticising it on two grounds — it failed to satisfy the principle of dominance in all instances (relying on people's ability to perceive dominated prospects and drop them from consideration during the editing phase); and it was not readily extended to prospects involving a large number of possible outcomes. Tversky and Kahneman proposed, instead, that a cumulative distribution function be applied separately to gains and losses. This would enable the theory to be extended to explain and predict uncertain as well as risky prospects incorporating large numbers of outcomes.⁵⁹

In *Cumulative Prospect Theory* risk aversion and risk seeking are determined jointly by a value function and by a *cumulative weighting function*, which is derived from the concept of *capacities* — a concept which is explained in the next subsection. In common with the 1979 version of *Prospect Theory*, the value function is concave above the reference point ($v''(x) \leq 0, x \geq 0$), while being convex below the reference point ($v''(x) \geq 0, x \leq 0$). These conditions indicate that the impact of a change diminishes with increasing distance from the reference point. The value function v is also assumed to be steeper for losses than for gains in accordance with Tversky and Kahneman's (1986) principle of *loss aversion*.

Tversky and Kahneman (1992)⁶⁰ explain *Cumulative Prospect Theory* in terms of the monetary outcomes of events, where an event is defined as a subset s of S , the finite set of states of nature. Only one event can occur, but *which* one is not known in advance by the decision maker. The set of outcomes, X , associated with a prospect is assumed to contain a neutral outcome ($x = 0$). All other members of the outcome set are to be interpreted as gains denoted

⁵⁹ Tversky and Kahneman, (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty", p. 299.

⁶⁰ Ibid, p. 300.

by positive numbers, and losses denoted by negative numbers. An uncertain prospect f can be seen as a function mapping S on to X such that $f(s) = x$.

The major step away from the 1979 version of *Prospect Theory* comes next. The elements s of S are sorted into partitions of S , denoted A_i , where all A_i are subsets of S . Prospect f , when its outcomes have been sorted into an increasing order, contains a sequence of pairs (x_i, A_i) which yield x_i if A_i occurs, where $x_i > x_j$, if and only if $i > j$. Positive outcomes are denoted by positive subscripts ($i = 0, 1, \dots, n$), and negative outcomes are identified by subscripts which are negative ($i = -m, m+1, \dots, -1, 0$).

When a prospect contains both positive and negative outcomes, the positive part of f , denoted f^+ , is obtained by letting $f^+(s) = f(s)$ on the condition that $f(s) > 0$. The negative part is labelled f^- and is determined in a similar fashion. To distinguish prospects apart according to the over-all utility of each, each Prospect f is assigned a number represented by the symbol cluster, $V(f)$. For example $V(f) \geq V(g)$ states that Prospect f is at least as good as, or better than, Prospect g in the decision maker's eyes.

Assuming a range of values for i such that $-m \leq 0 \leq n$, the value function for *Cumulative Prospect Theory* becomes⁶¹:

$$V(f) = V(f^+) + V(f^-)$$

Where:

$$V(f^+) = \sum_{i=0}^n \pi_i^+ v(x_i);$$

$$V(f^-) = \sum_{i=-m}^0 \pi_i^- v(x_i).$$

Where prospects are strictly positive or strictly negative the value function becomes:

$$V(f) = \sum_{i=-m}^n \pi_i v(x_i)$$

⁶¹ Idem.

Where $\pi_i = \pi_i^+$ if $i \geq 0$; and $\pi_i = \pi_i^-$ if $i < 0$.

2.5.2 Decision Weights in Cumulative Prospect Theory

In the value function equations in the preceding subsection, the decision weights π_i , like the prospects themselves, had superscripts denoting positive or negative values: $\pi_i = \pi_i^+$ or π_i^- .

The decision weights of *Cumulative Prospect Theory* need to be understood in terms of Choquet's (1955)⁶² concept of *capacity*. This is a non-additive set function which generalizes the standard notion of probability. For every partition, A in the set of states of nature, S , there is a *capacity*, W such that $W(A) \geq W(B)$ whenever partition A contains partition B . The idea here is that a given partition A may contain outcomes which are at least as good as a given outcome, x_i , or it may contain only outcomes which are strictly better than x_i (say, partition B), in which case, A contains B .

What is of interest is the difference, $W(A)$ minus $W(B)$. Tversky and Kahneman (1992)⁶³ describe the decision weight π_i^+ as the difference between the *capacities* of the subset of events in which the outcomes are at least as good as outcome x_i and the subset of those in which the outcome is strictly better than x_i . The decision weight π_i^- is defined in a similar fashion, being the difference between the *capacities* of the subset of events in which the outcomes are at least as bad as x_i and the subset of events in which the outcomes are strictly worse than outcome x_i . Tversky and Kahneman (1992) summed up their explanation of decision weights in *Cumulative Prospect Theory* as follows⁶⁴:

Thus the decision weight associated with an outcome can be interpreted as the marginal contribution of the respective event defined in terms of the capacities, W^+ and W^- . If each W is additive, and hence a probability measure, then π_i is simply the probability of A_i . It follows readily from the definitions of π and W that for both positive and negative prospects, the decision weights add up to 1. For mixed prospects, however, the sum can

⁶² Choquet, G., (1955), "Theory of Capacities", *Annales de l'Institut Fourier* 5, pp. 131 - 295. Cited by Tversky and Kahneman (1992), op. cit. n. 59, p. 300.

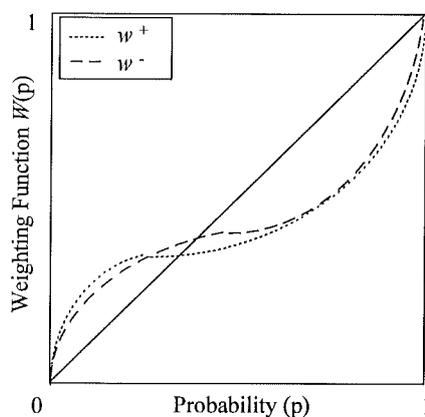
⁶³ Tversky and Kahneman, (1992), op. cit. n. 59, p. 301.

⁶⁴ Idem.

be either smaller or greater than 1, because the decision weights for gains and losses are defined by separate capacities.

When Tversky and Kahneman (1992) determined the shape of W^+ and W^- they found them to be dissimilar to the original *Prospect Theory* weighting. An initially concave curve develops convexity upon crossing the 45 degree line as depicted in Figure 2.4, where the horizontal axis measures probability and the vertical axis measures the two weighting functions.⁶⁵

Figure 2.4: Weighting Functions for Gains and Losses



Nevertheless, the information that Figure 2.4 conveys is that for both positive and negative prospects, decision makers tend to overweight low probabilities, underweight moderate and high probabilities, and show relative insensitivity to probability differences in the middle of the range. This is in line with the predictions of the 1979 version of *Prospect Theory*.⁶⁶

Although Dusenbury (1994)⁶⁷ investigated framing effects with respect to taxpayers' terminal tax filing behaviour in the context of *Prospect Theory*, nothing in *Cumulative Prospect Theory* contradicts the existence of the framing effects for which Dusenbury tested — and which are studied with reference to New Zealanders in the current study. Further consideration of framing effects is dealt with next.

⁶⁵ Ibid, pp. 312 - 313.

⁶⁶ Kahneman and Tversky (1979), op. cit. n.18, pp 280 - 284.

⁶⁷ The paper which was examined, replicated and extended by this study.

2.6 THE FRAMING OF DECISIONS

2.6.1 Framing Effects and Language-Induced Perceptions of Loss and Gain

Although the concept of framing was developed from analyses by Savage (1954) and Raiffa (1968) of anomalies running counter to the predictions of *Expected Utility Theory* aired by Allais (1953), it is Kahneman and Tversky in the papers they wrote together in 1979, 1981, 1984, and 1986 (all of which are cited in this chapter), who are responsible for the concept of decision frames which is discussed here.

Tversky and Kahneman (1986) provide a clear depiction of the power of decision frames in the following problem in which information is provided in a format conducive of disaggregated framing⁶⁸:

Problem 9 ($N = 150$):

Decision 1: Choose between:

Prospect S: A sure gain of \$240; (84%);

and

Prospect T: A 25% chance to gain \$1,000 and a 75% chance to gain nothing; (16%);

then

Decision 2: Choose between:

Prospect U: A sure loss of \$750; (13%);

and

Prospect V: A 75% chance to lose \$1,000 and a 25% chance to gain nothing; (87%).

The fact that most subjects chose S and V indicated they did not frame the decisions as a single decision even though they were told the decisions were concurrent. If the problem was

⁶⁸ Tversky and Kahneman (1986), op. cit. n. 36, p s255.

reconfigured so that it would be viewed as an overt choice among competing portfolio prospects, it became apparent the S & V portfolio was dominated by one of the others⁶⁹:

Portfolio S & V: 25% chance to win \$240 and a 75% chance to lose \$760;	(73%);
Portfolio T & U: 25% chance to win \$250 and a 75% chance to lose \$750;	(3%).

The failure of invariance was confirmed when the researchers found that the dominated prospect was almost always rejected when the problem was presented in its aggregated form. Tversky and Kahneman (1986) concluded that in the normal course of events people do not aggregate concurrent prospects or even view them with a common frame that would allow comparison of combined effects.

The construction of a decision frame is open to influence from a number of sources. One of these is the language used in the presentation of the prospects. By mere changes in the labelling of outcomes, the reference point of the decision frame may be significantly shifted. Tversky and Kahneman (1981) illustrated this point clearly when they manufactured a pair of problems from the single medical crisis scenario which is cited below. They provided a carefully worded pair of prospects for each, and presented the decision problems to samples of students at Stanford University and the University of British Columbia⁷⁰:

Problem 10 ($N = 152$):

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program W is adopted, 200 people will be saved;	(72%);
or	
If Program X is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved;	(28%).

⁶⁹ Ibid, p. s256.

⁷⁰ Tversky and Kahneman, (1981), op. cit. n. 38, p. 453.

The reference point Tversky and Kahneman's subjects tended to assume here was that the disease would kill all 600 people if left to run its course without medical intervention. Any lives saved by this intervention would consequently be viewed as a gain; and since a choice is to be made between a certain gain and a risky gain, one would expect the certain gain to be chosen in accordance with the principle of risk aversion. Indeed, 72 percent of Tversky and Kahneman's subjects chose the risk averse option. This finding supported the existence of risk aversion with respect to choices among gains.

In the second problem Tversky and Kahneman altered the wording (and only the wording) of the medical crisis solution prospects so that participants would assume a zero reference point at which the disease kills nobody; and that the choice is one between a certain loss and the risk of a greater loss. It is important to note that the wording of the scenario itself remained unchanged.

Problem 11 ($N = 155$):

If Program **Y** is adopted, 400 people will die; (22%);

or

If Program **Z** is adopted, there is a $1/3$ probability that nobody will die, and a $2/3$ probability that 600 people will die; (78%).

If decision makers were consistently risk averse, and if transitivity of preference were to be preserved, then Program **Y** would be chosen. This did not happen — transitivity was violated. More than two thirds of Tversky and Kahneman's subjects chose Program **Z**, which indicated risk seeking instead of risk aversion. This was in accordance with the predictions of *Prospect Theory*.

When Tversky and Kahneman investigated their subjects' responses to Problems 10 and 11 further, they found that both sophisticated and naive respondents tended to choose the transitivity-violating pair of prospects, **W** and **Z**. This experimental outcome was consistent whether the two problems were presented within a few minutes or widely apart in time. When Tversky and Kahneman confronted their subjects with their inconsistent answers, they found that they were unwilling — even on rereading the options — to relinquish the choice of risk aversion in Problem 10 and risk seeking in Problem 11; yet these very same people indicated

they wished to be consistent in their choice making (thus conforming with invariance). Kahneman and Tversky (1984), upon revisiting this pair of decision problems, commented⁷¹:

In their stubborn appeal, framing effects resemble perceptual illusions more than computational errors.

Kahneman and Tversky (1984) set up another pair of problems (here denoted Problems 12 and 13) which further demonstrated the absence of invariance in human decision-making. In this instance, the researchers proposed that a framing difference was caused by whether an initial outlay was perceived as a cost or a high-risk loss.⁷²

Problem 12 ($N = 132$):

Would you accept a gamble that offers a 10% chance to win \$95 and a 90% chance to lose \$5?

Problem 13 ($N = 132$):

Would you pay \$5 to participate in a lottery that offers a 10% chance to win \$100 and a 90% chance to win nothing?

These two problems share identical pairs of prospects. When they were presented with one intervening short filler problem, 55 of the 132 respondents chose inconsistent preferences; and 41 respondents rejected the gamble in Problem 12 while accepting the lottery in Problem 13.

2.6.2 Psychological Accounts

Tversky and Kahneman (1981)⁷³ looked beyond elementary decision problems which gave rise to gains or losses in a single attribute, to consider the process people use in editing and evaluating decision problems involving what they called *compound outcomes*. A compound outcome is one which entails a series of changes in a single attribute (for instance, a sequence of monetary gains and losses) or various concurrent changes in several attributes. In this

⁷¹ Kahneman and Tversky, (1984), op. cit. n. 2, p. 343.

⁷² Ibid, p. 349.

⁷³ Tversky and Kahneman, (1981), op. cit. n. 38, p. 456.

instance, the researchers proposed that decision makers use a specific outcome frame they called a *psychological account*.

A *psychological account* involves two requirements. To meet the first of these, decision makers edit their prospects into sets of elementary outcomes which can be evaluated jointly. The ensuing evaluations are, of course, subject to the influence of the manner in which the editing operations have been applied. The second requirement is that the decision makers have a concept of a reference outcome which they consider neutral or normal. Tversky and Kahneman (1981) argued that people facing complex problems with compound outcomes generally resort to evaluating prospects in terms of a *minimal account* which includes only the direct consequences of the decision made.⁷⁴

People commonly adopt minimal accounts because this mode of framing (i) simplifies evaluation and reduces cognitive strain, (ii) reflects the intuition that consequences should be causally linked to acts, and (iii) matches the properties of hedonic experience, which is more sensitive to desirable and undesirable changes than to steady states.

Tversky and Kahneman noted there are situations in which the possible outcomes of an act influence an account that has been set up as a result of some prior but related act. This may cause the decision problem of the moment to be evaluated in terms of a more inclusive account. They cite the well documented risk-seeking behaviour of punters placing bets on the last horse race of the day — a behaviour determined by the context of antecedent losses.⁷⁵ But they note, however, that more generally, a sunk-cost effect occurs when a decision involves bringing to mind an existing account which has a negative current balance. In other words and in the context of the current study, a person paying terminal tax is likely to regard prior payments of provisional tax as a sunk cost to be dismissed from consideration. This issue is investigated more fully in the next chapter.

⁷⁴ Ibid, p. 457.

⁷⁵ Idem.

2.6.3 Research involving Decision Frames outside the area of Tax

The concept of decision frames has attracted considerable interest from researchers operating in quite diverse fields; and some have been aware of *Prospect Theory* while others have observed and evaluated the phenomenon with little or no reference to the theory. Consequently the decision frame phenomenon could be regarded as a nexus by which all decision theory research becomes interconnected. Hence the body of research so far undertaken is vast; and it would be inappropriate to attempt to provide an exhaustive index of all of it. Instead a representative selection from 1979 onward is covered.

The year 1979 was seminal. Along with Kahneman and Tversky's first paper on *Prospect Theory* and Fishburn and Kochenberger's paper on the nature of utility functions,⁷⁶ an important paper by Grether and Plott was published.⁷⁷ Grether and Plott, in setting out to debunk the validity of psychology research in economics, laid out and tested thirteen possible variables which could be used for minimising the phenomenon of preference reversal. The one they found to be of greatest significance was what they called an *anchor*, consisting of the most prominent dimension or aspect of the decision-related object, from which the value of the object could be adjusted upward or downward to account for other aspects or features. Grether and Plott noted⁷⁸:

The particular dimension used as an anchor is postulated to be a function of the context in which a decision is being made.

This is the decision frame concept formulated independently and cast in different words. Grether and Plott's work was refined by Pommerehne, Schneider and Zweifel (1982)⁷⁹ and Reilly (1982)⁸⁰, who found in both cases that modifications in the design of experiments could

⁷⁶ Fishburn, P. C. and Kochenberger, G. A., (1979), "Two-Piece Von Neumann - Morgenstern Utility Functions".

⁷⁷ Grether, D. M. and Plott, C. R., (1979), "Economic Theory of Choice and the Preference Reversal Phenomenon".

⁷⁸ Ibid, pp. 627 - 628.

⁷⁹ Pommerehne, W. W., Schneider, F. and Zweifel, P., (1982), "Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination".

⁸⁰ Reilly, R. J., (1982), "Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design".

impact on the frequency of preference reversals but not eliminate them. Other early studies documenting the importance of framing of choices with reference to a target, or aspiration, level were made by Payne, Laughhunn and Crum (1980 and 1981).⁸¹

Focusing on the decision frame concept explicitly, Fischhoff (1983)⁸² found framing effects to be significant; but noted that there was no guarantee an individual would adopt the same frame as any other individual operating in the same context. The significance of this finding was echoed by Slovic and Lichtenstein (1983).⁸³

Arkes and Blumer (1985)⁸⁴ found evidence supporting Tversky and Kahneman's (1981) *sunk cost effect* mentioned in the preceding subsection. Further work was done in this area by Garland and Newport (1991)⁸⁵; and Whyte (1993)⁸⁶ incorporated an investigation of the *sunk cost effect* in his study of group decision-making.

Slovic, Fischhoff and Lichtenstein (1988), in examining problems posed by Tversky and Kahneman (1986) and other problems of a similar nature, suggested that individuals tend to perceive and think at a concrete level, only using information that is displayed explicitly in the formulation of the problem. Information which has to be logically deduced from the given display frame or created by some form of transformation of the given data (by the decision maker) escapes notice.⁸⁷ This finding was, to a degree, contradicted by Darke and Freedman (1993),⁸⁸ who showed that consumers are capable of framing their decisions at several different

⁸¹ Payne, J. W., Laughhunn, D. J. and Crum, R., (1980), "Translation of Gambles and Aspiration Level Effects in Risky Choice Behavior"; and (1981), "Further Tests of Aspiration Level Effects in Risky Choice Behavior".

⁸² Fischhoff, B., (1983), "Predicting Frames".

⁸³ Slovic, P. and Lichtenstein, S., (1983), "Preference Reversals: A Broader Perspective".

⁸⁴ Arkes, H. R. and Blumer, C., (1985), "The Psychology of Sunk Cost".

⁸⁵ Garland, H. and Newport, S., (1991), "Effects of Absolute and Relative Sunk costs on the Decision to Persist with a Course of Action".

⁸⁶ Whyte, G., (1993), "Escalating Commitment in Individual and Group Decision Making: A Prospect Theory Approach".

⁸⁷ Slovic, P. Fischhoff, B. and Lichtenstein, S., (1988), "Response Mode, Framing, and Information Processing Effects in Risk Assessment", p. 153.

⁸⁸ Darke, P. R. and Freedman, J. L., (1993), "Deciding Whether to Seek a Bargain: Effects of both Amount and Percentage Off".

levels by proving that both the dollar amount deducted and a percentage reduction in price operated as perceived gains from a neutral reference point.

Also in the area of consumer behaviour, Casey (1994)⁸⁹ developed and tested two descriptive models for determining maximum buying prices based on *Prospect Theory*. He found strong evidence of *loss aversion*, in that the observed maximum buying prices chosen by his subjects for the given experimental prospects were significantly below the prospects' expected values, even when the probability of a win was close to certainty. Casey concluded that individuals frame investments in risky alternatives as sure losses to be weighed against gains that are merely probabilistic.⁹⁰

Framing effects were found to exert an influence in intertemporal choice making by Loewenstein (1988).⁹¹ Intertemporal consumption shifts could be framed as *delays* of consumption or the bringing of consumption forward; but the impact of a *delay* frame was lessened if subjects were asked to state the present value of the expenditure in question at the two given points in time. Another investigation of intertemporal choice selection was made by Hirst, Joyce and Schadeewald (1994).⁹² They focused on the role that temporal contiguity and causal reasoning play in mental accounting for consumer-borrowing decisions; and, within the *Prospect Theory* paradigm, their particular interest was in Tversky and Kahneman's (1981) psychological accounts.⁹³ In pursuing their research objectives they confirmed, yet again, that decisions are influenced by how outcomes are framed.⁹⁴

Tversky, Slovic and Sattath (1988)⁹⁵ looked at framing from an entirely different point of view. Their paper found a significant discrepancy existed between responses to *gambling problems*

⁸⁹ Casey, J. T., (1994), "Buyers' Pricing Behavior for Risky Alternatives: Encoding Processes and Preference Reversals".

⁹⁰ Ibid, p. 747.

⁹¹ Loewenstein, G. F., (1988), "Frames of Mind in Intertemporal Choice".

⁹² Hirst, D. E., Joyce, E. J. and Schadeewald, M. S., (1994), "Mental Accounting and Outcome Contiguity in Consumer-Borrowing Decisions".

⁹³ See Subsection 2.6.2 above.

⁹⁴ Hirst, Joyce and Schadeewald, (1994), op. cit. n. 92, p. 146.

⁹⁵ Tversky, A., Slovic, P. and Sattath, S., (1988), "Contingent Weighting in Judgment and Choice".

posed as choices among given prospects on the one hand, and responses, on the other, to problems set up as *matching tasks* in which subjects had to assign a dollar value. This quite separate framing effect was termed the *elicitation effect*.

Elliot and Archibald (1989)⁹⁶ argued that the focus in research, up until the time of their study, had been on frames that had been externally imposed rather than generated by experimental subjects themselves from the supplied information. When they reworded Kahneman and Tversky's (1981) disease scenario⁹⁷ in order to elicit the subjective frame chosen by subjects, and then questioned the subjects on how they framed the choices they made, Elliot and Archibald found that decision makers independently imposed a frame on choice problems.⁹⁸ In running a second experiment on the same sample of subjects immediately afterwards, Elliot and Archibald also discovered that knowledge of risk preference (as determined by responses to four two-option gambling problems), was not as powerful a predictor of treatments chosen (as solutions for the disease scenario), as was knowledge of a subject's choice of frame.⁹⁹

Tversky, Slovic and Kahneman (1990)¹⁰⁰ found 90 percent of observed preference reversals in their experiments represented violations of invariance which could be subdivided into failures of description invariance (framing effects) and procedure invariance (elicitation effects). They perceived invariance failure to be a much greater problem for rational choice models than the failure of specific axioms such as independence or transitivity. This caused Tversky, Slovic and Kahneman to conclude that to construct a theory of choice that was both normatively acceptable and descriptively adequate was not possible.

⁹⁶ Elliot, C. S. and Archibald, R. B., (1989), "Subjective Framing and Attitudes towards Risk".

⁹⁷ See Problems 10 and 11 in Section 2.6.1 of this chapter.

⁹⁸ Ibid, p. 324.

⁹⁹ Ibid, p. 326.

¹⁰⁰ Tversky, A., Slovic, P. and Kahneman, D., (1990), "The Causes of Preference Reversal", pp. 214 - 215.

In a more recent study in which the focus shifted from decision-making to reasoning and justification, Shafir, Simonson and Tversky (1993)¹⁰¹ noted that reasons advanced by decision makers for their decisions also lend themselves to being modified by framing manipulations.

Research into the nature of decision frames has also been conducted of late at the University of Canterbury. Fountain (1992)¹⁰² studied framing effects with respect to quasi-rational consumer demand; and Fountain and McCosker (1993)¹⁰³ explored the impact of framing differences with respect to the transparency of risk on risky choices, and also with respect to the paradoxical common consequence effect posed by Allais (1953). The relationship between changes in risk transparency and risk attitudes was further explored by combinations of these researchers and others in working papers in 1994 and 1995.¹⁰⁴

By way of an endnote, it would seem that framing effects are universally pervasive; and evidence has been found for them even outside the world of human decision-making. Hamm and Shettleworth (1987)¹⁰⁵ ran an experiment on pigeons and found pigeons to be risk averse above a particular desired body weight and risk-seeking below that weight. The neutral reference point was the mass at which a bird was deemed to think that survival was assured.

Attention in the next chapter is focused on the application of *Prospect Theory* and decision frames in tax research.

¹⁰¹ Shafir, E., Simonson, I. and Tversky, A., (1993), "Reason-based Choice".

¹⁰² Fountain, J., (1992), *Quasi Rational Consumer Demand — Some Positive and Normative Surprises*, Discussion Paper No. 9202.

¹⁰³ Fountain, J. and McCosker, M., (1993), *Fans, Frames and Risk Aversion: How Robust is the Common Consequence Effect?*, Discussion Paper No. 9303.

¹⁰⁴ Fountain, J., McCosker, M. and Morris, D., (1994), *Are People Really Risk Seeking when Facing Losses?* Discussion Paper No. 9407; and Fountain, J., McCosker, M. and MacFarlane, B., (1995), *Framing and Incentive Effects on Risk Attitudes when facing Uncertain Losses*, Discussion Paper No. 9503.

¹⁰⁵ Hamm, S. and Shettleworth, S., (1987), "Risk Aversion among Pigeons".

3. TAX RESEARCH INVOLVING PROSPECT THEORY AND DECISION FRAMES

3.1 INTRODUCTION

It was inevitable that an interesting new descriptive theory of decision-making under risk emanating from the interface of psychology and economics would be taken up by students in the field of tax compliance research. Where *Expected Utility Theory* had entered the compliance field via the development of *economic deterrence* models, *Prospect Theory* has made its impact in a form of tax compliance modelling already known by the title, *fiscal psychology*. *Economic deterrence* models of taxpayer behaviour assume that taxpayers are rational economic agents maximising utility by accepting options with the highest expected value. This means taxpayers will evade whenever the return on evasion is greater than the expected cost of being caught.¹ *Fiscal psychology models*, prior to the advent of *Prospect Theory*, inductively examined the attitudes and belief structures of taxpayers in order to predict their tax-filing behaviour.² With the advent of *Prospect Theory*, a set of grounded premises has been provided from which deductive inferences may be drawn. However, there is neither a *normative*, nor a *descriptive* theory encompassing all framing behaviour as yet.

The chapter is divided into three discrete sections. Section 3.2 explores developments in compliance research which set the stage for investigation of *Prospect Theory* and framing effects in compliance, which is covered in Section 3.3. Section 3.3 deals with papers produced prior to, and at the same time as, Dusenbury (1994). Section 3.4 deals with the related issue of decision frame research with respect to the relationship between clients and tax professionals. Dusenbury's 1994 paper, the subject of this replication in a New Zealand environment, is studied in depth in the next chapter.

¹ This definition is adapted from Hasseldine, D. J. and Bebbington, K. J., (1991), "Blending Economic Deterrence and Fiscal Psychology Models in the Design of Responses to Tax Evasion: The New Zealand Experience", p. 302.

² *Ibid*, p. 304.

3.2 ORIGINS OF DECISION FRAME RESEARCH IN COMPLIANCE

The earliest approach to analysis of tax compliance behaviour was undertaken largely by researchers grounded in economics, who perceived the decision to comply or evade as a problem of net income maximization, which could be evaluated in terms of Von Neumann and Morgenstern's axioms for decision-making under uncertainty. The taxpayer's cardinal utility function had income as its only argument and the associated property of positive, but strictly decreasing, marginal utility indicated that the taxpayer could be expected to be risk averse.

When Allingham and Sandmo (1972) used this template to model the tax decision context and conducted a formal analysis of the resulting model, they found an increase in the probability of detection would always lead to a larger income being declared; and that an increase in the penalty rate would achieve the same effect. They concluded that the two policy tools available to taxation authorities — manipulation of detection probability and control of penalty tax rates — were interchangeable.³ A similar *Expected Utility Theory* tax compliance model, or *economic deterrence* model, was developed independently by Srinivasan (1973);⁴ and Allingham and Sandmo's model was developed and extended by Weiss (1976).⁵ In a paper which is distantly related to the central concern of this study, Yaniv (1988)⁶ investigated employer compliance behaviour relating to the reporting (and forwarding) of employee salary withholdings, and employee behaviour with respect to reporting total income and paying further tax. His model predicted an increase in evasion by employees as the withholding rate was reduced. Compliance research is still being conducted within the *Expected Utility Theory*

³ Allingham, M. G. and Sandmo, A., (1972), "Income Tax Evasion: A Theoretical Analysis", p. 330.

⁴ Srinivasan, T. N., (1973), "Tax Evasion: A Model", cited by Chang, O. H., Nichols, D. R. and Schulz, J. J., (1987), "Taxpayer Attitudes Toward Tax-Audit Risk", p. 300.

⁵ Weiss, L., (1976), "The Desirability of Cheating Incentives and Randomness in the Optimal Income Tax". Weiss extended Allingham and Sandmo's model and showed that when a labour supply variable was included, the model implied that "...it might be desirable (in the sense of increasing all agents' expected utility) if the taxing authority treats identical individuals differently." p. 1344.

⁶ Yaniv, G., (1988), "Withholding and Non-Withheld Tax Evasion".

paradigm, but unless some relevance has been perceived to the research objectives of this study, this body of research is largely ignored here.⁷

The gateway to consideration of compliance in more than just *Expected Utility Theory* terms was provided by Lewis (1982), who provided a critical review of existing social surveys of attitudes to evasion. Lewis, noting the “striking coherence”⁸ they furnished, developed a formal model in which psychological and sociological variables were applied to mapping the tax evasion process.⁹ This work was built upon by Groenland and van Veldhoven (1983),¹⁰ who provided a similar psychological framework. The structures provided by Lewis (1982) and Groenland and van Veldhoven were, however, in no way *Prospect Theory* models; but this was not the case with the models provided by Smith and Kinsey (1987). They incorporated, within their flow chart of the compliance decision process, a framing step in which the tax situation was coded as a loss or a gain.¹¹ They proposed that tax behaviour might not necessarily involve conscious decisions.

That *economic deterrence* model research conducted within the *Expect Utility Theory* paradigm did not provide universally satisfactory answers to all the questions relating to tax compliance decisions, became increasingly apparent during the 1980s. A representative piece of research bearing this message was conducted by Baldry (1986).¹² He contrasted results from a pair of experiments employing identical risk-payoff structures but which were garbed in dissimilar financial decision-making contexts. In one experiment, subjects were faced with an opportunity for tax evasion while in the other, the decisions to be made related simply to the

⁷ Peter Bardsley has published a paper, “Tax Compliance Research: An Economic Perspective on the Research Agenda”, in Volume 11(3) of the *Australian Tax Forum*, 1994 — a paper which provides a fairly current perception of the compliance problem in *Expected Utility Theory* terms. He notes (p. 284) that there has been a renaissance in decision theory research in response to criticisms levelled by Allais (1953) and Kahneman and Tversky (1979), and goes on to state: “[T]he apparent violations of expected utility theory can be and are being accommodated within the rational tax payer paradigm (though it remains to be seen whether they are empirically important).”

⁸ Lewis, A., (1982), *The Psychology of Taxation*, p. 121. More generally, see Part III, “Tax Evasion: the Link between Attitudes and Behaviour and Implications for Reducing Evasion”, pp 121 - 160.

⁹ *Ibid*, p. 156.

¹⁰ Groenland, A. G. and van Veldhoven, G. M., (1983), “Tax Evasion Behavior: A Psychological Framework”.

¹¹ Smith, K. W. and Kinsey, K. A., (1987), “Understanding Taxpaying Behavior: A Conceptual Framework with Implications for Research”, p. 645.

¹² Baldry, J. C., (1986), “Tax Evasion is not a Gamble”.

placing of bets. Every participant in the gambling experiment took up a risky position but Baldry recorded only 72 instances of evasion out of 104 observations in the tax experiment. Baldry recognised that gambling and tax evasion were seen in two quite different lights by the participants, and noted that this bifurcation of perceptions could not be explained in *Expected Utility Theory* terms.

Clotfelter (1983) ran a series of regressions on 47,000 tax returns for 1969 under the auspices of the IRS Taxpayer Compliance Measurement Program (TCMP). Clotfelter's dependent variable was the logarithm of income deemed to be underreported by IRS auditors and his primary focus of interest was the relationship between it and marginal tax rates. While Clotfelter noted that TCMP data omitted information on many personal or attitudinal variables, he found that income earners on wage incomes, who had taxes withheld from their pay packets, were associated with full reporting, and that any underreporting was offset by overreporting. He explained this finding in terms of two factors. The first was withholdings by employers; and the other was that employers passed employee income information on to the IRS. Clotfelter did not appear to be aware of the concept of framing effects and did not expand the discussion beyond ascribing fear of detection as the motive for compliance by this class of taxpayer. However Clotfelter did observe in a footnote that self-employed taxpayers owing additional taxes upon filing their annual returns tended to understate income more than other taxpayers; but he dismissed the issue on the ground that *tax balance due* was probably not an exogenous variable since the decision to underpay provisional tax and underreport terminal tax might well be made jointly.¹³

This point was also taken up by Cowell (1985),¹⁴ who extended Allingham and Sandmo's (1972) model to allow for trade-offs between income from which the tax authorities could withhold PAYE, and illegal income activities and leisure on the ground that a person's evasion opportunities could be improved by switching work effort from income-earning activities subject to tax withholdings to those activities that do not.

¹³ Clotfelter, C. T., (1983), "Tax Evasion and Tax Rates: An Analysis of Individual Returns", p. 368, Footnote 15.

¹⁴ Cowell, F. A., (1985), "Tax Evasion with Labour Income".

The existence of a differential compliance phenomenon relating to levels of tax withholdings was recognised widely in the 1980s. Wallschutzky (1984) noted it in an Australian context when samples of known tax evaders and others (with unspecified characteristics) agreed that self-employed taxpayers had a greater opportunity to evade than did salaried workers who were subject to withholdings from income at source.¹⁵ In the United States of America, Jackson and Milleron (1986) rated *tax withheld at income source* as one of the 14 most important variables from a list of 64 compliance variables identified by the IRS.¹⁶

When Milleron and Toy (1988)¹⁷ examined the views of 152 certified public accountants in the United States on the topic of tax compliance and how it might be improved, they found that *withholding* (along with *information reporting*) was considered to be the third most important factor affecting compliance. On a seven-point Likert scale, on which 7 denoted maximum importance, the CPAs awarded the *withholding* variable a 4.39 rating; and when they were asked whether use by tax authorities of the paired variables, *withholding* and *information reporting*, should be maintained, reduced or increased, the CPAs opted for maintenance at the current level.

When ten further possible means of increasing taxpayer compliance were introduced, the CPAs demoted the concept of increasing withholdings from income to eleventh place, below reduction of tax rates, reduction of the government's preponderant reliance for revenue on income tax, reduction of the range of possible deductions, increases in the probability of audits and increases in penalties for evasion.

Nevertheless, when Milleron and Toy contrasted support for *fiscal psychology* systems with systems founded on *Expected Utility Theory*-based economic deterrence models, the CPAs clearly indicated they considered the *fiscal psychology* approaches superior. Manipulation of

¹⁵ Wallschutzky, I. G., (1984), "Possible Causes of Tax Evasion", p. 381.

¹⁶ Jackson, B. R. and Milleron, V. C., (1986), "Tax Compliance Research: Findings, Problems, and Prospects", p. 126. Here Jackson and Milleron are citing the Internal Revenue Service *Taxpayer Compliance Measurement Program Handbook*, January 1984.

¹⁷ Milleron, V. C. and Toy, D. R., (1988), "Tax Compliance: An Investigation of Key Features".

the level of income withheld at source (or paid in advance) to meet future tax obligations clearly belongs in the *fiscal psychology* tool-box.

On the basis of this level of understanding of the impact of tax withheld at source and prepayments, Wallschutzky (1989) recommended that the principle of withholding tax at source be extended in Australia;¹⁸ and in the same year in New Zealand, resident withholding tax on interest income and dividends was added to the Inland Revenue Department's taxes for collection.

Summing up the shift away from exclusive reliance on *economic deterrence* models of tax compliance behaviour, Carroll (1992) stated¹⁹:

Although the economic model can be extended to include social consequences, at some point it stops resembling the economic model. On the one hand, the components of the consequence calculation become subjective and socially constructed rather than objective aspects of marginal tax rates and audit rates... On the other hand, even granting the subjective nature of risky consequences, taxpayers may not use consequence information in a way that is presumed by the rational economic model.

3.3 FRAMING EFFECT RESEARCH IN COMPLIANCE PRIOR TO DUSENBURY

Investigation of framing effects based on the level of taxes withheld prior to the filing of annual tax returns began in earnest, largely in the United States, in 1985, the year in which the concept was broadly disseminated in the popular journal, *Psychology Today* (Loftus 1985).²⁰ However

¹⁸ Wallschutzky, I. G., (1989), "The Effectiveness of Tax Office Enforcement Activities in Achieving Taxpayer Compliance", p. 501.

¹⁹ Carroll, J. S., (1992), "How Taxpayers Think about Their Taxes: Frames and Values", pp. 46 - 47.

²⁰ Loftus, E. F., (1985), "To File, Perchance to Cheat". This article also contained an interesting inset. A New York Times/CBS News Poll found that 45 percent of the population believed people like themselves paid too much; that 69 percent believed people with higher family incomes than themselves paid too little tax; and that 50 percent believed people with lower family incomes than themselves paid too much. It would seem that even fairness is susceptible to a framing effect. The trend of the poll's results was remained clearly discernible when the respondents were partitioned into those with family incomes above \$US25,000, and those whose family incomes were below that cut-off point. (The sample of 1500 people was canvassed in January 1985.)

it would be sensible to commence this subsection with a discussion of what is understood by the term *decision frame* in a tax compliance context.

3.3.1 Decision Frames in the Context of Tax

Carroll (1989) defined the term *framing effect* as the manner in which the editing processes of *Prospect Theory* work. Echoing Tversky and Kahneman (1981), he argued a problem is simplified by the selection (conscious or unconscious) of a particular set of definitions and viewpoints²¹:

In a general way ‘frames’ can be thought of as viewpoints or metaphors that help structure ill-structured problems, thus separating figure from ground, highlighting some aspects or a situation and hiding others.

Carroll provided two illustrations — one applying to the decision processes of individual taxpayers, and the second applying to both individuals and an entire political movement. In the first instance, Carroll saw that people making tax decisions could *frame* their behaviour as compliance and, in so doing, adopt a set of strategies designed for meeting the perceived obligation; or they could *frame* their behaviour instead, as avoidance, and adopt strategies for identifying legal loopholes and devious devices for tax minimization.²²

In the second instance, tax could be *framed* as a fair contribution to the running of the state and to the welfare of all citizens; or it could be *framed* as a measure of government waste and a loss both to society and to individual taxpayers. Carroll noted that Reaganomics of the 1980s appeared imbued with the latter decision frame. There is evidence to suggest that this frame may be influential in the Pacific Region as well. Wallschutzky (1984)²³ found 70 percent of the respondents to his mail questionnaire were dissatisfied with the way in which the Australian

²¹ Carroll, J. S., (1989), “A Cognitive-Process Analysis of Taxpayer Compliance” in *Taxpayer Compliance Volume 2: Social Science Perspectives*, ed. Roth, J. A. and Scholz, J. T., p. 242. This definition is ascribed to J. R. Russo and P. J. H. Shoemaker disclosed in a personal communication with Mr Carroll.

²² Scholz, J. T., (1985), “Coping with Complexity: A Bounded Rationality Perspective on Taxpayer Compliance”, *Proceedings of the Seventy-Eighth Annual Conference of the National Tax Association - Tax Institute of America*, Columbus, Ohio. This was cited by Carroll, op. cit. 21, p. 242.

²³ Wallschutzky, (1984), op. cit. n. 15, p. 379.

Federal Government spent their tax money; and 80 percent were unhappy with Australian income tax laws.

Furthermore, Carroll emphasised that taxpayers' rationality regarding compliance decisions is variable, and that this variability carries over into how planned and how pro-active their decisions actually are²⁴:

They may switch back and forth between self-aware and deliberate decision making and the habitual, unthoughtful carrying out of procedures set down earlier in the process. For most people, 'popping up' a level and considering premises as problematic is unusual.

Probing the concept a little further, Carroll noted that a number of taxpayer attitudinal types could be interpreted in terms of the basic motive or frame they adopt in furnishing their tax returns. He listed (1) honest taxpayers, (2) utility maximizers, (3) beaten taxpayers, (4) equity seekers and (5) taxpayers mobilized by need. Each would approach the decision problem differently.²⁵

These postulations were borne out in his own empirical research. Carroll (1992)²⁶ found evidence in a sample of 100 Boston taxpayers of ranges of frames, and also evidence of multiple frames held simultaneously by a single individual.²⁷ Carroll's method involved not only interviewing his subjects at both the start and the end of the 1988 tax year, but requiring them to keep a diary throughout the year as well, detailing their thoughts about their tax obligations. One respondent, for instance, stated²⁸:

I made \$1,500 more in 1987 than in 1986 and in 1987 I will only be paying about \$6 more in taxes... It appears the new tax laws have been working in

²⁴ Carroll, (1989), op. cit. 21, p. 254.

²⁵ *Idem*.

²⁶ Carroll, J. S., (1992), op. cit. n. 19.

²⁷ This finding is similar in some respects to Darke and Freedman's (1993) finding reported in Chapter Two, Subsection 2.6.3 of this study. The existence of a range of possible frames was also covered by Grether and Plott (1979), and Fischhoff (1983) — also mentioned in Subsection 2.6.3 — and also Tversky and Kahneman (1981), mentioned in Chapter Two, Subsection 2.6.1.

²⁸ *Ibid*, p. 51.

my benefit. In spite of actually having more money during the year, I still rather miss the large tax refund check... Still *feels* like a loss.

In this one statement the individual provides evidence of four frames. The first of these is a sense of total tax paid; the second, a benchmark frame in which this year's tax is measured against last year's tax; the third, an economic frame in which the time value of money, not lost to the taxpayer in the form of tax withheld, is recognised; and the fourth, a refund frame in which the absence of the traditional refund is regarded as a loss.

However, 60 percent of all disclosures made in diaries and interviews were codable in terms of a frame related to concerns about potential or actual refunds. This provides support for an observation made by Westat (1980), in a qualitative survey commissioned by the Internal Revenue Service twelve years earlier regarding withholdings. Westat noted²⁹:

People tend not to be aware of the amount of money paid through withholding. They do not 'miss what they never have.' People are motivated to avoid having to pay extra tax when they file. This money is 'missed' much more than the money withheld.

3.3.2 Prospect Theory Applied

Prospect Theory was applied for the first time to compliance research by Jackson and Jones (1985);³⁰ but in their study, the investigation of decision frames was only a very minor part. Jackson and Jones' main focus was on Kahneman and Tversky's subadditivity effect. To observe this effect, they set up two experiments testing whether taxpayers were more sensitive to the risk of detection than to the size of the penalty tax imposed if detection occurred.

Jackson and Jones found, when the probability of detection was low and the expected value of outcomes was equal for alternative prospects, the choice with the lower magnitude of penalty would be preferred. In other words, that taxpayers would be more sensitive to the magnitude

²⁹ Westat Inc, (1980), *Individual Income Tax Factors Study: Qualitative Research Results*, Prepared for the Internal Revenue Service, p. 11.

³⁰ Jackson, B. R. and Jones, S. M., (1985), "Salience of Tax Evasion Penalties versus Detection Risk".

of a penalty than to the probability of detection. The experimental subjects were university students studying commerce, who could be expected to have had some acquaintance with the concept of expected values. Nevertheless the students behaved in a manner which suggested they had a propensity to overweight low probabilities in accordance with *Prospect Theory's* subadditivity effect.

However, Jackson and Jones were aware of the relative imprecision of their experimental instrument. Upon asking the subjects who chose the indifference option in one of their contextual question sets why they did so, they found a multiplicity of motivations, some of which were at cross purposes to the researchers' aims.

But Jackson and Jones were not aware of a more important flaw. While they did not set out directly to investigate decision frames in a tax context, framing was an issue of considerable importance to their investigation and they were dismissive of it. They contended that the consistency of results over all their sets of questions indicated no significant difference existed between contextual and noncontextual decision frames³¹:

The results in the contextual framework support those in the noncontextual questions, leading to the conclusion that framing was not a substantial problem.

With respect to the imprecision of their experimental instrument, this conclusion is open to question.

The second *Prospect Theory* paradigm compliance study was conducted by Chang, Nichols and Schulz (1987);³² and again, the focus on framing effects was not yet a direct one. Chang, Nichols and Schulz set out to establish that switches between risk aversion and risk seeking behaviour occurred within the same group of experimental subjects.

³¹ Ibid, p. 13.

³² Chang, O. H., Nichols, D. R. and Schulz, J. J., (1987), "Taxpayer Attitudes toward Tax-Audit Risk".

The experiment required 56 MBA students to make choices in six hypothetical tax lotteries in which the expected values of the prospects were equal in value to each lottery's certainty alternative. Chang et al hypothesized that where a lottery and its certainty alternative were equal, a risk-neutral taxpayer would be indifferent between playing and not playing; but a risk-seeking taxpayer would opt to play the lottery. Chang et al argued that taxpayers' risk preferences could be inferred from the pattern of choices to play or not to play made by the subjects.³³

The highest incidence of participants saying *yes* to one of the six lotteries was 55.4 percent; but in all five other cases the incidence was less than 50 percent and ranged down to 26.8 percent. These results implied that on the whole, taxpayers are risk averse — a finding which supported *Expected Utility Theory* rather than *Prospect Theory*.

However, when they sorted responses according to whether the subjects viewed tax payments in the lotteries as a reduced gain or a sure loss, the results were remarkably different. In the case of the group perceiving reduced-gain, the highest incidence of a *yes* response to a lottery was 36 percent and the lowest was 9 percent. These people clearly displayed risk aversion. On the other hand, the group which had perceived the tax payments as a pure loss displayed clear-cut risk seeking. The highest incidence of a *yes* response to a lottery in their case was 82 percent; and only one of the six lotteries failed to attract less than a 50 percent *yes* tally. This constituted clear evidence of a framing effect arising from perceptions of gain and loss. Nevertheless the propensity towards risk aversion evident in the results produced by unsorted full sample did not allow Chang et al to refute the predominance of risk aversion in taxpayers.

Chang, Nichols and Schulz recognised several weaknesses in their study. Their sample size was small; and the detection probabilities (10, 50 and 90 percent) were unrealistically high as was one of the penalties (900 percent penalty tax). The use of high-level probabilities of detection was shown by Schepanski and Kelsey (1990)³⁴ to have a dampening effect on both

³³ Ibid, p. 301.

³⁴ Schepanski, A. and Kelsey, D., (1990), "Testing for Framing Effects in Taxpayer Compliance Decisions", Figure 2, p. 65.

risk aversive and risk seeking behaviour. Furthermore the lotteries were purely hypothetical. And, last, Chang et al did not investigate effects resulting from the *manipulation* of decision frames — they simply sorted their subjects' responses according to each individual's own perception of gain or loss. Consequently their investigation must be seen as the measuring of a single decision frame.

3.3.3 Gain/Loss Perceptions

The following year two major unpublished studies investigating the impact of gain/loss perceptions as a framing effect were produced; and in both, the neutral reference point for the framing effect was anchored to the level of taxes withheld prior to the tax return filing date. The first of these was the working paper, Hite, Jackson and Spicer (1988);³⁵ and the second was a large-scale study of archival American tax return data conducted by the Internal Revenue Service, cited as Cox and Plumley (1988),³⁶ which will be discussed in Subsection 3.3.4.

In their working paper, Hite, Jackson and Spicer required a sample of citizens called up for jury service in an undisclosed city in Colorado to give responses to questions associated with a scenario in which the jurors were told they had \$US1,000 income from informal employment unknown to the IRS and that reporting any or all of this so-far undisclosed income would have an unfavourable impact on their refund/tax bill. Effectively the jurors were being given an invitation to disclose whether or not they would choose to evade.

There were actually four different scenarios; but each juror looked at one only; hence Hite et al had four treatment groups to compare and contrast. The scenarios consisted of two instances in which provisional tax had been overwithheld and two in which it had been underwithheld; and each of these pairs broke into an instance in which there was a high monetary amount in

³⁵ Hite, P. S., Jackson, B. R. and Spicer, M. W., (1988), "The Effects of Decision Framing on Taxpayer Compliance".

³⁶ Cox, D. and Plumley, A., (1988) *Analyses of Voluntary Compliance Rates for Different Income Source Classes*, Unpublished report. IRS Research Division, Washington DC. This report was cited in the international study, Robben, H. S. J., Webley, P., Weigel, R. H., Warneryd, K., Kinsey, K. A., Hessing, D. J., Alvira Martin, F., Elffers, H., Wahlund, R. Van Langenhove, L. Long, S. B. and Scholz, J. T., (1990), "Decision Frame and Opportunity as Determinants of Tax Cheating", pp. 345 - 346 and by Webley P., Robben, H. S. J., Elffers, H. and Hessing, D. J., (1991) "Tax Evasion: An Experimental Approach", pp. 83 - 84.

question, and a scenario in which there was a relatively low monetary amount at stake. Hite et al's null hypothesis was that there was no difference in results between the different treatment groups.

Initial MANOVA testing indicated no significant difference existed between each group and any other; but that a number of the covariates in the analysis —gender, past evasion behaviour and perception of the likelihood of discovery— were significant at the 5 percent level. Hite et al recognised, however, that this result was potentially confounded by differences of pre-existing neutral reference points among subjects; so they divided the sample into three subsamples on the basis of the jurors' levels of income and ran the MANOVA analysis again. It was found that withholding status was a significant variable for the low income group (in which the income maximum was just short of \$US15,000): the low income earners showed a significant propensity to evade when underwithheld rather than overwithheld. The high income group furnished the opposite result. Contrary to the researchers' prior expectation, this group (income greater than \$US30,000) opted to report lower income when overwithheld than underwithheld. This latter result was significant at the 4 percent level. Over-all Hite et al's results were therefore inconclusive.

Hite et al attributed their mixed findings to a number of limitations inherent in their research approach. In the first instance they could not know if the jurors treated the scenarios in the same manner as their actual tax obligations. An amplification of this limitation was provided by Elffers, Weigel and Hessing (1987), who compared the self-reported attitudes towards compliance of Dutch taxpayers with their actual tax-filing behaviour (made available by the Dutch tax authorities). Elffers et al found that 69 percent of known evaders (detected and pursued by the authorities) denied non-compliance in the researchers' independent study.³⁷

³⁷ Elffers H., Weigel, R. H. and Hessing, D. J., (1987), "The Consequences of Different Strategies for Measuring Tax Evasion Behavior", cited by Hessing, D. J., Elffers, H., Robben, H. S. and Webley, P. (1992), "Does Deterrence Deter? Measuring the Effect of Deterrence on Tax Compliance in Field Studies and Experimental Studies", pp. 293 - 295. The 1987 paper was also published as, Hessing, D. J., Elffers, H. and Weigel, R. H., (1988), "Exploring the Limits of Self-Reports and Reasoned Action: An Investigation of the Psychology of Tax Evasion Behavior". With respect to these researchers' findings as reported in the 1992 version, it is to be noted that in the Netherlands, the tax system does not provide a strong deterrent to evasion from an economic point of view. Tax inspectors simply correct tax returns which are perceived to be incorrect, and then collect the extra tax found to be owing. Punishment only becomes possible if, after the detection and correction process has

A second limitation Hite et al recognised in their approach was closely related to the first limitation, outlined above. It was not clear whether the subjects were universally *able* to ‘put themselves in the role’.³⁸ This was regarded, however, as a minor shortcoming as the issue at stake was the directionality of choice rather than absolute choice of the dollar amount of income reported in the experiment.

The third limitation involved the essentially unknowable nature of the jurors’ actual neutral reference points: the experiment was conducted on the simplifying assumption that taxpayers view any refund as a gain and any terminal tax payable in terms of a loss; but long experience of a given level of refund over time may shift the reference point in a taxpayer’s mind to that given traditional level of refund.

Hite et al perceived a further possible confounding factor in their failure to present income and tax liability thresholds in the scenarios that would be necessarily meaningful to the jurors. A related flaw, however, is much more cogent. No detection rates or penalty amounts were specified in the study and subjects were merely told that the secondary income was not reported to the IRS and that it would be difficult to trace. In their paper published in 1993, White, Harrison and Harrell³⁹ suggested the lack of significance of Hite et al’s findings was explicable in terms of the implied minute probability of detection, since, according to *Prospect Theory*, extremely low probabilities are treated as indistinguishable from zero and discarded.⁴⁰ White et al noted that a framing effect which might possibly have a significant impact on tax compliance when detection probabilities are set at moderate levels could well be insignificant when the rate of detection is set at an extremely low level.⁴¹

taken place, inspectors become aware of further information providing a clear indication of an *intention* to evade taxes, (p. 298).

³⁸ Hite, Jackson and Spicer, (1988), op. cit. n. 35, p. 12.

³⁹ White, R. A., Harrison, P. D. and Harrell, A., (1993), “The Impact of Income Tax Withholding on Taxpayer Compliance: Further Empirical Evidence”.

⁴⁰ Kahneman, D. and Tversky, A., (1979), “Prospect Theory: An Analysis of Decision under Risk”, p. 275.

⁴¹ White, Harrison and Harrell, (1993), op. cit. n. 39, p. 67.

3.3.4 Rates of Voluntary Compliance

The second paper, Cox and Plumley (1988), also furnished evidence of a framing effect based on the prior tax withholdings as the neutral reference point. This paper was an in-house Internal Revenue Service production; and the sample in this instance consisted of 50,000 American tax returns for the year 1982 provided by the IRS' TCMP Survey of Individual Returns.

Defining their variable, *rate of voluntary compliance* as actual tax obligation disclosed and paid as a percentage of an individual's IRS-estimated total tax liability, Cox and Plumley found a consistent relationship between voluntary compliance and the size and sign of terminal tax balances. If the latter was viewed as a spectrum ranging from large refunds at one end, diminishing to zero in the middle, and from there increasing as bills to be paid from small to large at the other end, movement through the spectrum was matched by a tandem movement in the variable from relatively high *voluntary compliance* to significantly less compliance. Salary and wage earners requesting refunds larger than \$US1,000 were 96 percent voluntarily compliant whereas the rate dropped to 89 percent for their confreres who disclosed more than a \$US1,000 balance due. This drop was even greater when Cox and Plumley turned their attention to what self-employed individuals had recorded in their tax returns. While the rate of *voluntary compliance* was 95 percent for business people claiming refunds of more than \$US1,000, the rate for those recording the same magnitudes of terminal tax to be paid was only 70 percent.

These results were independently endorsed when Cox made the 1982 TCMP data available to Chang and Schultz (1990) who, apparently unaware of Cox and Plumley's paper, reused it to test the hypothesis⁴²:

H_A: Taxpayers with an over-withheld (refund due) position have a higher rate of actual tax compliance than taxpayers with an under-withheld (balance due) position.

⁴² Chang, O. H. and Schultz, J. J., (1990), "The Income Tax Withholding Phenomenon: Evidence from TCMP Data", p. 88.

Like Cox and Plumley, Chang and Schultz identified their salient variable as the *voluntary compliance rate*. When the returns were sorted into categories according to whether the TCMP thought a refund or terminal tax was due, the rate of voluntary compliance dropped smoothly from 97.1 percent to 95.3 percent as the size of claimed return dropped from greater than \$US1,000 to less than \$US100. This smooth decline in voluntary compliance continued as tax due increased from less than \$US100 to more than \$US1,000 — dropping from 95.1 percent to 90.1 percent.⁴³ Chang and Schultz then conducted a three-way ANOVA to determine if the withholding effect remained statistically significant once the variables, *source of income* and *adjusted gross income* were factored out. They found that there was a significant interaction between *source of income* and *withholding position*; and when they re-ran their initial test with the returns separated also by source of income, their results were almost identical with Cox and Plumley's.

Chang and Schultz did, however, feel there was a major limitation in exclusive reliance on TCMP data. While the results showed there was a strong relationship between the withholding phenomenon and the voluntary compliance variable, the direction of causality could not formally be identified.⁴⁴

3.3.5 Laboratory Analysis

After Cox and Plumley, the next investigation of decision frames with respect to compliance was back in the laboratory. Schadewald (1989)⁴⁵ set out to distinguish the gain/loss framing effect based on withholdings from the confounding factors that a more general empirical validation of the *Prospect Theory* value function involved. His two experiments provided scenarios containing a context of pre-existing financial plans and conditions which the subjects were expected to absorb and utilise in deciding how to file a set of income tax returns. His subjects were MBA students from the University of Minnesota.

⁴³ Ibid, p. 89.

⁴⁴ Ibid, p. 93.

⁴⁵ Schadewald, M. S., (1989), "Reference Point Effects in Taxpayer Decision Making".

Schadewald's first experiment was designed to determine whether taxpayers adopted these prior expectations of tax liabilities as the neutral reference point of their decision frames, or if they made use of a decision frame in which the neutral reference point was the current cash position at tax return filing time instead. He found that neither prior expectation nor current cash position was identified as the relevant decision frame by an outright majority of subjects.

Schadewald processed the decisions the subjects made in terms of two hypotheses⁴⁶:

- H₁:** More subjects will be risk seeking when the refund is overestimated than when it is underestimated. (This related to the use of a decision frame based on prior expectations.)

- H₂:** More subjects will be risk seeking when they are underwithheld than when they are overwithheld. (This related to the use of a decision frame based simply on current cash position.)

Schadewald found that the proportion of his subjects who opted for the risk-seeking tax position ranged only from 40 percent to 48 percent; and that there was no support for **H₁**. With respect to **H₂** he found that only 34 percent of his subjects were influenced by the manipulation of provisional tax withheld. Yet the responses to four non-contextual decision problems given as fillers between the two scenarios showed strongly that the subjects were risk averse for gains and risk seeking when facing the prospect of a loss. From this, Schadewald concluded that the experiment afforded no support for the hypothesis that a taxpayer's provisional tax withholding position alone significantly influences how the tax-filing decision is framed and what risk preferences are adopted.⁴⁷

This finding was also borne out in the results from an investigation along another avenue. Schadewald required his subjects to register, on a nine-point Likert scale, how strongly they preferred each of the prospects they had chosen. Given there was both a risky and a riskless prospect per decision problem, the range 1 - 4 indicated a decreasing preference for the riskless

⁴⁶ Ibid, p. 71.

⁴⁷ Ibid, p. 79.

option, while the range 6 - 9 was indicative of an increasing preference for the risky option; and the mid-point (5) represented indifference between the prospects. The mean preference ratings, when subjected to analysis of variance testing (ANOVA), showed that neither the effect of the prior expectation nor the effect of the amount withheld was significant.⁴⁸

In his second experiment Schadewald investigated the issue of explicit labelling of prospects — did they have a significant impact on the formation of a taxpayer's decision frame and the associated development of risk preferences? To answer this, he gave 130 MBA students from the University of Texas two further scenarios containing quite different contextual information, but which led to a choice among prospects containing the same probabilities and monetary outcomes as before. Schadewald's two hypotheses in this instance were⁴⁹:

H₃: More subjects will be risk seeking when prospective tax refunds are described as deviations from a prior expectation (which implies the outcomes are losses) than when the refunds are described as deviations from the current cash position (which implies the outcomes are gains).

H₄: The effect hypothesised in **H₃** will be stronger when the outcomes are also explicitly labelled as 'gains' or 'losses'.

Schadewald found little empirical support for **H₃**; but the evidence supporting **H₄** was statistically significant. Once outcomes were specifically labelled as *gains* or *losses*, his manipulation of the neutral reference point did have a stronger impact on the subjects.⁵⁰ However such explicit usages of gain and loss terminology can only occur in artificial circumstances. Furthermore, when the preference ratings associated with the second experiment's scenarios were subjected to ANOVA, the interactions predicted in **H₃** and **H₄** turned out to be statistically insignificant.⁵¹ Schadewald therefore concluded that *Prospect*

⁴⁸ Ibid, p. 74.

⁴⁹ Ibid, p. 76.

⁵⁰ Ibid, p. 77.

⁵¹ Ibid, p. 78.

Theory's concept of a decision frame containing a neutral reference point was “of little to no importance in the context of tax compliance.”⁵²

This conclusion rests, however, on the assumption that the 70 percent detection rate used in the study was still low enough not to dominate any noncompliance behaviour related to the level of provisional tax withheld. This assumption was considered in some depth by Schepanski and Kelsey (1990).

3.3.6 Improvement in the Realism and Assumptions of Scenarios

Schepanski and Kelsey (1990)⁵³ developed upon Schadewald's ideas by comparing the witting decision made to evade tax by claiming a non-allowable expense deduction in three different contexts — a *loss context*, a *refund context*, and a *final asset context*. Unlike Schadewald, they found significant framing effects; and also they determined that the strength of a framing effect may not be independent of the probability associated with the risky prospect offered in the concomitant decision problem. As the probability of detection rose to 45 percent, a framing effect which had been quite apparent when the detection probability was only 20 percent, tended to disappear. This phenomenon possibly accounts for why Schadewald's decision problems, with risky prospects containing a 70 percent disallowal probability, did not furnish evidence of significant framing effects.

Schepanski and Kelsey designed their experiment to monitor reactions to a \$US500 drop in funds. Associated with each of the three contexts were decision problems containing a riskless prospect in which the taxpayer bore the \$US500 drop, and a risky prospect in which the probability of detection was varied. The two researchers used separate groups of Iowa University undergraduate commerce students to work within each of the loss, refund and final asset position contexts.

⁵² Ibid, p. 80. Schadewald also noted, with respect to his first experiment, “...[that] the ineffectiveness of the reference point manipulations is consistent with utility theory and inconsistent with prospect theory.” (p. 76.)

⁵³ Schepanski and Kelsey, (1990), op. cit. n. 34.

When they compared results from the *loss* and *refund contexts*, Schepanski and Kelsey found that the *refund context* subjects exhibited a greater degree of risk aversion than did the subjects making choices in the *loss context*; but as the probability of detection increased, both the difference and the statistical significance of this difference declined.

Further, like Schadewald, Schepanski and Kelsey required their subjects to indicate on a nine-point Likert scale the strength of their risk preference. When they compared the mean preference ratings of the *loss context* and *refund context* subjects over the five decision problems, they found again that the refund context subjects were significantly more risk averse.

When Schepanski and Kelsey compared the choices made by the *loss context* subjects with those made by the *final asset context* subjects, they found the *loss context* subjects tended to be more risk-seeking; but the differences were of weaker statistical significance than in the *loss* versus *refund* comparison. Nevertheless this constituted evidence of significant framing differences. This was corroborated by evidence provided from a comparison of the *loss* and *final asset context* mean preference ratings. In other words, the wording of the scenario had an impact on how a subject framed the decision problems.

However, when Schepanski and Kelsey considered the nature of the responses made in each of the three contexts in isolation,⁵⁴ support for *Prospect Theory* in a tax compliance setting became much less clear-cut. The evidence from the *refund context* confirmed the theory's prediction of risk aversion; but the evidence from the *loss context* clashed with the theory's prediction of risk-seeking behaviour. When the probability of detection rose above 20 percent the loss scenario subjects exhibited mildly risk-averse behaviour. Dusenbury (1994)⁵⁵ asserted that Schepanski and Kelsey had confounded the decision frame effect with a variable denoting differences from expectation since each scenario involved an unexpected shift in tax-related circumstances.

⁵⁴ Ibid, p. 67. Schepanski and Kelsey noted that evidence from the *final asset context* could not be considered to be directly diagnostic of *Prospect Theory* per se on the ground that the value function was defined here in terms of final wealth states.

⁵⁵ Dusenbury, R., (1994), "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options".

3.3.7 International and Cross-Cultural Extensions

The next study was not so much a one-off investigation of the withholding-based framing effect, but a related family of studies which facilitated international and cross-cultural comparisons for the first time in this branch of compliance research— it was conducted by a group based primarily in Rotterdam in the Netherlands. Robben, Webley, Elffers and Hessing (1990)⁵⁶ hid their investigation of the framing effect of underwithheld and overwithheld provisional tax inside a two-period small business simulation conducted in 1988/89 in order to minimise what they termed the effect of ‘self-presentational concerns’ with regard to honesty in tax paying⁵⁷ — an issue Hite, Jackson and Spicer (1988) listed as a factor confounding results in their 1988 working paper.⁵⁸ The idea was that if simulation participants were not aware of a tax focus of what they were doing, then they would not make any effort to mask their true risk preferences when filing the incorporated tax return. The participants were 71 Dutch undergraduate students (most majoring in economics).

Robben et al gathered data relating to two quantitative measures: the frequency of tax fraud, and the amount of tax in Dutch guilders actually evaded. In addition they classified subjects into three groups according to their stated perceptions of their tax positions (*perceived payment*, *neutral position*, and *perceived refund*) after their first and second tax declarations as disclosed in a questionnaire at the end. With the sample partitioned in this fashion, Robben et al found, upon application of a one-way analysis of variance test, that participants who recalled having to make an extra payment evaded tax significantly more often in both periods than those who recalled receiving a refund.⁵⁹ But this result was based on groups ranging in size from 31 down to 11; and the researchers noted that one sixth of the group over-all could not recall in the

⁵⁶ Robben, H. S. J., Webley, P., Elffers, H. and Hessing, D. J., (1990), “Decision Frames, Opportunity and Tax Evasion: An Experimental Approach”.

⁵⁷ Ibid, p.360.

⁵⁸ Hite, Jackson and Spicer, (1988), op. cit. n. 35, p. 15: “First is the limitation of all research involving hypothetical choices: that we cannot know how these taxpayers would have acted on April 15 with their own returns and checkbooks in front of them.”

⁵⁹ Ibid, p. 359. $F = 2.43, p < 0.1$ with respect to the first declaration; $F = 3.28, p < 0.05$ with respect to the second declaration.

questionnaire whether they had made payments or received refunds. Therefore they regarded support for framing effects as positive, but weak.

A replication of this experiment on a sample of 88 undergraduates, with what were thought in advance to be minor modifications to the script, was conducted in 1989 and published by Webley, Robben, Elffers and Hessing (1991). The withholding decision frame variable was found this time to be significant at the 5 percent level.⁶⁰ The main difference between the experiment and its replication lay in the exact nature of items deductible for tax purposes in the business simulation. The original experiment presented unambiguous information as to what was deductible, the replication allowed for situations in which participants could find themselves expecting deductions which they *then* found were not legitimate.

Webley, Robben, Elffers and Hessing (1991) also published the results of a revised second replication, which was also conducted in 1989, this time on 85 Dutch taxpayers recruited from the Rotterdam phone book and from an adult education course in that city. In this version the taxpayers were confronted with their refunds receivable or terminal taxes payable so that they could have the opportunity to change their tax returns before officially filing them. This meant that the participants would be strongly conscious of their ultimate tax-filing options. No significant framing effect was found in this version of the experiment; and only half of the participants indicated that they had recognised the implications of their withholding situation as the researchers had intended. Yet in a concurrent British replication, on samples of students and non-student taxpayers, also conducted by Webley et al in conjunction with the University of Exeter and reported in the same publication as the researchers' two Dutch replications, a framing effect significant at the 5 percent level was detected. The researchers asserted that the minor adjustments necessary for grooming a Dutch experiment for English participants were not sufficient to violate the comparability of the replications. Hence the discrepancy in the results remained puzzling.

⁶⁰ Webley, P., Robben, H., Elffers, H. and Hessing, D., (1991), *Tax Evasion: An Experimental Approach*, p. 104. This replication is written up as NL03. (The 1990 experiment was also republished in this book as NL02.)

The puzzle deepened when further concurrent replications were conducted at three universities in the United States,⁶¹ the Stockholm School of Economics, the Universidad Complutense in Madrid and at the Free University of Brussels by Robben, Webley, Weigel, Warneryd, Kinsey, Hessing, Alvira Martin, Elffers, Wahlund, Van Langenhove, Long and Scholz (1990).⁶² The pooled sample (including subjects in the Dutch and British studies discussed above) contained 694 subjects in six different countries, of whom 547 filed accurate returns, 20 overpaid their tax by what was determined to be a significant amount (and who were discarded from further analysis), and of whom 127 significantly underpaid their tax. The framing effect hypothesis, repeated here for convenience, was supported at the 5 percent level of significance⁶³:

H_A: That subjects will engage in more tax cheating when low withholding taxes create a loss situation than when high withholding on the same tax liability creates a gain or refund situation.

Nevertheless the individual replications themselves failed to find significant effects for withholding status.⁶⁴ The researchers advanced the possibility that the subjects not being told that extra tax payments constituted a loss and not having any means of comparing personal performance with those of other subjects may have been a contributing factor.

3.3.8 Other 1990s Papers

Unlike Robben et al, Martinez-Vasquez, Harwood and Larkins (1992)⁶⁵ presented their subjects with tasks which addressed, in a direct and overt fashion, the issue of paying tax. Martinez-Vasquez, Harwood and Larkins tested the validity of prior withholdings-based decision frames in a set of experiments designed to broaden and deepen their understanding of Schadewald's

⁶¹ Syracuse, Stony Brook New York and the University of Illinois.

⁶² Robben, Webley, Weigel, Warneryd, Kinsey, Hessing, Alvira Martin, Elffers, Wahlund, Van Langenhove, Long, and Scholz, (1990), *op. cit.* n. 36.

⁶³ Robben, H. S. J., Webley, P., Weigel, R. H., Warneryd, K., Kinsey, K. A., Hessing, D. J., Alvira Martin, F., Elffers, H., Wahlund, R., Van Langenhove, L., Long, S. B. and Scholz, J. T., (1990), "Decision Frame and Opportunity as Determinants of Tax Cheating", p.347.

⁶⁴ *Ibid*, p.360.

⁶⁵ Martinez-Vasquez, J., Harwood, G. B. and Larkins, E. R., (1992), "Withholding Position and Income Tax Compliance: Some Experimental Evidence".

(1989)⁶⁶ finding that experimental subjects (and by extension, taxpayers) do not think in terms of a gain or loss measured from a neutral reference point determined by prepayments unless the gains or losses are explicitly labelled. In particular Martinez-Vasquez et al investigated the impact of an unintended or unexpected terminal tax liability (resulting from accidental underwithholding in terms of provisional tax forwarded to the tax authorities), relative to an expected liability of the same size resulting from an intentional level of underwithholding. Since the expected liability and unexpected liability situations were each encapsulated in both low detection/high penalty and high detection/low penalty rate scenarios, there were four matched scenarios; and each participant was assigned one only.⁶⁷

Upon subjecting their data to Chi-square testing, Martinez-Vasquez et al found no statistically significant difference between the behaviours elicited in the two situations. The subjects, who were MBA students at a university in the south-east of the United States, tended to choose the option associated with the highest expected value in every scenario — a result which did not vary whether a low probability of detection of evasion was paired with a high penalty, or a high detection probability was paired with a low penalty.

Martinez-Vasquez et al also examined the impact of a liquidity variable on taxpayer behaviour given expected and unexpected terminal tax balances due as a result of the nature of prior withholdings.⁶⁸ Two sets of subjects had liquidity in that they were given scenarios in which they were told they had the money in hand to pay their tax bill (expected or unexpected) and another two sets were told they did not have the cash to pay the liability and hence were illiquid. As in the researchers' first investigation, two matched pairs of scenarios were provided, differing in terms of the liquidity divide and in terms of either a three percent risk of evasion detection or no risk. A significant difference (at the five percent level) was detected only between the liquid and illiquid scenarios containing the three percent detection risk. The

⁶⁶ Schadewald, (1989), op. cit. n. 45. See Subsection 3.3.5 of this chapter.

⁶⁷ Ibid, pp 160 - 161.

⁶⁸ Ibid, p. 164. In their investigation of taxpayer liquidity, Martinez-Vasquez et al dropped consideration of expected and unexpected refunds because in this instance, decision frames involving perceived gains were irrelevant.

researchers did note, however, there was a possibility the low detection risk was deemed by subjects to be so minimal that measurement of the liquidity variable was confounded by an unmeasured *opportunity to evade without risk* variable.⁶⁹

Martinez-Vasquez, Harwood and Larkins also investigated the impact of a fiscal illusion variable. This involved checking the significance of a simple manipulation of the mode of payment allowed on a tax balance due (as a result of underwithholding): one group had to pay a lump sum while another group could spread payments out over the next year in the form of an increased rate of withholdings. No significant impact on the subjects' propensity to evade was detectable.

Although the Martinez-Vasquez et al paper was published two years in advance of Dusenbury (1994), Dusenbury appears to have been unaware of it, his paper having been substantially written before an initial submission date of August 1990.

However, Dusenbury was aware of, and cited, White, Harrison and Harrell (1993). Their study also investigated what happens when detection and penalty rates are varied with respect to each other. They found much stronger evidence in support of the influence of a provisional tax withholding framing effect.

Running their experiment on 81 full-time workers studying part-time at a university in the American south-east and a second sample of 175 undergraduate business students, they tested the basic *Prospect Theory* hypothesis⁷⁰:

H₁: Taxpayers are more likely to exhibit risk-seeking behaviour and claim an unsupported tax deduction in a tax due situation than in a tax refund situation.

⁶⁹ Ibid, p. 164.

⁷⁰ White, Harrison and Harrell, op. cit. n. 39, p.68. Because they used separate samples of full-time workers and students, White et al were also able to compare the performance of the two groups. They found the students to be reliable as surrogates for more experienced taxpayers in the analysis of the relative effects of behavioural variables; but that they were inappropriate if an analysis of absolute effects was called for. The students were found to be less aggressive than the experienced tax payers in full employment. This result conflicts with a finding on the same issue published by the international team of eleven researchers, Robben et al (1990), discussed earlier in this chapter. In the international team's pooled sample, it was found that students tended to be more likely to underpay taxes and do so by larger amounts.

Each participant was given nine scenarios to work through in which three levels of detection rate (5%, 30%, 55%) were matched against three different penalty levels (10%, 50%, 90% of tax evaded). This enabled White et al to test⁷¹:

H₂: As the audit detection rate increases, taxpayers are less likely to claim an unsupported tax deduction.

H₃: As the penalty amount increases, taxpayers are less likely to claim an unsupported tax deduction.

All three hypotheses were strongly supported: Chi-square testing indicated significance for all three at the $p < 0.0001$ level. There was also a significant three-way interaction among provisional tax withheld, detection probability and the size of penalty. This, White et al argued, indicated that a withholding framing effect existed; and that it had a greater impact when the expected values of outcomes were closer together than when relatively more dispersed. They found significant framing effects in seven of the nine detection rate/penalty size combinations — a finding that was robust when outcomes were both framed as gains or as losses.⁷²

3.4 FRAMING EFFECTS AND PROFESSIONAL TAX PREPARERS

Several studies have extended decision frame research to include consideration of the impact of tax withholdings on the relationship between professional tax return preparers and their clients.

Sanders and Wyndelts (1989)⁷³ investigated the impact on professional preparers of the framing effect relating to whether or not a terminal tax bill or refund was expected for a client. While

⁷¹ *Idem*.

⁷² *Ibid*, p. 75. See Table 5 in White et al's study. Only when a detection rate of 5% was combined with the lower two penalty sizes was no significant framing effect detected.

⁷³ Sanders, D. L. and Wyndelts, R. W., (1989), "An Examination of Tax Practitioners' Decisions Under Uncertainty".

Sanders and Wyndelts found evidence to support Kahneman and Tversky's (1979) *certainty effect*, they were unable to detect conclusive evidence of an influential decision frame.

Withheld provisional tax was shown to induce a significant framing effect on tax preparers' clients in a doctoral dissertation: Schisler (1992).⁷⁴ Without reference to professional advice, Schisler's client subjects were risk averse in the context of decision frames mapping gains and relatively more risk seeking where the frames mapped losses. But when Schisler investigated the impact on tax preparers, the posited framing effect vanished. Preparers were not more risk averse when their clients enjoyed prospects of large refunds than they were risk seeking when their clients faced the possibility of large tax bills. However, Schisler did find statistically significant support for a subtly different framing effect. It turned out that professional tax preparers were significantly influenced by the clients' degree of aggressiveness.⁷⁵ It is the client's degree of aggressiveness alone which is influenced by the framing effect of withheld tax; the preparers' concern was that contradiction of a client's wishes would result in a loss situation for the preparers themselves in terms of future custom lost.

Schisler's results implied that professionals with aggressive clients will therefore accept high levels of risk with respect to challenge and even prosecution by the tax authorities. Schisler (1995)⁷⁶ confirmed this result on a further set of actual professionals and proxy clients. But this client aggressiveness variable, albeit endogenous with respect to withholding-related framing effects, is irrelevant in a replication of Dusenbury (1994).

The next chapter is given over to an analysis of Dusenbury (1994), Dusenbury's hypotheses, and further hypotheses which made themselves apparent during the reading of the literature.

⁷⁴ Schisler, D. L., (1992) "An Experimental Examination of Taxpayers' and Tax Professionals' Aggressive Tax Preferences — A Prospect Theory Foundation".

⁷⁵ Ibid, p.42. The hypothesis here was: "H5: Tax professionals will be more aggressive when the client is more risk seeking and less aggressive when the client is conservative." Schisler used 125 Memphis State University MBA students as proxies for the clients of professional tax preparers, and ran his experimental scenarios on a sample of 119 professionals recruited from nine firms of chartered accountants operating in the Mid-west and in the South-east of the United States of America.

⁷⁶ Schisler, D. L., (1995), "An Experimental Examination of Factors Affecting Tax Preparers' Aggressiveness — A Prospect Theory Approach".

4. HYPOTHESES AND RESEARCH DESIGN

4.1 INTRODUCTION

At this point we are ready to analyse the object of this New Zealand replication, Dusenbury (1994), and move on to stating the precise objectives of the current research. Because any consideration of Dusenbury's hypotheses must inevitably be linked closely to an analysis of his research design, both are dealt with in this chapter. The research aims of the replication are closely associated with the replication's experimental design; so again, the aims and design are contemplated in tandem.

The prime purpose of the replication is to provide a cross-cultural comparison between North American and New Zealand taxpayers with respect to their risk preferences when preparing to file their tax returns. It is noted that changes to the current study's experimental instrument make the replication a conceptual replication rather than a pure one. Once the replication hypotheses are tabled, attention is turned to a number of extensions.

The chapter is divided into the following sections. Section 4.2 records Dusenbury's (1994) objectives and results. Section 4.3 then considers the academic relevance of replications in general, and a New Zealand replication and extension of Dusenbury's experiment in particular. The current study's hypotheses H_{1A} , H_{2A} and H_{3A} , which are to be examined in the replication are stated in Section 4.4; and Section 4.5, the final section, tables hypotheses H_{4A} to H_{9A} , which are examined in the extensions. Description of the research method is left to Chapter Five.

4.2 DUSENBURY (1994): HIS RESEARCH OBJECTIVES AND RESULTS

4.2.1 Dusenbury's Objectives and Research Design

Dusenbury (1994) investigated the decision frame phenomenon with respect to tax withholdings from a subtly different angle from those of earlier studies; but first it would be useful to provide an overview. In common with the experimental style of White, Harrison and Harrell (1993)¹ and Schepanski and Kelsey (1990),² Dusenbury held the level of income, tax rates and tax liability constant over three terminal tax-filing scenarios. He adopted this approach in order to investigate if shifts among different levels of provisional tax withholdings would be absorbed by participants as meaningful shifts of decision frame, with concomitant switches between risk aversion and risk willingness. It is important to note that Dusenbury set out to eliminate all possible alternative frames of reference so that participants in his experiment would potentially be influenced by one frame alone — the frame produced by the level of prior tax withholdings.³ His three tax scenarios were labelled for convenience: the *low pay* case in which, depending on what option the participant chose, a relatively small amount of tax had to be paid or even a refund claimed; the *high pay* case in which all the possible levels of reported income required payments of terminal tax; and the *refund* case in which all the options involved refunds.

In the *refund* and *high pay* tax cases, Dusenbury provided an unspecified item of income for which the tax treatment of associated expenses was in doubt. If a participant chose totally to ignore these expenses in filing his or her tax return, he or she would receive (in the *refund* case) a relatively low, but safe, refund, unchallenged by the tax authorities. If it was the *high pay* case, he or she would have to pay the largest stated terminal tax payment — again

¹ White, Harrison and Harrell, (1993), "The Impact of Income Tax Withholding on Taxpayer Compliance: Further Empirical Evidence".

² Schepanski and Kelsey, (1990), "Testing for Framing Effects in Taxpayer Compliance Decisions".

³ Dusenbury, R., (1994), "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options", p. 2.

TABLE 4.1
Summary Table Provided for *High Pay* Tax Case X.⁵

TAX FILING OPTIONS -- Circle One					
	(A)	(B)	(C)	(D)	(E)
AMOUNT DEDUCTED	NONE	SOME	HALF	MOST	ALL
PAYMENT DUE	PAYMENT	PAYMENT	PAYMENT	PAYMENT	PAYMENT
	\$800	\$700	\$600	\$500	\$400
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

unchallenged by the tax authorities. However, the expenses associated with the earning of the unspecified income item were implied to be not insubstantial; and in his training script, Dusenbury told his participants that, in choosing such a course of action (Option **A**, the riskless option), they might be paying too much tax.⁴ The alternative risky prospects involved claiming an expense deduction on the unspecified income item which, from Option **B** to Option **E**, became an increasingly larger proportion of the size of the unspecified income item itself. Also, with the increase in size of deduction claimed, there was a concomitant increase in the likelihood of a challenge from the tax authorities. The challenge risk rose from a 15 percent probability in Option **B** to 40 percent in Option **E**. This information was summarised for participants in a table at the foot of each scenario. For convenience, the tables for the *high pay* case and the *refund* case are reproduced here as Tables 4.1 and 4.2.

⁴ Dusenbury's sample case instructions contained this wording: "In Option A you include the total earnings. There is no risk of under-reporting your tax (though you may be overpaying)." See Appendix A.1.

⁵ See Appendix A.2.

TABLE 4.2
Summary Table Provided for *Refund Tax Case Y*.⁷

TAX FILING OPTIONS -- Circle One					
	(A)	(B)	(C)	(D)	(E)
AMOUNT DEDUCTED	NONE	SOME	HALF	MOST	ALL
REFUND CLAIMED	REFUND	REFUND	REFUND	REFUND	REFUND
	\$400	\$500	\$600	\$700	\$800
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

Given the above summary information, a constant level of income at \$27,780, a level of withheld tax in the *high pay* case of \$4,200, and a level of withheld tax in the *refund* case of \$5,400, Dusenbury's first hypothesis was as follows⁶:

H_{1A}: Participants will choose riskier options in the *high pay* case than in the *refund* case.

A vital aspect of Dusenbury's work was that he conducted both *within-subject* and *between-subject* testing. The *within-subject* approach allowed shifts of risk preference within an individual to be mapped. This was a distinct advance upon simple comparison of group means, and meant it was not necessary, as in Schepanski and Kelsey, to split the participants into mutually exclusive streams, each facing a different condition.

A logical consequence of *within-subject* testing was the requirement that the decision problems had to be directly comparable with each other. Dusenbury achieved inter-scenario comparability by jettisoning the use of identical expected values across the set of five prospects within a scenario in favour of establishing uniformity of expected values across equivalent prospects in the set of three scenarios.

⁶ Dusenbury, R., (1994), op. cit. n. 3, p. 6.

⁷ See Appendix A.2.

TABLE 4.3
Construction of Tax Cases in Dollars.

Cases	Option A	Option B	Option C	Option D	Option E
Panel A: The Cases as Linear Transformations[†]					
HIGH PAY	(-800, 1.0; 0, 0)	(-700, 1.0; -400, 0.15)	(-600, 1.0; -500, 0.25)	(-500, 1.0; -600, 0.33)	(-400, 1.0; -700, 0.40)
Shift	<u>+(600, 1.0)</u>	<u>+(600, 1.0)</u>	<u>+(600, 1.0)</u>	<u>+(600, 1.0)</u>	<u>+(600, 1.0)</u>
LOW PAY	(-200, 1.0; 0, 0)	(-100, 1.0; -400, 0.15)	(0, 1.0; -500, 0.25)	(100, 1.0; -600, 0.33)	(200, 1.0; -700, 0.40)
Shift	<u>+(600, 1.0)</u>	<u>+(600, 1.0)</u>	<u>+(600, 1.0)</u>	<u>+(600, 1.0)</u>	<u>+(600, 1.0)</u>
REFUND	(400, 1.0; 0, 0)	(500, 1.0; -400, 0.15)	(600, 1.0; -500, 0.25)	(700, 1.0; -600, 0.33)	(800, 1.0; -700, 0.40)
Panel B: Expected Values[‡]					
HIGH PAY	-800	-757	-725	-699	-680
LOW PAY	-200	-157	-125	-99	-80
REFUND	400	443	475	499	520
Panel C: Variances^{††}					
ALL CASES	0	20,400	46,875	80,000	117,600
<p>†In Panel A, the options are stated in the form $(x, p; y, q)$ where: x = payment due or refund associated with the taxable income reported, p = 1.0, (probability of the pay due or refund per reported taxable income), y = total cost of a subsequently detected understatement, q = the probability of a subsequently detected understatement.</p> <p>‡Expected value for any option $(x, p; y, q) = (xp) + (yq)$. Example: the expected value for High Pay, Option C is: $[-600 \times 1.0] + [-500 \times 0.25] = -600 - 125 = -725$.</p> <p>††Variances are $(y^2q) - (yq)^2$. Example: the variance for Option B in all levels is: $[-400^2 \times 0.15] - [-400 \times 0.15]^2 = 24,000 - 3600 = 20,400$.⁸</p>					

This approach meant that the five prospects in each scenario contained unequal expected values, increasing progressively from option A, which was riskless, to the riskiest option, option E. The outcomes of each prospect and their associated probabilities are formally laid out in Panel A of Table 4.3. Note in Table 4.3 that the initial choice made by a participant is associated with a probability of 1.0. This simply records the fact that the initial outcome, (x, p)

⁸ Dusenbury is using a variance formula which is the binomial formula for variance multiplied by a factor of y . The standard deviation of a binomial distribution, according to Ott, L. and Mendenhall, W., (1985), *Understanding Statistics* (Fourth Edition), p. 143, is $\sigma = \sqrt{[yq(1 - q)]}$. Therefore the variance of the distribution will be $[yq(1 - q)] = yq - yq^2$. By this formula the variance of the absolute value of the costs of the options are as follows: (B) = 51, (C) = 93.75, (D) = 132.66, (E) = 168.

is absolutely certain. However there is a delayed second part to the prospect, (y, q) representing the tax authority's decision to challenge or not to challenge the return. The expected value of each option is shown in Panel B of Table 4.3.

With reference to the axiomatic invariance of preference orderings to positive linear transformations according to *Expected Utility Theory*,⁹ Dusenbury made the point¹⁰:

It is important to understand the selection of a different option in any two cases constitutes a preference reversal.

Comparability of Dusenbury's scenarios was facilitated by making the expected values of each option A constant across all scenarios (and then B and so on); and then he disguised the equivalence by shifting the A-to-E set in each scenario by an increment of \$600. Since this was an additive linear transformation, the variance of each option was preserved across the spectrum of scenarios. This is shown succinctly in Panels B and C of Table 4.3. Dusenbury enlarged upon the significance of making each scenario's options the linear transformation of the options in the other scenarios¹¹:

The proof that preferring option C in the high pay case and option B in the refund case is inconsistent with expected utility theory depends on the existence of a functional representation of preference orderings. Let a greater than ($>$), or less than ($<$), symbol indicate preference. From the perspective of functional representation, if option C (high pay) $>$ option B (high pay), but option C (high pay) + \$1,200 $<$ option B (high pay) + \$1,200, then the slope of the function describing the preference ordering has changed signs due to the addition of a constant term. However, the slope of a function is unaffected by adding a constant term.

While Dusenbury provided three tax scenarios, he restricted his first hypothesis to consideration of the *high pay* and *refund* cases only. The reason for dropping the *low pay* case from consideration was that it was a mixed lottery which bridged the neutral reference point in

⁹ Dusenbury cited Luce and Raiffa (1967), *Games and Decisions*, p. 30. However this material has been covered in Chapter 2, Subsection 2.3.1 of the current study. See axiom C:a.

¹⁰ Dusenbury, (1994), op. cit. n. 3, p. 7.

¹¹ Ibid, pp. 7 - 8.

the value function. This meant there would be no obtainable meaningful result if the slope of the function describing ordinal preferences in this case was compared with those of preference functions associated with the other two cases, neither of which bridged the neutral reference point.¹²

Dusenbury also provided two non-tax scenarios in order to test framing effects further by altering the decision setting. One of these was a *medical insurance* case, chosen because the purchase of this kind of insurance involves a decision context known for risk aversion. The funds flow and the risks and payoffs in this case were identical with those of the *high pay* tax case. This gave rise to the second hypothesis¹³:

H_{2A}: Participants will choose less risky options in the health insurance case than in the *high pay* tax case.

In effect, the question asked in H_{2A} was: given that a framing effect existed with respect to the level of withholdings in the *high pay* tax case, could this framing effect be distinguished from that of a frame founded upon future money worries concerning future possible medical crises?

The second non-tax scenario was chosen for its association with risk-taking behaviours. It was a virtually contextless set of five gambling options, reminiscent of the two-prospect problems investigated by Kahneman and Tversky (1979),¹⁴ matching option for option, the risk and payoff combinations offered in the *low pay tax case*. Effectively, Dusenbury set up a *Prospect Theory* paradigm approximation of Baldry's (1986) experiment.¹⁵ The third hypothesis therefore was as follows over leaf.¹⁶

¹² Ibid, p. 9, Footnote 8.

¹³ Ibid, p. 7. In the medical insurance scenario, there was no equivalent of withheld tax, since for the taxpayer/citizen, the chosen policy's premium was made as a single annual payment. Also there was no given level of income. It could be argued that Dusenbury possibly assumed that the constant income level stated in the tax scenarios would carry across in the participants' minds — if income was a significant factor to them in framing this problem.

¹⁴ Kahneman, D. and Tversky, A., (1979), "Prospect Theory: An Analysis of Decision under Risk". Examples of these are listed in Chapter Two of this study, as Problems 1 - 8.

¹⁵ Baldry, J. C., (1986), "Tax Evasion is not a Gamble".

¹⁶ Dusenbury, R., (1994), op. cit. n. 3, p. 7.

H_{3A}: Participants will select riskier options in the gamble case than in the *low pay* tax case.

4.2.2 Dusenbury's Subjects

Dusenbury used 65 non-randomly selected participants in 5 groupings. One of these groupings consisted of eight members of a support group of an athletics team;¹⁷ another were ten supporters of a school band; the third were fourteen parents recruited from a group attending a university freshman orientation programme; the fourth and largest were twenty-two summer school attendees; and the fifth grouping were eleven members of a church congregation. What all these people had in common was that they were at least 30 years of age and had had more than two years' experience of full employment.

4.2.3 The Experimental Procedure

Each participant was given a training script with one scenario on it for familiarization, and a package containing a set of scenarios, pseudo money (stock cards) to the value of \$3,000 in \$100 notes, and an envelope in which to post payments made during the session. The first few minutes were given over to an oral introduction delivered by Dusenbury himself, or some other person in charge of the session; and in this time, participants could ask any questions they had about what was required of them.

When the experimental session began in earnest, the participants opened envelopes containing each of the five scenarios printed on a separate sheet. The order of progression was prescribed for every participant. To prevent the receipt of a refund by one participant having an influence on the choices of others in the session, everyone was required to do the *refund* tax case first. The remaining four scenarios were ordered in four different ways so that each choice appeared in each position once. Apart for the handing out of refunds requested, all outcomes were put on hold till all decisions on all scenarios had been recorded (unseen by others) on the scenario

¹⁷ Ibid, p. 9. Dusenbury's term for this was "an athletic booster group".

sheets and returned to each participant's envelope. According to Dusenbury, this overcame the problem of possible decision contamination. He noted¹⁸:

Because the resolution of each outcome was not decided until all choices were made, choices were not affected by outcomes.

The purpose of this secretiveness was to achieve the condition of *privacy*, which is one of four preconditions laid out by Smith (1987) for a well-conducted economic laboratory experiment¹⁹:

This precept is used to provide control over interpersonal utilities (payoff externalities). Real people may experience negative or positive utilities from the rewards of others, and to the extent that this occurs we lose control over induced demand, supply and preference functions.

Dusenbury made sure that privacy was maintained by providing a minimum of one supervisor for every three participants. This also had the effect of streamlining the experimental procedure since any difficulties experienced by participants could be seen to very quickly.

Once a participant had worked through all five scenarios, he or she was required to spin a random outcome generating device, known as a spinner, to determine an outcome for every risky prospect chosen. The spinner selected outcomes in accordance with their probabilities (p). If the spinner's needle recorded an unfavourable result (to the participant), he/she was required to pay the penalty stated in the chosen prospect. Hence all penalty payments were paid by participants at the end of the session.

4.2.4 Monetary Incentive

At the start of their laboratory session, participants were given a fund of \$3000 in pseudo \$100 bills with which to make payments in accordance with the decisions they made in the five scenarios. They were told during the introductory briefing they would receive a real cash payoff of \$US0.50 for every \$100 of pseudo money they still controlled, after the outcomes to

¹⁸ Dusenbury, R., (1994), op. cit. n. 3, p.10.

¹⁹ Smith, V., (1987), "Experimental Methods in Economics" in *The New Palgrave: A Dictionary of Economics*, Vol. 2., p. 248.

any risky decisions had been determined, at the end of the fifth scenario by a random outcome generator — in this case, the spinning of the spinner.

The purpose of the incentive was twofold. In the first instance, the handling of the experimental paper money (pseudo money) would render the scenarios and the business of paying tax and receiving refunds more tangible to the participants; and in the second, they would be encouraged, by the possibility of real gains to themselves, to take the decision problems seriously rather than dismissing them as some kind of game or other abstraction unrelated to actual financial decision-making. Dusenbury's aim was to make the tax-filing activity in the experiment as real as possible to the subjects. If they paid close attention to the details provided in the scenarios and made careful and considered decisions on the basis of that information, then the experiment could be said to have *external validity*.²⁰

Dusenbury's achievement of external validity was testable in terms of the results associated with testing his second and third hypotheses. He noted that if his design and procedures induced responses lacking in external validity, then participants would exhibit identical preferences in lotteries which were structurally identical.²¹ But, as recorded above, Dusenbury did not find this to be the case.

4.2.5 Questionnaire

Dusenbury also required his participants to answer a debriefing questionnaire. In this, information was gathered on a number of variables relevant to the tax scenario decisions. These variables included participants' tax filing experience in years, their self-reported non-

²⁰ The use of the \$0.50 per \$100 incentive met Smith's (1987) requirements for salience (a clear connection between act and outcome, and between outcome and reward) and dominance (the reward to the participant outweighs the participant's costs in terms of time and effort complying with the experimental requirements). See Smith, *op. cit.* n. 19, p. 248.

²¹ *Ibid.*, p. 7, Footnote 7: "The context contrasts also provide evidence regarding the influence of the experimental procedures on the participants' choices. The experimental design and procedures include highly structured lotteries, experimental funds, and resolution of uncertainty via a spinner. If the design and procedures induce responses lacking external validity (i.e., if subjects attend only to the incentive scheme and ignore the decision context), then participants should have the same preferences in structurally identical lotteries. The gamble case and low pay case are structurally identical as are the health insurance case and high pay case."

compliance, participants' attitudes towards taxpaying, the frequency of participants' tax-due and payment-due tax filings, and their perceptions of actual tax audit risks.²²

4.2.6 Dusenbury's Results

Although both repeated-measure ANOVA and MANOVA tests enabled Dusenbury to reject all three hypotheses, he reported only the ANOVA statistics:

1. Taxpayers, operating within a decision framework in which the level of prepaid provisional tax was low enough for all options to be seen as loss prospects, were significantly more risk-willing than the same taxpayers operating within a decision framework in which there was a perception of choices among gain prospects ($p = 0.01$).
2. The mean riskiness of prospects chosen in the high payout tax scenario was significantly higher than in the health insurance case ($p = 0.01$). This possibly indicated the existence behind the surface of the decision frame of an ideology stressing a greater degree of concern for family physical welfare than on one's obligation to the state.²³
3. The mean riskiness of prospects chosen in the gamble case was significantly higher than in the low pay tax case ($p = 0.0001$).

When Dusenbury conducted an analysis of his laboratory experiment results in conjunction with variables about which information was solicited in his debriefing questionnaire, he found that only one of these variables had a significant relationship with options selected in the tax scenarios. This variable was tax filing experience in years. Taxpayers with more filing experience were found to prefer riskier filing positions than less experienced taxpayers in both the high terminal tax bill and refund cases.

²² Dusenbury (1994), *op. cit.* n. 3, p. 13.

²³ Dusenbury does not state this, but it is implicit in his hypothesis.

4.2.7 Links with Earlier Research and Dusenbury's Unique Contribution

In what respects was Dusenbury (1994) signally different from earlier *Prospect Theory* paradigm studies in compliance incorporating a withheld tax variable? Like White, Harrison and Harrell (1993), Dusenbury subjected his participants to a set of comparable scenarios. However, where White et al presented the choice to be made in each case in the form of ten-point Likert scale in which the extreme points represented a definite preference for one of the two prospects offered, Dusenbury offered participants a choice of five different courses of action, each containing a discrete level of risk and payoff, as summarised in Subsection 4.2.1.

Furthermore, in a similar fashion to Robben, Webley, Elffers and Hessing (1990) in part, Schisler (1992 and 1995) also in part, and White et al (1993), Dusenbury used a sample of taxpayers who were not primarily students; but unlike the earlier researchers, he also provided the cash incentive scheme described in Subsection 4.2.4 above, whereby participants earned real returns on the results of their efforts.

At this point we move from analysis of decision frame research in general and Dusenbury (1994) in particular to consideration of the 1995 New Zealand replication of Dusenbury's experiment.

4.3 WHY A REPLICATION?

4.3.1 Definition

There are two types of replication, *pure* and *conceptual*. A *pure* replication is an exact reproduction of the original study using the original raw data and method. This is useful as a cross-checking device, as a replicant, approaching the issues with a set of expectations not necessarily congruent with those of the original researcher, will provide validation if the results turn out to be identical; and may uncover inadvertent errors if the results are not identical. The replication tabled in the current undertaking, however, is *conceptual*. A *conceptual* replication

is essentially a variation of an original study using a different sample or different constructs of the empirical variables to determine the robustness of the findings in the original study.²⁴

4.3.2 Purpose

Borg and Gall (1983)²⁵ argued that studies conducted in what are known as the social sciences need to be replicated before their findings should be accepted. Along with the cross-checking function, Borg and Gall cited the need to delimit the population for which the findings hold true, and the need to check for trends or changes over time. They also argued that the use of a different method provides an excellent way of checking the validity of a study. The current replication performs several of these functions. In the first instance, the subjects were groups of taxpayers living in Christchurch, New Zealand, which afforded a cultural comparison with Dusenbury's, North American taxpayers, half of whom were associated in some way with the University of Florida. And in the second instance, the original study's laboratory instrument was adapted so that it would relate to existing New Zealand income tax rates and be easily decipherable to ordinary New Zealanders.

The academic validity cross-cultural comparison was also underscored by Jackson and Milleron (1986)²⁶:

More basic research is needed before we can begin to comprehend the phenomena of tax evasion. ...Little work has been done on comparative tax compliance between countries, although an evaluation of the studies of tax evasion in other cultures could potentially provide insight into the projected effects of changes entertained. ...Another fundamental issue is the congruency of the tax system with the basic ideologies of the public. ...In this regard, the experience of other countries may be relevant in evaluating changes designed to improve compliance in the United States.

²⁴ Fischer, C. M. and Russell, J. D., (1991) "Research Note: Replication of Empirical Tax Research", p. 74.

²⁵ Borg, W. R. and Gall, M. D., (1983), *Educational Research: An Introduction*. This work was cited in Fischer and Russell, (1991), *ibid*, p.75.

²⁶ Jackson, B. and Milleron, V., (1986), "Tax Compliance Research: Findings, Problems and Prospects", p. 156. Jackson and Milleron's study has already been referred to, in brief, in Chapter 3.

Hasseldine, Kaplan and Fuller (1994) underlined the importance of conceptual replications by prefacing their study in the following terms²⁷:

The purpose of this note is to present descriptive evidence of the characteristics of tax evaders and to identify factors associated with tax evasion behaviour in New Zealand. This study extends prior work by focusing on New Zealand individuals, responding to calls by Lewis [1982, p. 139] and Jackson and Milleron [1986, p.156] for further international comparative studies.

4.4 THE REPLICATION HYPOTHESES

4.4.1 Hypotheses

It is the intention of this study to replicate the tests of the three hypotheses laid out by Dusenbury (1994). For convenience, these are repeated here.

H_{1A}: Participants will choose riskier options in the *high pay* case than in the *refund* case.

H_{2A}: Participants will choose less risky options in the health insurance case than in the *high pay* tax case.

H_{3A}: Participants will select riskier options in the gamble case than in the *low pay* tax case.

4.4.2 Research Design Modifications

Investigation of these hypotheses in terms of New Zealand taxpayers is a conceptual replication. The reason for this is that several modifications were made to the materials presented to subjects in the laboratory experiment. In preparing the scenarios for consumption by New Zealand participants, the tax rate was changed from Dusenbury's 20 percent — which he said was low — to the actual rates at which tax is levied by the New Zealand Government at

²⁷ Hasseldine, D. J., Kaplan, S. E. and Fuller, L. R. (1994), "Characteristics of New Zealand Tax Evaders: A Note", p. 80.

the present time via its Inland Revenue Department, which are 24 cents in the dollar to a maximum income of \$NZ30,875, and 33 cents in every dollar of income above that ceiling. All other figures were then reworked in order to make them fit the level of overwithholding or underwithholding implicit in the dollar amount of the terminal tax adjustment stated in the scenario's riskless prospect (option A).

Also, as a result of feedback from fellow graduate students in the pretesting process, the level of withheld tax was given much greater prominence; and the size of the departure from the assumed neutral reference point was explicitly stated in terms relating to the given level of income in a fashion reminiscent of (but not identical to) Schadewald's (1989) third and fourth hypotheses.²⁸ This was done because it seemed a little odd that the variable under examination, level of tax withheld, was presented as an inconspicuous figure requiring an act of computation or conceptual leap on the part of the participant to link it with data provided in the tax-filing options, when all the information necessary to understand the implications of options A to E came in a prepackaged and pre-digested form in the summary table. The following extracts contain Dusenbury's original information from the *refund* case and the equivalent version used in the New Zealand replication:

Dusenbury (1994)

The task is to file your income tax return. The correct amount of your income tax is uncertain because the correct amount of one deduction item is in doubt. You must choose the amount of this item to report. Your taxable income is \$27,780 not counting this item. Your withholding is \$5,400.

In Option A you deduct none of this item, and there is no risk of under-reporting your tax. In Options B, C, D, and E you deduct more and more of this item, and both your refund and also your chance of under-reporting increases...

Dusenbury (1994) Tax Case Y.

²⁸ See Chapter Three, Subsection 3.3.5. With respect to his fourth hypothesis, Schadewald found a significant framing effect when departures from a neutral reference point were stated explicitly as gains or losses. However the New Zealand replication of Dusenbury's (1994) does not make explicit use of the gain/loss labels. More in keeping with the ideas expressed in Schadewald's third hypothesis, the withholding position is expressed in terms of its deviation from tax payable on the stated level of income earned in the year. No mention is made of prior expectations or any surprise factor in the replication. See Schadewald, M. S., (1989), "Reference Point Effects in Taxpayer Decision Making".

Replication (1995)

The task is to file your income tax return. The correct amount of your income tax is uncertain because the correct amount of one item you could claim a rebate on is in doubt.

You must choose the amount of this item to report.

Your taxable income is \$24,166 not counting this item. **You have already made PAYE payments totalling \$5,400** which is **\$400 too much** on your income (ignoring the possible rebate).

In **Option A** you do not claim the rebate, and there is no risk of under-reporting your tax.

In **Options B, C, D, and E** you claim more and more of the rebate item's cost, and both your refund and also your chance of under-reporting increases...

New Zealand Replication (1995) Tax Case Y.²⁹

A related aspect of scenario wording and formatting is discussed below in Subsection 4.5.2, under the heading, 'Summary Syndrome Extension'.

Another change to the scenarios related to the nature of the participants. It could not be assumed that everybody had had experience in filing taxes on income earned from self employment. Therefore, the nature of tax withholdings was expanded to deal with tax withheld by employers on behalf of employees, in accordance with the Pay As You Earn (PAYE) system.³⁰ The point was made in the introductory speech that PAYE and provisional tax were both terms for tax gathered by the taxation authority before the time of filing — and that any differences between them were not salient. Participants were to consider themselves self-employed in some of the scenarios and employees in others (employees in fact, in Tax Case Y). The presence of an extraneous framing effect arising from the switch between PAYE and

²⁹ The use of bold text in this excerpt follows the formatting of the experimental instrument. These particular clauses and phrases were printed this way in the scripts read by the participants.

³⁰ The main purpose of this was to prevent participants from developing a sense of alienation. This was achieved by providing at least one scenario in which the withheld tax type they were familiar with, was used.

provisional tax could be tested for with information obtained in the questionnaire on participants' employment history.

Perhaps a more serious change from Dusenbury (1994) was the jettisoning of Dusenbury's uniform income level. This act was a by-product of the recalculations designed to incorporate New Zealand tax rates and an explicit level of discrepancy between tax withheld and total tax liability. The aim was to provide a plausible tax position which could be checked by a sceptical participant with a calculator. However, since no questions were asked about the relationship between the stated income and the tax position, either the structure was deemed logical, or none of the participants felt the need to do the check.

But, what is the effect on the replication of varying levels of income? The only hypothesis potentially affected is H_{1A} ,³¹ but these changes do not invalidate cross-cultural comparisons of the findings with respect to this hypothesis if the following line of reasoning is valid. Given that there is a higher level of income in the *high pay* case than in the *refund* case, and the *refund* case is always the first case encountered, one might expect a framing effect based on income to bias the participant in the direction of risk aversion. The reason for this is that income has risen and a rise is framed as a gain. The effect of this would be to dampen the risk willingness *Prospect Theory* predicts in a loss frame context. The loss frame in the *high pay* case is predicted by Dusenbury (1994) on the basis of the implied underwithheld position existing in this scenario. Therefore, while the putative income framing effect may have a moderating impact on a detected level of risk willingness in the *high pay* tax case, it will not have confounded it unless no risk willingness is detected at all.

Nevertheless, as in Hite, Jackson and Spicer's (1988) study, the subjects in the current experiment will be sorted by household income level to determine whether this is the source of any extraneous influence, be it potential or actual, on risk preferences. This is made possible

³¹ Replication of Dusenbury's other two hypotheses is not affected on the ground the neither the health care case nor the pure gamble contain income information.

by several questions in the experiment's end-of-session questionnaire on the size and nature of household income.³²

4.5 EXTENSION HYPOTHESES

Fischer and Russell (1991) pointed out that a conceptual replication of a tax study may also involve new elements or an extension which may potentially provide a substantive contribution to the compliance literature.³³ It is the intention of this study not only to expand Dusenbury's investigation in geographical terms by running his experiment in a New Zealand context, but also to extend the scope of the experiment itself, by comparing the risk profiles of taxpayer subjects with ample cash resources with those operating on a much tighter cash budget. To this end the number of subjects recruited to undergo the experiment was increased from Dusenbury's 65 to 132. This extension is explained in further detail in Subsection 4.5.1.

A number of secondary extensions are then considered. In Subsection 4.5.2 an extension relating to a possible propensity on the part of subjects to confine their decision-making largely to information laid out in each scenario's summary table is investigated. Next, in Subsection 4.5.3, the stability of framing effects over near-identical tax situations is evaluated; and the final extension, tabled in Subsection 4.5.4, investigates the assumption (inherent in Dusenbury's experiment) that subjects' value functions are of a consistent shape congruent with that posited by *Prospect Theory*.

³² Hite, P. S., Jackson, B. R. and Spicer, M. W., (1988), "The Effect of Framing Biases on Taxpayer Compliance", p. 13. Hite et al were concerned that their results may have been confounded by subjects adopting reference points relating to their own incomes rather than in the data provided by the researchers' experimental scenarios.

³³ Fischer and Russell, (1991), op. cit. n. 24, p. 76.

4.5.1 Cash Flow Extension

There were two motivations for this extension.³⁴ Dusenbury (1994) noted a possible confounding factor in his experiment resulting from a frame hinged on the availability to participants of liquid cash reserves³⁵:

The experimental fund was always positive, which might have diminished risk taking.

Perhaps Dusenbury was amplifying this, when in the final paragraph of his paper he said³⁶:

An extension of this study would be to determine whether risk preference shifts are greater or smaller under alternative risks, tax rates and amounts at issue.

In the second instance, a *cash on hand* frame was one of the multiple frames detected by Carroll (1992) in his investigation, in Boston, of taxpayer thinking via interviews and diary entries.³⁷ And, as discussed in Chapter Three, Subsection 3.3.8, a liquidity variable was demonstrated by Martinez-Vazquez, Harwood and Larkins (1992) to have an impact on compliance in conjunction with a three percent risk of detection.³⁸

For the purpose of the cash flow extension, each subsample of New Zealand subjects was partitioned into two equal groups. One of these was given a high initial float, while the other received a substantially lower fund of pseudo money. Consequently, half of the 132 people who undertook the experiment (66) had a float of \$3,500 or more;³⁹ and the other 66 were given only \$2,100. This lower amount was set at a level calculated to cause a participant making

³⁴ Note that the terms *flow* and *float* are both used with reference to this extension. These terms are so closely allied in meaning that they may be used interchangeably.

³⁵ Dusenbury, R., (1994), op. cit. n. 3, p. 13.

³⁶ Ibid, p.15.

³⁷ Carroll, J. S., (1992), "How Taxpayers Think about Their Taxes: Frames and Values", p. 51. A few other aspects of Carroll's study have already been described in Chapter Three, Subsection 3.3.1.

³⁸ Martinez-Vazquez, J., Harwood, G. B. and Larkins, E. R., (1992), "Withholding Position and Income Tax Compliance: Some Experimental Evidence". This study was cited in Chapter Three, Subsection 3.3.8.

³⁹ The 30 high float participants in the initial subsample received \$4,200; but this level of initial funding gave rise to high stocks of residual funds in their hands at the end of the experiment, so the level was reduced to \$3,500.

consistently risk averse responses to barely break even. The decision-making of the two groups could then be compared to gauge the extent to which a decision frame involving lack of liquidity impacts on a decision maker's risk profile. This gave rise to the following hypothesis:

H_{4A}: Participants with a potentially inadequate cash flow will choose riskier options in the *high pay* case than chosen by participants with high cash flow.

4.5.2 Summary Syndrome Extension

This extension was formulated as a result of perceptions arising from feedback from the pretesting procedure,⁴⁰ and from an awareness that earlier studies in tax compliance had experienced problems with the failure of participants to engage themselves in the simulated taxpaying roles required of them.⁴¹ One of Dusenbury's concerns was that subjects might make their financial decisions purely on the basis of what they read in each scenario's summary table. Dusenbury stated that one of the reasons for setting up his second hypothesis (contrasting the *high pay* tax case with a *medical* context), and his third hypothesis (contrasting the *low pay* case with a straight *gamble*), was that he wanted to provide evidence of *external validity*,⁴² a concept discussed above in subsection 4.2.4. Indeed, the inclusion of a monetary incentive in the experiment was intended to strengthen the participants' focus on *all* of the financial information so as to make the most appropriate financial decision.

Whether or not there is a monetary incentive element in the experiment, there is likely to be a tendency on the part of experimental subjects to minimise the processes by which they arrive at a decision — especially when the information which has been provided requires careful

⁴⁰ Once it became apparent that there was a phenomenon to watch for, it was quietly, but keenly observed. Over the period in which the laboratory sessions were run, a conviction of the extension's validity, as a focus of research, grew.

⁴¹ In the conclusions to their study, Hite, Jackson and Spicer (1988) recorded, as one of the study's limitations, the inability of participants "to put themselves in the role". See Hite, P. S., Jackson, B. R. and Spicer, M. W., (1988), "The Effects of Decision Framing on Taxpayer Compliance". This was cited in Chapter Three, Subsection 3.3.3. Also see Martínez-Vázquez, Harwood and Larkins (1992), *op. cit.* n. 38, cited in Chapter Three, Subsection 3.3.8. Martínez-Vázquez et al found that their MBA student subjects tended to ignore contexts and uniformly chose options associated with the highest available expected value. Furthermore, Schadewald (1989) found that subjects tended not to think in terms of gains or losses measured from a neutral reference point determined by prepayments unless the gains or losses were explicitly stated. See Schadewald (1989), *op. cit.* n. 28, cited in Chapter Three Subsection 3.3.5.

⁴² *Ibid.*, p. 7, Footnote 7. See Footnote 20 of the current chapter.

consideration, and when a number of decision problems, structured in a standardised format, are to be solved sequentially. Simon (1976), in his discussion of substantive rationality,⁴³ argued that humans learn procedural rules which fit a given problem context, which then become heuristics to be applied in identifiably similar problem contexts. What makes a context similar is the recurrence of a given significant pattern.⁴⁴ The format of the five original Dusenbury scenarios (and the reworded New Zealand versions) provided a set of recurring patterns; and the adoption of a heuristic would minimize the effort expended on extracting information from the body of data in each case.

The fact that there is a learning curve, which participants must scale in order to be able to perform the experimental tasks competently, is recognised in the provision of a sample case in the training script. In their own tax-paying lives, the participants presumably have not before encountered a neat one-page puzzle of this sort, including specifications of alternative risks and payoffs; so, in spite of all efforts to produce verisimilitude, the scenario on the piece of paper is something novel for even the most experienced of taxpayers. Consequently, participants absorb what they need to know in order to make decisions in this novel environment by listening to the session organiser's introductory speech and/or by reading the instructions in the sample case, or more generally, in the training script. However, it is human nature to 'cut to the chase'; and — as stated in the law of the apocryphal Mr Parkinson — when all else fails, *then* read (or reread) the instructions. The learning process is facilitated — but *not* completed — in the initial scan through the training script's written material while listening to the verbal explanation. The process is completed when the participant feels he or she understands enough to get on with the tasks set.

In effect, the participant selects what he or she considers to be salient points from the apparent plethora of detail provided; and sets about addressing the initial decision task with a set of tools

⁴³ Simon, H. A., (1976), "From Substantive to Procedural Rationality", pp. 130 - 131: "Behavior is substantively rational when it is appropriate to the achievement of given goals within the limits imposed by given conditions and constraints. Notice that, by this definition, the rationality of behavior depends upon the actor in only a single respect — his or her goals. Given these goals, the rational behavior is determined entirely by the characteristics of the environment in which it takes place."

⁴⁴ *Ibid*, p. 136.

determined by that act of selection. If the selection turns out to be less than adequate in the participant's eyes, then he or she returns to the written explanation and instructions, or asks the session organiser (or one of the other supervisors) a question aimed at remedying that comprehension deficit. This was confirmed to be the case by the elucidation questions asked in every New Zealand replication laboratory session, since in this Antipodean experiment, participants were given to understand they could ask questions at any time.⁴⁵

It is posited in this extension that participants may have been primarily influenced by as few as just two salient information items; and perceived that adequate decision-making could be done on the basis of these two items alone. Once this perception was acted upon, it effectively provided a heuristic for deciphering all of the decision problems. It is further argued that this minimized set of information items could give rise to a non-tax-withholdings-related decision frame which would still result in decisions essentially congruent with those recorded by Dusenbury (1994).

The first of these items was the scenario's *title*. Every scenario had a title; and for most participants, this could be expected to be the first thing they read. It is argued that the very mention of the three-letter word, *tax*, would be enough to provide an initial mindset different from those induced by the words *health care* (in the New Zealand replication, *medical insurance*) on the one hand, and *gamble* on the other.

The second item was the *summary table* at the bottom of each scenario. The very nature of the figures provided in that table gives rise to framing in terms of a loss or a gain. In Tax Case X, a set of payments due is provided, ranging from a riskless maximum of \$800 down to a minimum payment of \$400, with a 40 percent probability of a further \$700 demanded later. This, without any other information, may set up a loss frame in the mind of a participant. By contrast, in Tax Case Y, a set of refunds is provided, ranging from a riskless \$400 to a risky

⁴⁵ That participants had a reasonable right to information and the right to withdraw from the experiment at any time was one of the requirements imposed by the University of Canterbury's Human Ethics Committee.

\$800 which has a 40 percent chance of later being reduced by \$700 to a net \$100.⁴⁶ This alone may set up a *gain frame* in the mind of a participant. Hence the difference between an individual's responses to the *high pay* tax case and the *health care plan* case may well come down to reactions to the words in the scenario titles. Although it was not investigated, the same argument applies in the case of distinctions between the *low pay* tax case and the *choose a gamble* task.

The information required to investigate the information minimization phenomenon, henceforth called the Summary Syndrome, was gathered in two extra questions inserted in every package of scenario scripts at the end of the fifth scenario. These questions were:

My choice of option was influenced by the level of PAYE said to be withheld in **Tax Case Y**:

- Not at all — I ignored it.
- Only in a slight fashion.
- It was significant in my thinking.
- It was quite important.
- Of major importance — it totally shaped my thinking.

My choice of option was influenced by the level of provisional tax said to be withheld in **Tax Case X**:

- Not at all — I ignored it.
- Only in a slight fashion.
- It was significant in my thinking.
- It was quite important.
- Of major importance — it totally shaped my thinking.

For simplicity, the variable obtained by asking the first question was called INFLY (Influence of Y's story-line); and following the same pattern, the second variable became INFLX. Answers to these two questions made several approaches to analysis of the Summary Syndrome possible, depending how the variables were employed; but the hypotheses in both instances were the same:

⁴⁶The tax filing option tables of these two scenarios are reproduced in Tables 4.1 and 4.2 in Subsection 4.2.1 of this chapter, and also in Appendix B, Section B.3.

H_{5A} : Participants who claim to have ignored withheld tax data in the *high pay* and *refund* tax cases will make filing decisions in which a non-withholdings-related differential decision frame effect is discernible.

4.5.2.1 Partitioning By Levels of Declared Influence

It was possible to separate the 132 pairs of responses to Tax Cases X and Y into ten categories — one for every level of INFLX and one for every level of INFLY. When this was done, the partitioned groups of Tax Case X and Y responses could be retested in terms of H_{1A} in order to determine if the partitioning made any difference to the H_{1A} result.

4.5.2.2 Declared Influence Incorporated as a Categorical Variable

The second method for analysing the impact of INFLX and INFLY was to incorporate them as categorical variables in the repeated-measures analysis of variance procedure employed to detect risk level shifts between *high pay* Tax Case X and *refund* Tax Case Y. Accordingly, the observation levels of INFLX and INFLY were collapsed from five to two. The natural line of division within the influence questions fell between the second and third response boxes, which allowed for $INFLX < 3$, and $INFLY < 3$ (not influenced); and $INFLX \geq 3$, and $INFLY \geq 3$ (influenced).

If INFLX and INFLY, recoded in this manner, could be shown to have an *insignificant* relationship with any detected shift between the two tax-related decision problems, then it would be clear that participants could act in accordance with the existence of a decision frame in which the level of stated tax withholdings had no input. Such a result would give grounds for rejecting the null form of H_{5A} .

4.5.3 Risk Profile Stability Extension

This extension explores the assumption implicit in laboratory experiments such as Dusenbury (1994) that participants will be consistent in their decision-making when presented with choices which are alike in both decision context and in the format and scale of the available

Financial Datum	Tax Case X	Tax Case W
Income before doubtful item	\$40,000	\$41,148
Provisional Tax Paid Up	\$9,621	\$10,000
Level of Underwithholding	\$800	\$800

options. *Prospect Theory* predicts that people will display this form of consistency.⁴⁷ The extension was made possible when two extra scenarios were inserted into the experimental instrument on the eve of the first laboratory session.⁴⁸ One of these, *high pay* Tax Case W, provided a sample of 83 experimental responses in which a statistical comparison was made possible between two essentially similar decision problems. The case which the last-minute insertion, Case W, could be compared with was *high pay* Tax Case X. This comparison was facilitated by the fact that the stated level of underwithholding was identical in both cases, and the set of options provided in the summary table the two scenarios was also identical. However, the levels of provisional tax said to have been paid to date, and the levels of stated income in the two scenarios differed by a relatively small amount, as shown in the Table 4.4.

As in the situation arising in the testing of H_{1A} , the influence of any framing effect caused by income will be in the direction of risk aversion on the ground that *Prospect Theory* predicts an increase in income will be perceived as a gain. (In terms of *Expected Utility Theory*, all levels of wealth are associated with risk aversion.) Given that the income and withholdings in the two scenarios are likely to be viewed as similar, and that a decision frame based on losses from a

⁴⁷ *Prospect Theory* predicts that contexts giving rise to perceptions of loss will induce a greater degree of risk willingness; whereas contexts giving rise to perceptions of gain will induce a greater degree of risk aversion. See Chapter Two, Subsection 2.4.2.2, which explains the theory's value function.

⁴⁸ While these two scenarios, *refund* Tax Case V and *high pay* Tax Case W, are not a part of the study as it was initially envisaged and planned, they were included because the opportunity to run them on a recruited sample was too good to miss. They were cobbled together in a short space of time as a result of a cross-fertilization conversation with several academic members of the University of Canterbury's Economics Department involved in *Decision Theory* research. However, the insertion was insufficiently well-prepared for any serious investigation to arise from it with respect to the current study; but it does serve as a starting-point for future compliance laboratory experiment research, which is explored in Chapter Eleven. A cutting and pasting error occurring in the preparation of Tax Case W, undetected until eighty-three participants had been processed, made the current extension possible. As such, the extension was an unexpected windfall. The final 49 subjects received scripts with the cutting and pasting error removed; but as the insertion, as initially conceived, was conceptually flawed, these 49 Tax Case W results were not used in any form in the study.

neutral reference point should produce evidence of risk willingness in both cases, the following negatively framed hypothesis is proposed:

H_{6A}: The risk preference exhibited by participants in the *high pay* case, Tax Case X will be different from the risk preference exhibited by the same participants in the *high pay* case, Tax Case W, (First 83 scripts).

Hence, if the null hypothesis cannot be disproved, then the evidence may be taken as supporting risk preference stability.

A second hypothesis may also be investigated in order to provide a little more depth to what may be understood from the results obtained in the testing of **H_{6A}**. This simply involves investigating a variant of **H_{1A}** with respect to the 83 participants who received the appropriately-formatted Case W scenarios. This extra hypothesis is worded:

H_{7A}: Participants will choose riskier options in *high pay* Tax Case W than in *refund* Tax Case Y.

A significant case-related difference detected with respect to this hypothesis will provide confirmation that responses to Case W were indeed of a strongly similar nature to responses to Case X.

4.5.4 Value Function Extension

This extension investigates one of Dusenbury's (1994) operating assumptions. It is concerned with the smoothness of the value functions expressed by participants in their second, third, fourth and fifth choices of options associated with each decision problem. According to Kahneman and Tversky (1979),⁴⁹ the value function is a smooth S-shape with a point of inflection at the neutral reference point. This shape was retained in Tversky and Kahneman's (1992)⁵⁰ update of the theory as *Cumulative Prospect Theory*. According to *Expected Utility*

⁴⁹ Kahneman, D. and Tversky, A., (1979), op. cit. n. 14.

⁵⁰ Tversky, A. and Kahneman, D., (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty".

Theory the function is a smooth concave curve. The issue at stake here is whether *Prospect Theory* (or for that matter, *Expected Utility Theory*) provides an adequate explanation of the participants' behaviour in a tax-filing setting. Since this study is concerned primarily with a *Prospect Theory* perspective, it is the existence of the S-shaped curve which will be investigated in terms of the smoothness of the choice orderings declared by the participants.

At the end of each of the five scenarios adapted from Dusenbury (1994), the participants were required to record their ordinal preferences before moving on to the next case⁵¹:

My second choice would have been: _____
 My third choice would have been: _____
 My fourth choice would have been: _____
 My fifth choice would have been: _____

If it can be shown that their preferences followed a smooth progression away from the most favoured option, then the evidence can be said to support the contention that the *Prospect Theory* value function is a reasonable approximation of the participants' value functions. If a smooth progression away from the subjects' first choices cannot be found, then the *Prospect Theory* value function does not provide a valid approximation of the subjects' value functions.

Assuming Option C is a subject's first preference, a smooth progression to the least favoured option may be recorded in two manners:

The zigzag configuration: C - B - D - A - E; (or C - D - B - E - A)

Or

The two sweeps configuration: C - D - E - B - A; (or C - B - A - D - E)

By coding the alphabetical displacements out from the letter designated as first preference (circled in the summary table and acted upon financially), the smoothness of the value function — or lack of it — can be ascertained. In the case of the zig-zag configurations, the

⁵¹ See Tax Cases X, Y, Z, and the Medical Insurance and Gamble Cases set out in Appendix B, Section B.3.

displacements map to 1 - 2 - 2 - 3 - 3. In the case of the two sweeps configuration, the displacements become 1 - 2 - 3 - 2 - 3.

Examples of a non-smooth progression, assuming Option C is the first choice are:

$$\begin{aligned} \mathbf{C - A - B - D - E} &= \mathbf{1 - 3 - 2 - 2 - 3} \text{ (A is further from C than is B)} \\ \mathbf{C - A - D - B - E} &= \mathbf{1 - 3 - 2 - 2 - 3} \text{ (A is further from C than are D, B)} \\ \mathbf{C - A - E - B - D} &= \mathbf{1 - 3 - 3 - 2 - 2} \text{ (A, E are further from C than are B, D)} \\ \mathbf{C - E - A - B - D} &= \mathbf{1 - 3 - 3 - 2 - 2} \text{ (E, A are further from C than are B, D)} \\ \mathbf{C - E - A - D - B} &= \mathbf{1 - 3 - 3 - 2 - 2} \text{ (E, A are further from C than are D, B)}. \end{aligned}$$

Where any integer is not an immediate ordinal neighbour of the previously listed integer, the value function is not smooth. On this basis, every preference ordering can be shown to be either smooth or broken; and a variable can be constructed with two levels of observations. A value of 1 would indicate smoothness; and a value of 2 would indicate a value function contained a discontinuity.

Furthermore, the size of the break in the value function may be measured. If the maximum jump from the first preference to the next is measured on a five-point interval scale, a jump from A to E can be reported as a level 4 observation ($E - A = 5 - 1 = 4$), while a jump from A to B would register as a level 1 observation implying smoothness. Observation levels 2, 3 and 4 would indicate the range of breaks of different magnitudes in the value function.

As well as performing a simple count of both of these sets of measurements, it would be of interest to know if the New Zealand taxpayer participants display a consistent pattern of smooth value functions over the different tax contexts. This involves making a comparative analysis of observations for each of the three tax scenarios. The existence of inconsistent patterns would indicate that *Prospect Theory's* value function might not be an accurate model of participants' value functions. The hypothesis is stated in its alternative form as follows:

H_{8A} : Subjects will *not* display a consistent incidence of smooth value functions across all three Dusenbury-related tax-filing decision problems.

This hypothesis takes as its null form the assumption that smooth value functions are the naturally predominant type.

The investigation could then be broadened to measure the consistency observable over all five financial decision contexts adapted from Dusenbury (1994). A second hypothesis covering this possibility is H_{9A} :

H_{9A} : Subjects will *not* display a consistent incidence of smooth value functions across all five Dusenbury-related decision problems.

The next chapter deals with the first aspect of the method applied to the research design described in this chapter. This aspect is the organisation of session procedures.

5. SESSION PROCEDURES

5.1 INTRODUCTION

Attention is now turned to the operation of the experiment. Chapter Four provided an initial understanding of the hypotheses. This and the next two chapters essentially furnish an expansion and deepening of this insight. Because this expansion is comprehensive in its detail, Chapter Five restricts itself to dealing with the running of the experimental sessions and considerations of remuneration, while the experimental instrument (comprising the seven decision problems, their related scenarios, and the nineteen-question questionnaire) are dealt with in Chapter Six. Chapter Seven then completes the three-chapter explanation of the method employed in the study by providing an in-depth explanation of the sample's participants in terms of their characteristics and recruitment.

Chapter Five itself is divided into three sections. The first of these, Section 5.2, covers the running of the sessions, while the second, Section 5.3, takes a look at the obligatory peripheral documentation required in the running of the experiment under the auspices of the University of Canterbury. A peripheral document called the *tally sheet* is also discussed in this section. The final section, Section 5.4, reviews the reasoning behind the shift from Dusenbury's 50 cents per residual \$100 stock card incentive to the 30 cent per residual \$100 pseudo-money banknote utilised in the current experiment.

5.2 SESSION FORMAT

5.2.1 Introduction

The laboratory experiment was conducted on various week-day evenings and Saturday mornings between the middle of July and late September. It was drawn out over this length of period as a result of difficulties in recruiting the requisite number of people.

It turned out that processing what was eventually 132 volunteers required the running of 13 sessions ranging in size from 20 participants down to two. The first nine sessions were held in

a room in the Department of Accountancy, Finance and Information Systems at the University of Canterbury; but of the final four, two were held at locations at some distance from the campus, which, on the ground of geographical convenience, facilitated the recruitment of participants from a school support group and a Lincoln University postgraduate class. The other two sessions were held at a kindergarten because the organizer of participants from that source said that it would be the only way that she would be able to recruit the final 13 people needed to fill the agreed quota of 40.

The format of this section is as follows. Subsection 5.2.2 deals with the nature of the session handouts and the seating arrangements, while Subsection 5.2.3 examines the procedures followed in the running of the sessions. Then Subsection 5.2.4 deals with considerations which arose with respect to the use of the cash float. This last subsection is further subdivided to allow for coverage of the pseudo currency used in the experiment, the partitioning of the sample to accommodate the *cash flow* extension, and modifications to the running of the sessions occasioned by the *cash flow* extension.

5.2.2 Session Kit and Seating

When participants entered the room in which the session was held, they each sat at a desk at which there was a stack of \$100 bills in pseudo-money, a white envelope with the words *Payment Envelope* on it, and a stack of four printed documents. These documents comprised a sheet titled *Purpose and Nature of the Study* (one page), the training script (six pages), the booklet of seven scenarios with the end-of-session questionnaire stapled to it (together totalling 13 pages), and a tally sheet for computing what was owed to a participant's organization as a result of the outcomes of the decisions made. A fifth document, a debriefing sheet, was handed separately to individuals at an appropriate point in the closing stages of the session.

The experiment kits were arranged far enough apart for subjects not to be able to see what others would be doing. The participants were also overseen by a number of roving supervisors; the minimum number of these was one for every three people undertaking the experiment. The supervisors at the various sessions were drawn from the same small pool of pre-trained

voluntary helpers, who, in turn, had participated earlier as subjects in the pretesting of the instrument phase.

5.2.3 Procedure

All sessions were run in an identical fashion. Each commenced with an introductory speech which amounted to an expanded oral version of the contents of *Purpose and Nature of this Study* document and the training script. Participants were then given the opportunity to pass comments and ask questions.¹ They were also told they could ask questions at any time during the experiment; but that these must be addressed to the supervisors. They were also told the supervisors would help them understand what was going on in a scenario; but that they would not make their decisions for them. The participants were then invited to work through the experiment at their own pace, or reread the preliminary documents before doing so, if they so wished.

The seven decision problems and associated tick-in-box questions took most people just over an hour to work through. (Although it was stressed that it was not a time trial, one individual completed the whole thing in less than 15 minutes; but this outlier was balanced by another who took two hours.) Where a choice entailed receiving a refund, the participant was required to raise a hand so that a supervisor could supply the appropriate amount of pseudo-money. Where a choice involved making a payment, the participant was required to file the stated amount in the *payment envelope*.²

After completing the scenarios in the strict order in which they had been presented, each participant moved on to the closing questionnaire which contained 19 questions seeking information on demographic details and attitudes to compliance. The questionnaire is discussed in Chapter Six.

¹ In some instances participants felt free to interject during the introductory speech.

² Since this activity had all the attraction of a school exam for most people, cups of tea and coffee, and in a couple of sessions even biscuits, were provided to ease the cogitation process.

Either during or after the questionnaire phase, each participant's decisions were checked by a supervisor who ascertained that the correct amounts had been posted into the *payment envelope*. The supervisor then invited the participant to determine the outcome of his or her risky decisions by making a draw out of a bag of 100 numbered chips for every risky decision chosen in the course of the experiment. If the number on the chip drawn was less than, or equal to, the risk level associated with the selected prospect, the participant was required to pay up the extra money stated. The chip was returned to the bag after each draw so that every outcome was determined independently from the full complement of 100 equally pickable chips. Both the numerical probabilities of the choices and the numbers of all chips drawn were recorded on the specially formatted tally sheet (described in more detail in Subsection 5.3.5).

At some point after a participant had completed the seven decision problems, and usually after the determination of outcomes for any risky choices, he or she was also handed the debriefing document required by the University of Canterbury's Human Ethics Committee. A more detailed examination is made of this document, together with the *Statement of Purpose* document and the tally sheet, in Section 5.3, below.

5.2.4 Cash Float Considerations

5.2.4.1 The Pseudo-Money

As stated in Chapter Four, Subsection 4.2.3, Dusenbury (1994) used stock cards with \$100 printed on them as the currency in his experiment. Each of his participants received a stock of 30 of these cards, totaling \$3,000 (pseudo-dollars). The \$100 bills used in the New Zealand experiment were quite money-like in that they were created from a \$NZ50 note which was scanned and modified by computer.

5.2.4.2 Partitioning and Cash Float Calculation

Because it was hypothesized that the level of available cash might create a framing effect based on liquidity (See Chapter Four, Subsection 4.5.1), the sample of participants was partitioned into subjects who received a generous cash float, and those who received a float which would be barely adequate. Symmetry of design was maintained by partitioning each of the four

subsamples (as defined by recruitment origin) as nearly as possible into two equal groups. Since there were seven decision problems involving pseudo-money transactions, the high float participants each received \$3,500 (and in the case of the first subsample, \$4,200), and the low float subjects were given only \$2,100. There were 66 participants in each group.

The initial decision to provide \$4,200 (pseudo-dollars) to each participant in the high cash float subsample was based on a quick rule of thumb. If Dusenbury (1994) gave his subjects a \$3,000 float for five decision problems, then an experiment with seven problems would involve handing participants a float of \$4,200 (calculated as a multiple of \$600 per session). This, however, ignored the fact that one of Dusenbury's scenarios entailed a refund, whereas two of the New Zealand study's scenarios contained tax refund situations. Once the first session for the first subsample of participants by recruitment origin was underway, it became clear that this float was unnecessarily generous; so the level of high cash float funding was arbitrarily cut back to \$3,500 for participants recruited from all other sources.³

The low cash float participants received only \$2,100 (pseudo), which was barely sufficient pseudo dollars for a participant taking no risks to break even. It was difficult in advance to determine how miserly to make this low float; so it was set at a level which would return the participant \$100 (pseudo) if he or she chose the riskless option in every decision problem.⁴ There are several interesting issues connected with the setting of such a cash-starvation-inducing cash float, which will be explored in Chapter Eleven.

5.2.4.3 Modifications to Session Procedure

Several issues needed to be taken into consideration with the adoption of the *cash flow* extension. Initially the idea of incorporating a borrowing facility was contemplated, with loan

³ In terms of later statistical analysis of the results, this decision created a greater degree of complexity than was initially envisaged; but the analytical complications were surmountable; and the economising did save a relatively impoverished student from a greater level of payout.

⁴ This equates with the participant's earning only \$NZ0.30 (above the set minimum donation) for his or her organization. The issue of set donations is discussed in Subsection 5.4.3 and in Chapter Seven, Subsection 7.2.2. Note that the participants were only peripherally aware of the set minimum donation; and although it was mentioned in the oral introduction to sessions, it may easily have escaped the notice of some of them.

repayments (including interest) to be made at the end of the session. This was discarded as an unnecessary complication which had been ignored by the participants in the pretesting.

A second issue was whether the two streams of participants should be segregated into separate laboratory sessions. Time and personnel management constraints ordained that sessions tended to contain mixtures of high float and low float participants.

A necessary corollary to this management constraint was an added candour constraint imposed by the University of Canterbury's Ethics Committee. Permission for a laboratory experiment involving human subjects would not be granted if that experiment involved practices which left the participants feeling tricked or betrayed in some fashion. If people taking part in the experiment became aware (without having had the issue explained to them in advance) that they had been given different levels of cash float with which to complete their tasks, such a sense of having been abused could well have arisen. Hence it was necessary that the details of a high and a low float were disclosed in the introductory speech and in the preliminary documentation. This candour had an impact on the results of the experiment, the manner of which is explained in Chapter Ten, Section 10.2.

5.3 PERIPHERAL DOCUMENTATION

5.3.1 Introduction

Because the replication of Dusenbury (1994) in a New Zealand context involved the running of a social laboratory experiment in which subjects disclosed information which they might not divulge if they were aware, in advance, of its nature, it was necessary to obtain a formal clearance from the University of Canterbury Human Ethics Committee (hereafter called the Ethics Committee) before it could proceed. The Ethics Committee gave this clearance on condition that certain disclosures (including the one mentioned in Subsection 5.2.4.3 above) be made in the introductory speech and in the format provided in the following documents. While the *tally sheet* was not a requirement of the Ethics Committee, it is included in this section on the ground that it too was a peripheral document which facilitated the efficient running of sessions with a view to satisfying the interest of outsiders. In this instance the outsiders were

the organisations providing volunteers in return for cash remuneration, as explained in detail in Chapter Seven. Determination of the rate of remuneration is discussed later in Section 5.4.

5.3.2 Statement of Purpose

The first document participants looked at when they entered the room in which their session took place was a single sheet titled *Purpose and Nature of the Study*.⁵ The Ethics Committee required that this document was presented and explained at the beginning of each session; and that a matching, fuller written explanation was provided at the end. The Committee emphasised that the purpose and the nature of the study had to be disclosed to recruits to a degree that would prevent them from feeling deceived and abused once they committed themselves to participation. This entailed making a written guarantee that personal anonymity would be respected at all times, and that the participants would have access to the findings in due course, which is set out below:

The session will last about 60 minutes or up to 90 minutes at the very most. All your answers will be anonymous.

There is a questionnaire to fill out at the end of the session which will be used in the statistical analysis of the experimental results. Neither your personal results nor what you disclose in your answers to the questionnaire will become available to anybody other than the researcher (who only knows you as a number). A summary of results will be available at the end of the project (November) for anyone who wishes to obtain one; but neither individuals nor organisations will be identifiable in the end product. (There are a number of separate organisations contributing participants.)

Also it was necessary to admit at the outset that the primary focus was on decision-making with respect to income tax:

This work is for a Master of Commerce thesis project involving decision making under uncertainty. While much of the experiment relates to tax payments, there is no connection between the researcher and the IRD other than that he is an ordinary taxpayer with a typical low student income. The IRD has no access to information from this experiment that could identify any individual in any way.

⁵ This handout sheet, "Purpose and Nature of the Study", is to be found in Appendix B, Section B.1.

Participants were also informed they were free to withdraw their consent and that in doing so they could walk away knowing that nothing they had recorded would be kept or used.

5.3.3 Ethics Committee Rubric

The act of formally consenting to be a participant was deemed to have occurred when participants opted to take part in the experiment and completed the ensuing questionnaire. The consent process required that participants have their attention drawn to the opening lines of the experimental *training script* immediately after perusal of the introduction sheet, *The Purpose and Nature of this Study*. These opening lines took the form of the following rubric, which was a standard declaration provided by the Ethics Committee for instances in which personal anonymity ordained that no signature should be given:

You are invited to participate in the research project with the working title, 'Personal Financial Judgements' by completing the following experiment and questionnaire. The aim of the project is to determine how people handle tax and medical decisions in situations where the law is not easily definable. Both parts of this undertaking are anonymous, and you will not be identified as an informant without your consent. You may at any time withdraw your participation, including withdrawal of any information you have provided. By completing the experiment and questionnaire, however, it will be understood that you have consented to participate in the project, and that you consent to publication of the results of the project with the understanding that anonymity will be preserved.

The *training script* itself is explained in Chapter Six, in Subsection 6.5.1.

5.3.4 Debriefing Document

The *debriefing* document was, in effect, a repetition of the *Purpose and Nature of this Study* document presented at the commencement of proceedings. The salient difference was that it disclosed a little more detail; but the provision of this detail had to be tempered, however, with

the possibility of behavioural contamination of participants attending later sessions, recruited from the same source. The compromise level of detail was as follows⁶:

As you will have gathered from the experience you have just had, the main issue in this experiment was how people evaluate the choices they have to make in various financial contexts. You have just worked through five tax scenarios, a medical insurance scenario, and a gamble. The research question was, do people treat these quite different financial undertakings in an identical fashion, or do they approach each in a measurably different fashion? In particular the scenarios were set up in such a way that the predictions of *Expected Utility Theory* (from the discipline of Economics) and *Prospect Theory* (from the discipline of Psychology) could be tested in a tax-paying context.

Since this debriefing sheet was intended to be a document which participants could take home, it was not inserted between the experimental instrument and the questionnaire (which were stapled together), even though the logical time for reading it was after completion of the scenarios and before commencement of the questionnaire. Most people read it while waiting for others to finish and have their results computed and checked; and only about half of the participants bothered to take it away.

5.3.5 Tally Sheet

The record sheet was a means by which the summary of the monetary changes associated with decisions taken in the course of the experiment could be standardised for all participants. This was to facilitate computing of the correct level of residual pseudo-money each participant held, and conversion of this figure into New Zealand currency. The New Zealand currency figures

⁶ The full document, titled *What this was all about (in Brief)*, [sic] is printed in Appendix B, Section B.6. There was little concern that disclosure of this information would contaminate the behaviour of volunteers participating in later sessions. In the first instance, much of the disclosed material was presented at the start of each session; and in the second instance responses to the eighteenth and nineteenth questions in the questionnaire provided for an analysis of the level of possible contamination. Question 18 solicited the participants' prior level of awareness of *Expected Utility Theory* and Question 19 did the same for *Prospect Theory*.

TABLE 5.1				
A Sample Tally Sheet				
Case	Option chosen	Risk level	Chip drawn	(H)
Tax Case Y	C	25	24	500
Medical Insurance	C	25	73	
Tax Case X	A			
Gamble	C	25	30	
Tax Case Z	B	25	14	400
Tax Case V	B	25	62	
Tax Case W	A			
AMOUNT IN FUND AT END:				1400
Amount owed to Organization XXX⁸				\$NZ 4.20

recorded at the foot of each record sheet could then be tallied to determine the total remuneration payable to each of the organisations providing volunteers.⁷

The record sheet was a single sheet of paper with a table on it listing the various scenarios and containing a number of blank cells in which a summary of a participant's decisions and the risk levels associated with them could be recorded. To cater for choices of risky options, there was a column for recording the numerical outcome determined by a random dip into the bag of 100 numbered chips. Where an unfavourable outcome was recorded, there was provision for recording the monetary penalty. In Table 5.1, which is a reproduction of the record sheet as filled out by one of the participants, a \$500 penalty in pseudo-dollars was paid on Tax Case Y, and a further (pseudo) \$400 penalty on Tax Case Z.

⁷ While the remuneration payout calculation could have been done with a pocket calculator and a pencil, all data points recorded on the record sheets were entered into a spreadsheet so that the possibility of miscalculation was minimised. This was feasible as the payments to the organisations did not fall due until late September.

⁸ This panel contained the name of the actual organization to which the participant's \$NZ 4.20 was forwarded.

TABLE 5.2 (Tear-off Slip)	
AMOUNT IN FUND AT END:	1400
Amount owed to Organization XXX	\$NZ 4.20

Note the (H) in the top right cell. This records that the person started the experiment with a high cash float.⁹ From this information, a supervisor could calculate that the person started with \$3,500 in pseudo-dollars, made (pseudo-dollar) payments of \$2,300 and received refunds to the value of \$1,100, (not shown, but easily calculated)¹⁰ and was liable for the above-mentioned \$900 in penalty payments (shown in the far right column); so therefore should have (pseudo) \$1,400 after settlement. The pseudo-dollar residue was converted at 30 cents per pseudo \$100 bill, hence 14 bills left converted to \$NZ4.20.

Beneath the table there was a tear-off slip (see Table 5.2) which duplicated the information contained in the final two rows.

This tear-off slip was given to participants for their personal use; and the main table was kept for data entry into a computer spreadsheet, and for auditing purposes.

5.4 THE MONETARY INCENTIVE

5.4.1 Introduction

The current study incorporates a monetary incentive because a monetary incentive was used by Dusenbury (1994). However, the current study provided a payout rate of \$NZ0.30 per \$100 in

⁹ This actually appeared in the top right corner of the sheet; but since the table has been reproduced rather than the table *in situ* on the sheet, it made sense to show it here for the purpose of demonstration.

¹⁰ These calculations required a knowledge of the payments associated with the particular options, which were available on the spot to supervisors and to the researcher at all times. This background data is available to the reader in Appendix B, Section B.5, which contains the seven scenarios.

pseudo-money retained by participants at the end of a session, while Dusenbury paid out at the more generous rate of \$US0.50 per \$100. Why the difference? The reasoning behind the decision to reduce the monetary incentive rate of payout is reviewed in Subsection 5.4.2; and the details of the scheme adopted in the New Zealand study are disclosed in Subsection 5.4.3.

5.4.2 Monetary Incentives in Prior Studies

Monetary incentives have been considered, evaluated, and either incorporated into or excluded from the methods chosen by researchers for nearly as long as laboratory experiments have been conducted in the area of human financial decision-making under risk. Grether and Plott (1979) compared impacts on preference reversal obtained in the presence and in the absence of a small monetary stake on the basis that¹¹:

Theories about decision-making costs do suggest that unmotivated choice behavior may be very different from highly motivated choice behavior...

They noted that such differences, as late as 1979, had remained essentially unexplored.¹² Their attempt to eliminate preference reversals by providing an incentive for taking the decision-making process seriously was emulated by Pommerehne, Schneider and Zweifel (1982),¹³ who provided larger incentives. Neither group was able to eliminate the preference reversal phenomenon; and as a result, the phenomenon came to be considered more robust than had been previously thought. However, these early studies did not shed much light on the validity of the incentive variable itself. This remained as an interesting side issue: did these early experiments show monetary incentives were valid, but that transitivity of preferences was nevertheless violated? Or were monetary incentives irrelevant? And, if this were so, would monetary incentives become relevant if only they were made large enough? Thaler (1986) observed that what he called 'systematic mistakes' with respect to preference reversal

¹¹ Grether, D. M. and Plott, C. R., (1979) "Economic Theory of Choice and the Preference Reversal Phenomenon", p. 624. This paper was also cited in the current study in Chapter Two, Subsection 2.6.3.

¹² Idem.

¹³ Pommerehne, W. W., Schneider, F. and Zweifel, P., (1982), "Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination". This text was also cited in the current study in Chapter Two, Subsection 2.6.3.

phenomena would not necessarily disappear if a sufficiently large enough monetary incentive was provided, since the validity of monetary incentives was “an assertion unsupported by any data.”¹⁴

The validity of one form of high incentive payouts was seriously questioned in the same year by Holt (1986).¹⁵ Holt was concerned with an analysis of the random-lottery incentive system in which payouts were awarded for only one of a number of decision task outcomes in an experiment; and where the one cash-rewarded task was selected afterwards by some means of random selection. He showed that such a system introduced a bias. Starmer and Sugden (1991),¹⁶ however, supported the random-lottery incentive system; and found that Holt’s bias did not occur if the data was interpreted in accordance with *Prospect Theory*.

Tversky and Kahneman (1992)¹⁷ quite firmly debunked the usefulness of monetary incentives of any size. They commented, with reference to their 1992 paper in particular and more generally to the entire sequence of their earlier joint papers, that they were aware of little discernible difference between experimental results gained from subjects paid on an incentive basis and those of subjects paid a flat rate for participation. They agreed with Smith and Walker (1992)¹⁸ that monetary incentives might be a means of eliminating careless mistakes; but that incentives¹⁹:

...are neither necessary nor sufficient to ensure subjects’ cooperativeness, thoughtfulness, or truthfulness.

¹⁴ Thaler, R. H., “The Psychology and Economics Conference Handbook: Comments on Simon, on Einhorn and Horgarth and on Tversky and Kahneman”, p. S280.

¹⁵ Holt, C. A., (1986), “Preference Reversals and the Independence Axiom”.

¹⁶ Starmer, C. and Sugden, R., (1991), “Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation”, p. 977 (particularly).

¹⁷ Tversky, A. and Kahneman, D., (1992), “Advances in Prospect Theory: Cumulative Representation of Uncertainty”, p. 315 - 316.

¹⁸ Smith, V. L., and Walker, J. M., (1992), “Monetary Rewards and Decision Cost in Experimental Economics”. Unpublished manuscript, Economic Science Laboratory, University of Arizona. This paper was cited by Tversky and Kahneman (1992), op. cit. n. 17, p. 316.

¹⁹ Tversky, A. and Kahneman, D., (1992), op. cit. n. 17, p. 316.

However a set of comparative studies conducted in the People's Republic of China and in North America did provide evidence of a tangible monetary incentive effect. Empirical evidence of the differential impact of huge and minor monetary incentives on experimental subjects was provided by Kachelmeier and Shehata (1992).²⁰ They found that the offering of high monetary incentives to students at Beijing University did make a significant difference to the students' behaviour as laboratory experiment participants when the incentive payoff was roughly equal to triple a Beijing student's total monthly income of 60 yuan (¥60). When risky payoffs were set at ¥10 for a range of different probabilities, the researchers found, with increases in the probability of winning, that the level of risk aversion shown by the students was correspondingly greater than the level exhibited by the same group when the payoff level was only ¥1.²¹

The cross-cultural contrast was provided by Kachelmeier and Shehata also running versions of the experiment on a number of North American student samples. In one Canadian sample, the Chinese experiment, with its contrasting massively high and peppercorn payoffs, was replicated on a strictly hypothetical basis only; and subjects were paid a flat fee of \$CAN15. On another campus, students were paid in full for \$CAN1 trials (a peppercorn payoff) and at the rate of \$CAN1 when the hypothetical payouts were \$CAN100. In both of these samples, the students exhibited systematically increasing risk aversion as the probability of winning increased towards certainty; but there was no significant difference between the low-value set and high-value set results. In other words, the sensitivity shown by the Beijing students to a huge

²⁰ Kachelmeier, S. J. and Shehata, M., (1992), "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China", p.1123. The students participated in a series of two-step lottery games in which each student had a chance of winning 10 yuan (¥10 = one sixth of their normal monthly income) at a stated probability in the second step; but was required in the first step to set a minimum selling price on his/her right to participate in the second step. This minimum selling price set by the student was effectively that student's certainty equivalent to the risky ¥10 winnable in the second step. Kachelmeier and Shehata used the ratio of the certainty equivalent divided by the expected value of the lottery in the second step as a measure of risk aversion. The two-step lottery worked as follows. If upon a first draw from a set of 100 cards representing monetary outcomes (incremented at one tenth of a yuan from ¥0.10 to ¥10) the card drawn had a value equal or higher than the participant's selling price, the student was paid off at the selling price. If, on the other hand, the card had a lower value, the student participated in a draw for the ¥10, determined by a draw from a second set of 100 cards determining probability outcomes. The student then collected either ¥10 or nothing.

²¹ The ¥1 lotteries were run before the ¥10 ones. It is also important to note that the risk aversion under study here was relative risk aversion — that is, a tendency towards risk aversion. In all trials undertaken by all ten participant groups, the finding was that for low-probability prospective gains, participants were strongly risk seeking, which was consistent with the predictions of Kahneman and Tversky's (1979) possibility effect. What Kachelmeier and Shehata observed was almost an exponential decrease in this propensity for risk-seeking. This is evident in Figures 1 and 2, p. 1125 and in Figures 3 and 4, p. 1134.

increase in value of real payoffs was absent when the payoffs were scaled to low levels or were rendered hypothetical. The verdict, then, was that large incentives may heighten a risk preference; but they apparently do not reverse it; and the choice between small monetary incentives and the use of flat fees does not make a significant difference to the behaviour of participants.

5.4.3 Setting the Monetary Incentive in the Current Study

The current study made use of both a flat fee *and* a low remuneration rate monetary incentive. Although the scheme is described in brief in the next two paragraphs, the main body of detail is in Chapter Seven, Subsection 7.2.2.

Given that funding restraints prevented the setting of the size of a monetary incentive in the current New Zealand study at anything more than a peppercorn level, the evidence cited in the previous subsection suggests that any influence on behaviour associated with a payout rate of \$NZ0.30 would be unlikely to be different from that associated with a rate of \$NZ0.75 (the rough equivalent of Dusenbury's (1994) use of \$US0.50). The issue thus became one of balancing a reasonable return on attendance and on effort against the total size of the recruited sample.

The first hurdle to be surmounted was that of recruitment. Because some participants were likely to end up with a debt as distinct from a surplus, it was necessary to ensure that there was a flat fee component so that volunteers would participate in the first place. The requirement for the flat fee was that it needed be high enough to be sufficiently attractive to a fund-hungry organisation for it to produce willing participants (\$NZ8.33 per person to the first organisation and \$NZ10 per person to the other three). The total of all flat fee payments to organisations was \$NZ1,240.

The second hurdle was to provide a rate of New Zealand dollar remuneration for decisions made in the course of the experiment which, on the one hand, was sufficient to provide a stimulus for participants to regard the decision problems seriously; and on the other, was affordable to the researcher. Given the evidence marshalled in the previous subsection, the rate

did not need to be very high. Dusenbury (1995) covered much the same ground with the words²²:

My notion in this task is to give the subjects some involvement, some way to count, some context for making the cognitive effort required to evaluate choices consisting of many risky options. My guess is that it is the existence, not the magnitude, of the scoring system that matters.

Initially the sample was envisaged as being 120 people, and was later extended to 132. In advance there was no knowing what the average New Zealand dollar payout on the participants' experimental outcomes would be. Dusenbury (1994) reported his average was \$US8.75. The New Zealand study's average (net of the flat fee) was \$NZ3.79,²³ which equates with \$US2.50. The total payout (including flat rate bulk funding) was \$NZ1,748.20, which equates with \$US1,153.81.²⁴ When the total cost of the New Zealand cash incentive was divided by the number of participants (134 before two people's results were excluded), the full average cost per participant was \$NZ13.05, which equates with \$US8.61. This last figure is very close to Dusenbury's own average payoff of \$US8.75.²⁵

The focus is now turned to the second major ingredient of the experiment's method: the main experimental tool. In Chapter Six the scenario scripts and also the questionnaire are closely reviewed.

²² Dusenbury, R., (1995), A personal communication via email.

²³ This figure was computed as follows (Each amount is the total for one of the four subsamples, the order being the Anglican congregation, the kindergarten support group, the primary school support group, and the Lincoln University graduate student group): [$\$262.50 + \$138.60 + \$63.30 + \43.80]/134. Note that the denominator includes two participants who were subsequently dropped from consideration.

²⁴ The exchange rate used in this calculation was \$NZ1 = \$US0.66, which was the approximate value of the exchange rate in late 1995 and in January 1996. (New Zealand has a floating exchange rate which, measured in terms of basis points, varies from hour to hour with the American dollar.)

²⁵ Dusenbury, R. (1994), "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options", p. 9.

6. THE EXPERIMENTAL INSTRUMENT

6.1 INTRODUCTION

This section is devoted substantially to a discussion of the rewording of the scenarios used in the New Zealand laboratory experiment. The method by which the style of Dusenbury's five scenarios was reformatted for use in an Antipodean context was touched upon in Subsection 4.4.2 of Chapter Four. Attention is now drawn to the reasoning behind the design change, with the changes to the various scenarios explained in depth. This chapter also delves into the format of an extra scenario, Tax Case W; and provides a description of the end-of-session questionnaire, explaining the goals for which it was crafted.

In terms of the chapter's layout, the impetus for redesigning the scenarios is discussed in Section 6.2, while in Section 6.3, an explanation is given of various departures from New Zealand tax law made in order to reduce complexity in the experiment. In Section 6.4, attention is given to the issue of *case order*. This section contains a justification for the sequence in which participants received and answered the six scenarios. The next issues to be dealt with are the revisions to both the preliminary training script and to the sample scenario explained within it, and are contained in Section 6.5. In the next part, Section 6.6, closer attention is paid to the story-lines¹ employed in the five customised experimental scenarios; while, in Section 6.7, Dusenbury's additional rating question is reproduced and commented upon, along with notice of the inclusion of Summary Syndrome-related questions. In the following section, Section 6.8, the extra scenario, Tax Case W, is explained. Finally, in Section 6.9, the scope, contents and desired aims of the end-of-session questionnaire are traversed.

¹ The story-line is the short explanation of the taxpayer's context. It provides information on income, the amount of tax withheld, the level of over- or underwithholding, and the details of the nature of the item for which the correct tax filing disclosure is uncertain.

6.2 LESSONS FROM PRETESTING

The original training script and scenario scripts used by Dusenbury (1994) were pretested on a small group of New Zealand students and graduates in early July 1995. Eight of the pretest participants were classmates enrolled in the Master of Commerce degree course in the Department of Accountancy, Finance and Information Systems at the University of Canterbury; and the other three were a computer science graduate, a history graduate and an economics undergraduate. The format and language of the North American scripts were also commented upon by the Ethics Committee which had also been called upon to assess the material, albeit for quite a different purpose.²

What emerged was a clear consensus that the scenarios used by Dusenbury (1994) were couched in language which the New Zealand pretest participants and Ethics Committee members found difficult to penetrate. They said that rewording was necessary if the experiment was to be run successfully on taxpayers with a range of different reading competencies. Several of the (highly literate) pretesters admitted that after an initial attempt at thoroughness, they found it easier to ignore the scenario information and make their assessments solely on the financial information presented in each case's table. It was clear from this that the Dusenbury (1994) experimental instrument — in a New Zealand context anyway — failed to provide a clear connection in the pretesters' minds between the level of withholdings and the tax choices they needed to choose among. It was not that such a connection does not occur in the normal course of events; it was that the connection was buried in data that was thrown out as unrewarding. This, in fact, instigated the line of thought which matured into the formulation of the Summary Syndrome hypothesis, H_{5A} , described in Chapter Four, Subsection 4.5.2.

The problem was less a matter of jargon than it was of subtle differences in the nature of regional English. Quite technical information imparted with clarity in American English to a

² The Ethics Committee was formally required to assess the experimental procedures with respect to its ethical treatment of subjects. (The approval of this committee was a precondition of the New Zealand undertaking.) Suggestions made by members of this committee were passed on by their spokesperson, a professor in the University's Department of Philosophy and Religious Studies.

North American public is not necessarily so communicated in an Australasian context. Not only terms may be understood in a slightly different way, but phraseology may be substantially different as well. There was general agreement that the length of the instructions also acted as an impediment to easy comprehension.

This preliminary feedback brought about three modifications in the research design:

1. The training script and scenario scripts were reworded as described in the ensuing subsections.
2. The nature of the information provided about withheld taxes was changed to make them more familiar to New Zealanders.
3. Participants were required in two new questions to register the level of attention they paid to scenario information on a five-point Likert scale (which provided the data for addressing the Summary Syndrome hypothesis).³

6.3 SIMPLIFICATION OF TAX LAW IN THE EXPERIMENT

Over and above the considerations just contemplated, the recasting of the North American scenarios to suit New Zealand conditions also required simplifications of a more purely practical nature. Inevitably there had to be a trade-off between the requirements of the New Zealand Income Tax Act 1994⁴, which is a complex piece of legislation, and the need for clarity and simplicity in the experimental script. The experiment needed to be easy for subjects to understand and perform, yet at the same time retain the credibility of actual tax situations.

As a consequence, the participants were told in the introductory speech that, while the experiment used existing New Zealand tax rates, the regulations implied in the scenarios were not necessarily the same as those enforced by the Inland Revenue Department (which inevitably change over time). In fact, scenario requirements were less stringent and less complex. An instance of this sort of simplification could be found in participants being invited

³ These two questions were explained in Chapter Four, Subsection 4.5.2.

⁴ This Act, which came into law on 1 April 1995 (at which time this study was already under way), replaced New Zealand's Income Tax Act 1976.

to make claims with respect to items of income or expenditure which were not properly documented. In reality, there is little scope for claiming deductions or rebates in the absence of comprehensive documentary evidence. A second instance concerned unfavourable outcomes associated with risky prospects. If a decision was challenged at the end of the experiment (by the drawing of an adversely numbered chip from the bag of 100), the challenge equated with a summary adjustment made on the spot.⁵ By contrast, there is a drawn-out (and potentially expensive) process by which taxpayers may contest IRD rulings if they are challenged in actuality.⁶

6.4 ORDER OF PRESENTATION

Dusenbury (1994) recognised that his experimental subjects, left to their own devices, were likely either to mutually consult, or to pay at least some attention to how other participants were responding to each decision problem as it came up; and that this development would give rise to a confounding variable — the peer influence variable discussed by Jackson and Milleron (1986).⁷ Dusenbury minimised peer influence decision contamination by requiring his subjects to answer the *refund* scenario, Tax Case Y, first, so that everybody received refunds at roughly the same time; and then by jumbling the remaining four scenarios into four different orders so that no two participants could reasonably expect to be dealing with the same problem at the same time. This policy was followed and also explained in advance in the New Zealand study so that participants understood that it was their personal responses, uninfluenced by those of their peers, which were required. Hence also the participants were told they could communicate only with the experiment supervisors — and not their neighbours. Since the

⁵ In this respect the experiment set up an arrangement similar to that in the Netherlands, as explained by Helsing et al: “The Dutch tax system does not really have a strong deterrent to evading. ... When a tax form is found to be incorrect, the tax inspector simply corrects the tax form and collects the tax associated with it. ... The question of willful evasion of taxes does not enter the process at this stage. ... In fact, a punishment is only possible when, after the tax form has been processed and cleared, new information comes to the attention of the tax inspector that gives him or her reason to correct the amount declared after all and when there are enough grounds to suspect an intention to evade taxes.” Helsing, D. J., Elffers, H., Robben, H. S. J. and Webley, P., (1992), “Does Deterrence Deter? Measuring the Effect of Deterrence on Tax Compliance in Field Studies and Experimental Studies”, p. 298.

⁶ However, there has been some movement towards improvement of the disputes resolution process, as witnessed by recent legislative proposals.

⁷ Jackson, B. and Milleron, V., (1986), “Tax Compliance Research: Findings, Problems, and Prospects”, p. 136.

experimental scripts were stapled together as a booklet, it was feasible to record the ordering of the seven decision problems to enable testing for the influence of a possible scenario ordering effect on participants' risk preferences. The preliminary documents and the scenarios are now each commented upon in turn.

6.5 TRAINING OF PARTICIPANTS

6.5.1 Training Script

The training script began with the *statement of consent* in the form of a rubric required (and provided) by the Ethics Committee,⁸ covered in Chapter Five, Subsection 5.3.3 above.

The purpose of the training script was to provide enough information about the experiment as an exercise involving risky choices for participants to feel confident they knew what they were doing when they embarked upon the experiment itself. Many of the main points were converted into a bulleted form for ease of comprehension.

Since the most common form of scenario involved tax decisions, it was stated that the tax choices under investigation did not involve illegal behaviour. Instead they were choice problems in which a decision maker could not know in advance what was a legitimate solution and what would not be acceptable to the tax authority.

The script also contained Dusenbury's explanation of the penalty payments in terms the paying of taxes owed plus an additional \$300 "to cover all monetary and nonmonetary costs such as lost work time, professional service fees, anxiety over the deficiency etc".⁹ To reduce information overload, this information was not repeated in the tax scenarios.

Also, in keeping with Dusenbury's original script, the revised version contained a brief mention of the medical insurance and gambling scenarios, and provided a sample case involving day

⁸ See Appendix B, Section B.2.

⁹ For Dusenbury's training script, see Appendix A, Section A.1. For the revised training script used in the current study, see Appendix B, Section B.2.

care income. In keeping with the scenarios in general, however, the sample case was substantially reworded. This quite major change is discussed in the next subsection.

6.5.2 Sample Case

Apart from the change in presentation explained in Chapter Four, Subsection 4.4.2, the sample case provided as a dry run in the training script was left as it was used by Dusenbury (1994). The case concerned the impact of extra income earned looking after small children at home. The costs associated with earning this extra cash were not documented; and the participants were required to estimate how much of these undocumented costs they would claim, given the knowledge that Options **B** to **E** involved running various levels of risk that the tax authority would challenge the claim.

The session supervisor explained this case verbally; and participants could choose whether they provided an actual answer to the problem and to the concomitant preference order question. The sample case served as a useful bridge between the verbal and written instructions and details of the first experimental scenario, *refund* Tax Case **Y**, which is described next.

6.6 SCRIPTS USED IN THE CONCEPTUAL REPLICATION

6.6.1 Refund Tax Case **Y**

As explained in Section 6.4 above, the first scenario to be dealt with, in the experiment proper, was the *refund* tax scenario, Tax Case **Y**. Where Dusenbury used a story-line involving a questionable deduction, the New Zealand replication equivalent used a story-line about a questionable item with respect to the availability of a rebate. It is recognised that a deduction of costs incurred in the earning of income reduces the level of income upon which tax is levied, while a rebate is a direct reduction of tax payable; and if the prevailing tax rate is 24 cents in the dollar, a dollar of rebate equates with a reduction of income actually taxed by the sum of \$4.17. However, the participants' attention was not drawn to this hidden aspect of the information; and in fact it is probably irrelevant. The issue at stake was whether an

overwithholding of PAYE explicitly spelt out as \$400 overpaid (ignoring the rebate), would cause participants to claim some or all of that questionable rebate in order to increase their refunds. In other words, the rebate was presented as a tool for redressing a known overwithholding.

It is also important to note there is no element of surprise or divergence from expectation built into this, or any of the other scenarios. This eschewal of thwarted expectation distinguishes Dusenbury (1994) and the present study from Schepanski and Kelsey (1990).¹⁰

However, further explanation of the decision to incorporate a rebate in the scenario would be cogent at this point. While the items on which rebates may be claimed under current New Zealand tax law have been pruned significantly to only very few in number, most taxpayers are more conversant with rebates than they are with the deduction of costs. In fact, the right to deduct expenses is now granted only to taxpayers in self employment.¹¹ Since Tax Case Y was everybody's first decision problem, the more familiar rebate concept provided a more immediately familiar problem to resolve.

The level of gross income stated in Tax Case Y was \$24,166, which is fairly low in New Zealand terms. In February 1993 the average wage/salary income was \$28,897.¹² The average annual income over all industries in May 1995 had risen to \$36,259.¹³

The income and withholding figures used in the other tax scenarios were allowed to differ from the \$24,166 for the sake of verisimilitude, as were the levels of tax said to be prepaid. While Dusenbury was motivated to keep his income levels uniform by a desire to avoid framing

¹⁰ Schepanski, A. and Kelsey, D., (1990), "Testing for Framing Effects in Taxpayer Compliance Decisions". Also see Dusenbury, R., (1994), "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options", p. 3.

¹¹ The exception to this situation is that costs associated with having returns prepared by a tax professional are still reclaimable.

¹² This figure is computed from information provided in the *New Zealand Official Yearbook 95*, pp. 330 - 331. In February 1993 the average weekly income for wage/salary earners over all sectors, and based on average ordinary time, was, in New Zealand dollars, \$555.72, which equates with an annualised figure of \$28,897. In terms of the sexes, males on average earned \$619.40 weekly (or \$32,208 per annum); and females earned \$478.01 (or \$24,856 annually).

¹³ This figure was calculated from the weekly income figure for all earners provided by Statistics New Zealand in *Key Statistics October 1995*, Table 4.02, p. 44. Total average earnings, when the sexes were undifferentiated, were \$697.29 per week. This publication also noted that all wage/salary rates (inclusive of overtime) had risen 1.3 percent in the year to March 1995, which was an increase from the 1.2 percent rise recorded for the year ending December 1994. (p. 42).

effects from income impacting as a possible confounding variable, in the New Zealand experiment it was considered more important to make the scenarios as individualistic as possible to avoid fuelling the Summary Syndrome effect discussed in Chapter Four, Subsection 4.5.2. If as many as possible of the particulars were fresh in each scenario, then maybe information absorbed with respect to previous scenarios would not be impounded into later decisions — so long as the story-line information in the scenario of the moment was adequately digested.

On the same basis, levels of tax withholdings were also allowed to vary from scenario to scenario; but in Tax Case Y, Dusenbury's original overwithholding of \$5,400 was used. However, the more important figure was the additional figure printed in bold, which stated the level of overwithholding — in this case, \$400. The aim of this information was to provide fertile ground for the formation of a gain-related decision frame.

6.6.2 High Pay Tax Case X

In the *high pay* scenario, Tax Case X, the story-line was not materially altered from that used by Dusenbury (1994). The exceptions were the dollar amounts of existing tax withholdings and the stated level of income. The level of withheld tax in this instance was \$9,621, which was stated in bold print to be an underwithholding of \$800. The income stated in Tax Case X was \$40,000. While slightly higher than the average income figure of \$36,259 computed for all industry groups for May 1995 and cited in the previous subsection, it was not unacceptably higher.

To prime the subjects that they were not salary earners but self-employed in this scenario, the deductible item was explicitly labelled a 'business expense' and the withheld tax was referred to as provisional tax, which is the New Zealand term for taxes prepaid by firms and self employed individuals. This switch between salaried employment and self employment in the scenarios was considered feasible on the ground that there have been significant shifts in the nature of New Zealand employment recorded over the past decade; and people who have been

employed by the state or by private enterprise may well have also had experience in self employment.¹⁴

6.6.3 Low Pay Tax Case Z

In the *low pay* tax scenario, Tax Case Z, both the stated income figure, the level of withheld tax and the story-line remained unaltered from Dusenbury (1994). However, the wording was altered so that uniformity of format would be preserved across all of the New Zealand scenarios. The withholding is referred to as “provisional tax already paid up”.

6.6.4 Medical Insurance Case

Dusenbury (1994) talked in terms of buying a health care plan. Since in New Zealand parlance people think in terms of medical insurance policies, this small change in terminology was put into effect. There is salient dollar amount over- or underwithheld in this scenario. It is designed to be contrasted with Tax Case X on the basis that the riskless and risky prospects are identical while the story-lines differ dramatically.

6.6.5 Gamble Case

Apart from the change in format standard in the New Zealand scenarios, no significant alterations were made with respect to the gamble. Again, there is no salient monetary amount over- or underwithheld; and again the crucial issue is comparability with a tax scenario (Case Z), with which it shares identical risks and returns.

¹⁴ The total number of employed persons increased overall by 17,300 (1.2 percent) in the five years to the end of 1994. This increase breaks down into a decrease in wage and salary earners by 20,400 (1.7 percent) and a corresponding increase in self-employment by 21,700 (a 13.7 percent increase). Not only that, but there has been an increase in the number of employers by 9,300 (7.9 percent) and in the number of people not officially drawing wages while working in a family business (6,300 or a 64.3 percent increase). Source: *New Zealand Official Yearbook 95*, p. 319.

6.7 ADDITIONAL RATINGS QUESTIONS

At the end of whichever scenario constituted the fifth decision problem, Dusenbury (1994) asked his subjects to complete one additional rating question. They were required to rank the five scenarios (he called them judgment situations), according to their perception of the riskiness associated with each relative to all of the others. To preserve strict transitivity in the responses, Dusenbury required that no two or more scenarios could be given the same risk level. Hence the cases were to be ranked on a five station spectrum from **A** (least risky) to **E** (most risky). Because the **A** to **E** spectrum — given the imposed transitivity requirement — could be confused with the responses elicited in the scenarios, the spectrum was converted to a **1** to **5** scale.¹⁵ The question was¹⁶:

For this rating please **use each rating (1,2,3,4 and 5) once and only once**. Assign the 1 (the least risky option) to the Judgment Situation for which you are least willing to take risk, and so on. This means you should assign the 5 (the most risky option) to the Judgment Situation for which you are willing to take the most risk.

Please look back at all five of these Judgment Situations and compare them to make this additional rating.

Use each rating (1,2,3,4,5) once and only once:

TAX CASE X	_____
CHOOSE A GAMBLE	_____
TAX CASE Y	_____
PAY MEDICAL INSURANCE	_____
TAX CASE Z	_____

This question solicited an overview of the participants' perceptions of the decision problems; but because of the imposed transitivity, it did not usefully meet the need it apparently met in the North American study, which was to provide a secondary measure for detection of risk-level differences between responses to the various cases. Routinely during the various

¹⁵ Initially, Dusenbury's (1994) use of alphabetical ranking was used; but this caused many participants to ask for clarification. (Further comment is made in the text in the next paragraph.)

¹⁶ The bold type was used in the instrument as it was presented to participants. The aim, as ever, was to drive the most important information home strongly.

sessions, participants would query the strict transitivity aspect, saying that they could not meaningfully distinguish between this scenario and that (or those). In some instances, they ignored the instructions and ranked decision problems equally anyway. The question was nevertheless included for completeness.

At this point the two new questions relating to the Summary Syndrome extension (reproduced and discussed at length in Chapter Four Subsection 4.5.2) were inserted. The research question in this extension was as follows: do participants react to laboratory scenarios in the way researchers intend that they react, or do they develop their own simplified methods for solving the decision problems, which subtly subvert the nature of the experiment? Once this question was dealt with, participants were invited to move on to the experiment's final two decision problems.

6.8 EXTRA SCRIPTS

6.8.1 Introduction

The two extra scenarios, *high pay* Tax Case **W** and *refund* Tax Case **V**, which were added to the main body of the experimental instrument, but which are not to be viewed as necessarily being a formal part of the study, were formatted in the same manner as the customised replications of the five Dusenbury (1994) scenarios. Since *high pay* Tax Case **W** was formally incorporated into the study via the Risk Profile Stability Extension by providing the data required for testing H_{6A} , a description of Tax Case **W** would be useful here.

6.8.2 High Pay Tax Case W

The version of Tax Case **W** described here is the one received by the first 83 subjects. It was useful because the risks associated with Options **B** to **E**, in this version, were unchanged from those of the five replication scenarios.

Case **W** was another business expenses case although the word *business* was not used. The questionable item was a major piece of equipment on which it was unclear how much

depreciation could be claimed. The stated level of taxable income, if this depreciation was left unclaimed, was \$41,148; and the level of provisional tax prepaid was \$10,000. In ball-park terms, this level of provisional tax, which is the highest in all of the tax scenarios in the experiment, could be considered not to be appreciably different from the \$9,621 in Tax Case X, which is the two cases with which Case W is directly compared. By contrast, the other tax cases, Y and Z, have much lower levels of withholding at \$5,400 and \$4,800 respectively. Kahneman and Tversky (1984) note, that for both gains and losses, the absolute difference between two small amounts has much less impact on decision makers than the same absolute difference between two much larger amounts. A \$379 difference between \$9,621 and \$10,000, for example, has but a fraction of the impact of the difference between a withholding of \$100 and one of \$600.¹⁷ The same argument applies to the level of stated income in Tax Cases X and W.

As ever, the most accessible piece of information is the level of underwithholding, which, at \$800, is identical with the underwithholding in the matching scenario, Tax Case X.

Once the peripheral material, Tax Cases V and W, had been completed, participants were directed to turn their attention to the questionnaire.

6.9 QUESTIONNAIRE

6.9.1 Introduction

Participants were told that the purpose of the questionnaire was to provide a background of facts and opinions which would be compared with responses to the decision problems. In fact, the questions were asked on the ground that the information provided might have some bearing on how participants arrived at the decisions they made in the course of the experiment. The answers were solicited in a tick-in-box format, which, given that almost every question

¹⁷ Kahneman, D. and Tversky, A., (1984), "Choices, Values and Frames", p. 342. Kahneman and Tversky call this phenomenon an instance of *psychophysical measurement*; and they suggest that subjective value is a concave function of the size of a gain. They note that the same generalization applies to losses as well.

involved a participant's identifying his or her position on a defined spectrum, meant that the responses could be interpreted with ease in terms of a Likert scale. The confidentiality of the participants' responses was stressed.

The 19 questions were divisible into three categories. One category asked for factual information about the participant and his or her background; the second solicited a set of opinions on issues such as the fairness of the health system, the morality of gambling, and the morality of non-compliance; and the third category required the participant to disclose his or her perceptions of IRD tax-gathering practices and the incidence of taxpayer compliance in general. The questions were presented in a haphazard order to minimise perceptions of any systematic direction in the line of enquiry which might require a set of politically correct answers. The categories of questions are now examined in turn. Since not every question is reproduced in the following subsections, the questionnaire is printed in full in Appendix B, Section B.4.

6.9.2 Category One: Demographic Questions

The first category solicited background facts. While all eleven of these questions were likely to furnish cogent responses, of particular interest was Question 3, which was about participants' tax-filing experience in terms of years in which returns had been filed. Dusenbury (1994) found this variable had a significant impact on preferred levels of risk in his tax filing judgement situations. He found that taxpayers with greater experience chose riskier options in both the *high-pay* and *refund* cases.¹⁸

The full list of demographic questions in the category is provided in Table 6.1.

Also of immediate interest is Questions 13, which solicits information regarding household income. Answers to this question enabled the participants to be segregated into income groups whose risk profiles could be examined separately in order to detect any personal income-

¹⁸ Dusenbury (1994), op. cit. n. 10, p. 13.

TABLE 6.1
Demographic Questions in Numerical Order.

Q1	Gender;
Q2	Age;
Q3	Tax filing experience in years;
Q9	Frequency of payment-due tax filings;
Q12	Household income by number of contributors;
Q13	Household income by (New Zealand) dollar amount;
Q14	Nature of income earning experience, i.e., source of employment;
Q15	Education;
Q17	Incidence of professional tax-filing assistance;
Q18	Knowledge of <i>Expected Utility Theory</i> ;
Q19	Knowledge of <i>Prospect Theory</i> .

sourced influence on the risk preference shifts examined in the hypotheses.¹⁹ As stated in Chapter Four, Subsection 4.4.2, this precaution has already been taken in *Prospect Theory*-based compliance research by Hite, Jackson and Spicer (1988).²⁰ Again, income level was a variable which has been discussed in depth in Jackson and Milleron's (1986) survey of the literature, along with age and sex.²¹

One of the motivations behind Question 14, *Nature of Income Earning Experience*, was similar to the income level issue. The underlying point of interest was whether the risk preferences of participants who earned salaries or wages were measurably different from those of self-employed participants. Clotfelter (1983)²² and Wallschutzky (1984)²³ reported that self-employed taxpayers are perceived to have much greater scope for non-compliance; and the phenomenon is discussed in considerable detail by Jackson and Milleron (1986).²⁴ It is to be

¹⁹ Question 12 was formulated as a question-in-waiting which could be used in an analysis of the responses to Question 13 if the significance of an income-related variable (or social prejudice variable) required that its various observation levels be more closely interpreted. However, the variable generated in Question 13 was found to be relatively insignificant; therefore exploration of this ramification was not deemed relevant.

²⁰ Hite, P. S., Jackson, B. R. and Spicer, M. W., (1988), "The Effect of Framing Biases on Taxpayer Compliance".

²¹ Jackson and Milleron, (1986), op. cit. n. 7, pp. 130 - 134.

²² Clotfelter, C. T., (1983), "Tax Evasion and Tax Rates: An Analysis of Individual Returns", p. 368. Also see Chapter Three, Section 3.2.

²³ Wallschutzky, I. G., (1984), "Possible Causes of Tax Evasion", p. 381. Also see Chapter Three, Section 3.2.

²⁴ Jackson and Milleron (1986), op. cit. n. 7, pp. 134 - 135. Also cited in Chapter Three, Section 3.2.

noted, however, that these researchers concerned themselves with the fact there was a difference in tax avoidance and evasion opportunities; but they did not focus on the possibility of an employment-related conditioning effect giving rise to measurable differences in propensity for risk.

Most of these demographic questions required a tick in a box representing the appropriate range — be it age, filing experience, and so on — and the sequence of boxes could be reinterpreted as a Likert scale. The question which was different was Question 15. Higher education levels have been linked with increased non-compliance, as documented in the overview of compliance variables published by Jackson and Milleron (1986).²⁵ In the current study the question asked, is does education, as an extraneous variable, have an impact on risk preferences in a tax setting? The question was worded in the following way:

Q15: Education:

For this question, tick **every** box which applies to you. Which of the following describes your educational history:

- School education
- Completed a polytechnic qualification
- Completed a university qualification
- Completed some other form of tertiary education qualification
- Self-taught in your area of expertise

In order to encourage participants to reveal the information best suited for use as a maximum level of education variable, the Question 15's boxes were arranged in a jumbled order to avoid disclosure of any belief, on the researcher's part, that any level was superior to any other level; and the participants were asked to fill in *all* of the boxes which were relevant to them.²⁶ The Likert scale underlying the question ranged through stations on the following spectrum, where school education equated with 1 and the university qualification equated with 5.

²⁵ Ibid, pp. 132 - 133.

²⁶ This facilitated the systematic selection of a participant's single highest educational level according to the researcher's criterion rather than the participants' various (probably equally valid) interpretations.

- School education
- Self-taught in your area of expertise
- Completed some other form of tertiary education qualification
- Completed a polytechnic qualification
- Completed a university qualification

Because they are relevant to any formal description of the sample used in this experiment, responses to the gender, age, tax-filing, income source, and knowledge of *Prospect Theory* and *Expected Utility Theory* questions are presented in summary form in Chapter Nine, Section 9.2. This section also presents summary information on the frequency of end-of-year tax bills experienced by the participants.

6.9.3 Category Two: Solicitation of Attitudes

6.9.3.1 Introduction

The second category of questions solicited information about attitudes. These are laid out in Table 6.2. Again the purpose was to determine if any underlying predisposition significantly modified the choice of responses made with respect to the decision problems.

TABLE 6.2	
Questions Soliciting Participants' Attitudes.	
Q4	Attitude to health insurance;
Q5	Attitude to tax with respect to changes in the public health system;
Q6	Attitude to non-compliance by higher income earners;
Q7	Attitude to non-compliance by lower income earners;
Q8	Feelings about actual taxpaying;
Q11	Feelings about gambling.

The questions in this category were more complex in that they required the participants to commit themselves ideologically to a site in the given spectrum allowed for answers. In order that this spectrum be disclosed, each of the questions is repeated in full here in the following subsections.

6.9.3.2 Attitudes Stemming from Changes to the New Zealand Public Health System

Question 4 was asked because the public health system in New Zealand is at present being restructured.²⁷ Given the level of public debate on the issue, is possible that participants may hold strong views on the system's present state, which might influence their response to the *paying medical insurance* scenario. Those who have trust in the public health system, as it is currently structured, could be expected to be more risk willing in terms of the options in the scenario than those who lack this trust. It will therefore be useful to partition the experimental results by sites or groupings of sites on the spectrum provided in the question:

Q4: Attitude to health insurance:

Given the current state of the public health system in New Zealand, how worthwhile do you consider a personal health insurance scheme to be for you or other members of your family?:

Please tick the box which most closely approximates your judgment of it:

- It is ridiculous
- It is not very necessary
- I am neither for nor against
- I think it is a reasonable thing to have
- I would not be without it

The next question, Question 5, covered the same ground from a slightly different angle. When the sample was partitioned by the different responses made to this question, was it possible to detect discernible differences in participants' risk profiles with respect to paying for medical insurance? And were there discernible differences in risk profiles with respect to the decisions made in the tax scenarios?

²⁷ Until the late 1980s, citizens could rely on the public health system to meet their emergency medical requirements; but it did so in what has been argued to be a slow and economically inefficient fashion, conducive of long waiting lists and wastage of resources. An analysis of this situation, and the changes which have since been effected, is irrelevant to the current study.

Q5: Attitude to fairness of tax with respect to changes in the public health system:

Given the level of tax levied at present in New Zealand, do you consider that changes in the public health system are:

- unacceptable
- have some, but overall questionable merit
- undecided
- are basically OK, but have some flaws
- acceptable

6.9.3.3 Attitudes towards Non-Compliance

There are three questions in this subsection. The first two, Question 6 and Question 7, solicit information on attitudes towards evasion; and both entail recognition of a fairness variable, which is yet another compliance variable recognised and overviewed in Jackson and Milleron's (1986) encyclopaedic study.²⁸ Questions 6 and 7 are used in three different ways:

1. Is there a connection between the implied fairness variable contained in Question 6 and propensity for risk willingness over the tax scenarios?
2. Is there a connection between the implied fairness variable contained in Question 7 and propensity for risk willingness over the tax scenarios?
3. Do people who choose the *never* box in both Question 6 and Question 7 exhibit total risk aversion with respect to the tax scenarios? If so, what inference may be drawn from this? Is there any clear connection between these people and those who have indicated that they disapprove of gambling in their response to Question 11?

The formats of these questions were as follows:

Q6: Attitude to non-compliance #1:

Given the tax laws that are in place in New Zealand at the present time, would you consider it acceptable for a taxpayer earning over \$35,000 to under-report income that he/she would have the power to conceal?

²⁸ Jackson and Milleron, (1986) op. cit. n. 7, pp 137 - 138.

- never
- seldom
- on average
- mostly
- always

Q7: Attitude to non-compliance #2:

Given the tax laws that are in place in New Zealand at the present time, would you consider it acceptable for a taxpayer earning less than (say) \$18,000 to under-report income that he/she would have the power to conceal?

- never
- seldom
- on average
- mostly
- always

The third question in this grouping was Question 8. This question canvassed participants directly about their attitude towards terminal tax bills. It relates to the *Prospect Theory* concept of payouts being framed as losses relative to a neutral reference point. The closer a subject scored towards 5 on the implicit Likert scale, the less valid the link becomes between tax payouts and perceptions of personal loss; and the more apparent becomes the existence of another possibility in the form of a patriotic duty variable. Participants who consistently choose the riskless prospect, Option A, may well be more strongly influenced by this sense of duty than by the posited framing effect. The question enabled this species of participant to be separated from the main body and the two groups to be tested separately as well as together.

Q8: Feelings about actual taxpaying:

When you have found yourself in the position of owing more tax to the IRD, which single box of the following would come closest to most accurately describing what you have experienced?

- I am very unhappy: This is a loss to me and an imposition.
- This is something I feel moderately unhappy about; but it is necessary.
- No strong feeling.
- I'm moderately happy to do the right thing.
- I'm happy to fulfil this requirement: it is my contribution to society.

6.9.3.4 Attitude on Gambling

There is one question in this subcategory, Question 11. This question set out to provide an insight into responses made in the *Choose a Gamble* case. This issue was connected closely with the testing of Dusenbury's third hypothesis in the experiment.

Because one of the subsamples was recruited from an Anglican congregation, the possibility existed that a disproportionately high number of the volunteers from this source might have strictly negative views on the morality of games of chance.²⁹ There was the possibility, of course, that people with views of this sort were also recruited via the other three contributing organisations. Question 11 enabled this category of participant to be quantified both in terms of the full sample and of the four subsamples.

The question was as follows:

Q11: Feelings about Gambling:

This question is about how you view the rights and wrongs of gambling. Your attitude to spending say \$5 a week on a Lotto ticket is most closely approximated by which single box:

- I would never buy a ticket because gambling is always wrong.
- I am happy to support a charity that funds good causes; but I consider regular ticket-buying to be not right for me.
- I do not have an opinion either way.
- I am quite happy to have a regular flutter of this sort; but if Lotto was banned, I would not miss it.
- I feel good about playing Lotto and would resent it if it were banned.

In the light of their responses to this question, the extent of their impact on the sample could be measured.³⁰

²⁹ This possibility stems from the fact that active membership of a religious denomination implies the likelihood of a strong sense of ethics. Gambling has traditionally been viewed by Christians as a sin.

³⁰ The fact that people of this nature exist and pay taxes is beyond doubt. It is an interesting question — which cannot be answered here — as to what percentage of the taxpaying public this group actually constitutes. But only a fully random sample could provide a reliable answer to this.

6.9.4 Category Three: Perceptions of Taxpayer and IRD Behaviour

The two remaining questions are grouped in a separate category because they solicit information regarding the participants' assumptions. They are listed in Table 6.3:

TABLE 6.3	
Questions Soliciting Participants' Tax-Related Perceptions.	
Q10	Perception of likelihood of an audit by the tax authorities;
Q16	Awareness of general compliance with tax laws.

Both questions are concerned with beliefs about the incidence of certain behaviours.

The first, Question 10 was formatted as follows:

Q10: Perception of likelihood of an audit by the tax authorities:

At present you believe that the IRD fully audits (as distinct from just checking the arithmetic) what proportion of tax returns that cross its desks?
Tick the box which most closely approximates your own estimate:

- | | | |
|-------------------------------------|-------------------------------------|--------------------------------------|
| <input type="checkbox"/> 1 percent | <input type="checkbox"/> 5 percent | <input type="checkbox"/> 10 percent |
| <input type="checkbox"/> 15 percent | <input type="checkbox"/> 20 percent | <input type="checkbox"/> 25 percent |
| <input type="checkbox"/> 30 percent | <input type="checkbox"/> 35 percent | <input type="checkbox"/> 40 percent |
| <input type="checkbox"/> 45 percent | <input type="checkbox"/> 50 percent | <input type="checkbox"/> 55 percent |
| <input type="checkbox"/> 60 percent | <input type="checkbox"/> 65 percent | <input type="checkbox"/> 70 percent |
| <input type="checkbox"/> 75 percent | <input type="checkbox"/> 80 percent | <input type="checkbox"/> 85 percent |
| <input type="checkbox"/> 90 percent | <input type="checkbox"/> 95 percent | <input type="checkbox"/> 100 percent |

This question takes the study into the domain of *Expected Utility Theory*. According to Allingham and Sandmo's (1972) model, detection probability (audit risk) is one of the valid policy tools available to a government seeking to minimize non-compliance.³¹ The evidence marshalled by Jackson and Milleron (1986)³² providing empirical support for the variable's impact on taxpayers is mixed. Noncompliant taxpayers have tended to have been associated with lower detection rate perceptions than those associated with compliant taxpayers; but

³¹ Allingham, M. G. and Sandmo, A., (1972), "Income Tax Evasion: A Theoretical Analysis", p. 338.

³² Jackson and Milleron, (1986), op. cit. n. 7, pp 140 - 141.

Warneryd and Walerud (1982)³³ found no significant connection, and Spicer and Thomas (1982)³⁴ found a significant correlation only when the subjects of their study were given precise information of the risk of detection. Nevertheless, it is of value to review the decision-making recorded in the current study for any extraneous influence traceable to underlying beliefs with respect to the probability of an audit by the IRD.

The purpose underlying Question 16 is similar. In this instance, participants were asked to disclose their concept of what is normal taxpayer behaviour.

Q16: Awareness of general compliance with tax laws:

You believe (or have at least the gut feeling) that taxpayers in general, where they have freedom of action to do so, will tend to:

- Fully report all taxable income to the very last cent.
- Report all taxable income give or take the last few dollars.
- Sometimes under-report taxable income.
- Usually under-report taxable income.
- Under report as much and as often as they possibly can.

That perceptions of what other people do have an influence on taxpayers' levels of compliance has been long known. Jackson and Milleron (1986) summarized two aspects of this influence. In the first instance, both *Deterrence Theory* advocates and exponents of *Reference Theory* have examined the link between compliance and familial and social peer group influence; and in the second instance, compliance has been linked with geographical location.³⁵ Porcano (1988)³⁶ also found that perceptions of existing evasion were significantly related to past evasion and also to intentions of evading in the future. In a New Zealand setting, Hasseldine, Kaplan and Fuller (1994)³⁷ found the number of evaders known to a respondent to be a significant predictor for the under-reporting of income. Collection of data on this variable provides the opportunity to link peer influence (in general, as distinct from across the table

³³ Warneryd, K. and Walerud, B., (1982), "Taxes and Economic Behavior: Some Interview Data on Tax Evasion in Sweden". This was cited by Jackson and Milleron (1986), op. cit. n. 7, p. 140.

³⁴ Jackson and Milleron, (1986), op. cit. n. 7, p. 40. Although Spicer and Thomas' (1982) paper was cited by Jackson and Milleron in the body of their text, they neglected to record this paper in their bibliographical references.

³⁵ Ibid, p. 136.

³⁶ Porcano, T. M., (1988), "Correlates of Tax Evasion", p.58.

³⁷ Hasseldine, D. J., Kaplan, S. E. and Fuller, L. R., (1994), "Characteristics of New Zealand Tax Evaders: A Note", pp. 87 - 88.

during the experiment) with participants' risk profiles. The purpose is to confirm (or disconfirm) the influence of the variable in a new context, which is whether or not it has any impact on the hypotheses adapted from Dusenbury (1994).

At this point the focus shifts to the third and final major ingredient of the experiment's method, which is the recruited sample. This is discussed in Chapter Seven.

7. THE SAMPLE

7.1 INTRODUCTION

The focus of the research method is now turned to the sample. Since this is quite a large topic, this chapter has been divided into five sections; and these, in turn, have been further subdivided. Section 7.2 deals with the strategy adopted for recruitment of the sample in general; while in Section 7.3, the focus narrows to provision of recruitment and session details with respect to each of the four subsamples. The need for modifications, once the experimental sessions were underway, is dealt with in Section 7.4; and the final section of the chapter, Section 7.5, provides sample size specifications with regard to the hypotheses laid out in Chapter Four.

In general, the thrust across these subsections entails a narrowing from broad to fine details, commencing with consideration of why a non-random sampling method was adopted.

7.2 RECRUITMENT STRATEGY

7.2.1 Introduction

Two issues are dealt with in this section. A justification for the recruitment of a non-random sample is given; and the actual recruitment strategy is explained.

7.2.2 Why the Sample was Non-Random

The subjects in the New Zealand experiment were chosen on a non-random basis. Non-random recruitment is standard in laboratory research concerning compliance in taxation matters on the ground that random recruitment via local parliamentary electoral rolls, business directories or telephone books would be difficult, given popular perceptions of tax research as something in which the Inland Revenue Department (or other national/state taxation authority) might have a less than disinterested hand.

The inclusion of the cash flow extension in the New Zealand experiment provided a second, more cogent reason for non-random recruitment. If some subjects found themselves cash starved to the extent of running a large risk of ending the experiment in debt, a rational independent participant could be expected to walk out rather than complete the experiment. In this respect, the experiment involved a trade-off between the exigencies of *external validity* as assured (debatably) by the monetary incentive (which became a demand for money if a participant ended up in the red), and the risk of bias inherent in a non-random sample. The monetary incentive was mentioned in brief in the discussion of Dusenbury's (1994) research design in Chapter Four, Subsection 4.2.4, and was further discussed in Chapter Five, Sections 5.2 and 5.4.

7.2.3 Recruitment Technique

The sample consisted of 132 Christchurch residents with tax-paying experience, recruited in four batches.¹ Each batch was to be provided by an organisation whose members were known to be actively involved in fundraising on its behalf. The experiment was marketed as a relatively effortless means by which the organisation could raise funds in return for the sacrifice of part of a morning or evening by an agreed number of member volunteers. The technique, as described so far, was used in the recruitment of part of Dusenbury's (1994) sample. Dusenbury negotiated a group donation arrangement with two sets of subjects, amounting to 18 people. Instead of being paid off personally at the conclusion of their session like the other 47 participants, these people were given a chit recording their personal US dollar outcome, and the cash was forwarded directly to their organisation.² (The other 47 were also recruited from bodies of people who held some form of group membership in common; but had agreed to take part as independent agents.)

¹ As explained in Section 7.4, this figure is the net figure of useable recruits. Two extra people performed the experiment; but their scripts were discarded when it was found they had not had any tax return-filing experience.

² Dusenbury, R., (1994), "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options", pp. 9 - 10.

With respect the monetary incentive requirement, the recruitment strategy had a further twist. In order to accommodate the possibility of some participants being responsible for actual monetary losses, the organisations providing volunteers were told they would receive a minimum sum. This minimum in the case of three of the four subsamples, was equal to ten dollars per person recruited. The organisations were advised that this sum would be augmented at the rate of \$0.30 per pseudo \$100 held by the participants at the end of the simulation. The organisations were also told that this increase would be moderated by the deduction of \$0.30 for every pseudo \$100 owed by individuals who ended up in debt. In this manner, individuals could be put under pressure to perform for real monetary stakes, knowing that unfortunate, maybe misjudged, decisions might result in real monetary losses, and that the organisation they represented stood to gain an amount net of these losses.

7.3 ORGANISATIONS

7.3.1 Introduction

The participant groups provided by the four organisations from this point will be called subsamples; and the term sample will pertain to the full complement of 132 subjects. Each of the four subsamples of participants will be reviewed in turn.

7.3.2 The Church Subsample

The first organisation to be approached was the congregation of an Anglican church in one of the city's more affluent suburbs,³ as represented by a liaison person from the Vestry Committee. This committee guaranteed 60 participants in return for a \$500 minimum donation. In the negotiation it was agreed that the optimum arrangement would be three laboratory sessions to be held at the University of Canterbury, with an attendance of twenty church members at each of them. The difficulty — even for an organisation with a large and committed following — of co-opting and corralling this number of eligible and willing people

³ Avonhead, Christchurch.

was aptly illustrated by the lengths to which the Committee found it was necessary to go in order to accommodate those who agreed to take part. It was necessary to run six sessions ranging in size from twenty participants down to two in order to get the full complement of sixty participants agreed upon.

7.3.3 The Kindergarten Subsample

The second organisation agreed to supply forty recruits in return for a minimum sum of \$400, increasable and modifiable as described above. This body was the committee elected to oversee and support a kindergarten situated in the relatively affluent suburb in which the University itself is situated.⁴ Again the committee members charged with finding suitable volunteers able to spend an evening at the University of Canterbury, which is within two kilometres of the kindergarten, encountered significant difficulties. Three sessions were run, as before, at the University, with attendances of fourteen, ten and three, respectively. However, obtaining the final eight kindergarten parents/supporters required a change of tactic, since no further kindergarten associates were prepared to volunteer for a tax-related study in what was referred to as the University's 'alien environment'.

Two further sessions were held at the kindergarten itself a full month later — a morning session of three, and an afternoon session of eight mothers, several of whom agreed to join in at the last minute.⁵

7.3.4 The School Subsample

The third organisation was a parent support group associated with a primary school in a relatively poor suburb.⁶ In this instance only one session was held. Twenty parents (19

⁴ Ilam, Christchurch.

⁵ If there was a difference between these last two sessions and those held at the University, it lay in occasional interruptions from either sons and daughters in the kindergarten class wandering in from outside, or from babies in carry-cots or on rugs at the participants' feet. Nevertheless these women were still valid subjects. Of the thirteen who completed the experiment under these conditions, only one had to be discarded on the ground of filing inexperience; and of the other twelve, six were currently holding down employment, four had university degrees, six more had other tertiary education qualifications, and six of the twelve had each accumulated between 16 and 20 years of experience in filing tax returns.

⁶ North New Brighton, Christchurch.

mothers and one father) participated in the experiment in the school staffroom. The minimum donation agreed upon in this instance was \$200.

7.3.5 The Lincoln University Subsample

The fourth group recruited for the experiment was provided by the Centre for Resource Management at Lincoln University, located on the edge of Christchurch. It consisted of fourteen graduate students who were processed in a single session. This group, like the other three, was keen to raise funds for some common purpose. Twenty volunteers were requested, fifteen turned up at the appointed time, and one withdrew. (She was the only person to do so in the course of the study.) The minimum payout was established at \$140.

7.4 MODIFICATIONS TO THE SAMPLE OF PARTICIPANTS

Initially it was intended to run the replication and the existing extensions on a sample of 120 tax return-filing members of the public, obtainable from the first three organisations listed above. When it became obvious the three organisations already contributing an agreed upon number of volunteers were having difficulty recruiting their full quota, participants were solicited from a fourth, hitherto unconsidered organisation. This action, it was hoped, would guarantee a minimum size of 120 subjects.

Once the running of the sessions was underway, a further reason for increasing the size of the full sample became apparent. A small number of the participants already recruited turned out to be unsatisfactory in some manner. At this point it was also impossible to know how many more would be found to be unsatisfactory on grounds not perceived until the results were entered into a spreadsheet and subjected to a preliminary analysis.

Soliciting volunteers from this fourth group raised the total count of participants to 134, but this was reduced to 132 by the discarding of two participants on the ground of zero years tax filing experience disclosed in response to the experimental questionnaire. One of these was from the

church congregation, and the other, from the kindergarten parent and supporters group. These were the only two participants discarded for any reason from the sample.

7.5 SAMPLE SIZE WITH RESPECT TO THE HYPOTHESES

7.5.1 Introduction

Because the hypotheses laid out in Chapter Four are, for various reasons, testable on samples (or subsamples) of a range of different sizes, this issue is dealt with here in a separate section. The size of the sample appropriate to the hypotheses is explained in each instance, along with the reason for its selection. In all instances, the size data is disclosed at recruitment subsample level as well as in terms of the full available sample.

It is to be noted that the replication hypotheses, H_{1A} , H_{2A} and H_{3A} , along with the Summary Syndrome hypothesis, H_{5A} Risk Profile Stability hypotheses, H_{6A} and H_{7A} , are assessed by means of *within-subject* tests. This means that every observation in one of the relevant data sets is paired with a matched observation in the other set, provided by the same participant. The primary requirement for inclusion in the sample is that there must be two observations available from each individual.

By contrast, the Cash Flow hypothesis, H_{4A} , requires *inter-subject* comparisons, which means that these compared samples contain different subjects. The remaining two hypotheses, H_{8A} and H_{9A} , which relate to the Value Function extension, require only one sample, which is reduced to a binomial distribution. The requirements of each individual hypothesis are now considered in turn.

7.5.2 Sample Size and the Replication Hypotheses

The three replication hypotheses, H_{1A} , H_{2A} and H_{3A} , which were formally stated in Chapter Four, Subsection 4.4.1, require a sample of participants who face similar experimental conditions to those faced by Dusenbury's (1994) subjects. Given that the initial extension

TABLE 7.1
Sample Sizes with Respect to the Replication Hypotheses.

SAMPLE	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
TOTAL	132	59	39	20	14
LOW (\$2,100)	66	29	19	10	8
COMBINED HIGH	66	30	20	10	6
HIGH (\$3,500)	36	0	30	10	6
HIGH (\$4,200)	30	30	0	0	0

undertaken in this study involves varying the level of the pseudo-dollar cash float, the initial conceptual replication exercise should therefore ensure the cash float level is constant. To this end, either the *low cash float* level of \$2,100 (experimental pseudo-money) or the *high float* of not less than \$3,500 (pseudo) could be used to test these hypotheses.⁷ Of these, the *low float* grouping of 66 participants provides the less problematic replication sample on the ground that the float is invariant.

The high float grouping may be viewed as a feasible alternative on the ground that \$3,500 is a minimum *high float* (\$4,200 was the maximum). If a significant group difference is detected between the \$3,500 participants and the \$4,200 participants, then the replication hypotheses could, instead, be tested on two smaller samples. These smaller samples contain 30 subjects (in the case of the \$4,200 group) and 36 (for the \$3,500 group). The *low float* group of 66 will constitute the primary sample for these hypotheses; and the *high float* group — either as a single homogeneous group of 66 or as two heterogeneous groups of 30 and 36 — will be used as a secondary sample or samples. The size data is summarized in Table 7.1.

⁷ These cash float levels were explained in Chapter Five, Subsection 5.2.4.2.

7.5.3 Sample Size and the Extension Hypotheses

7.5.3.1 Cash Flow Extension

The Cash Flow Extension,⁸ discussed in Chapter Four, Subsection 4.5.1, requires a direct comparison of *low cash flow* and *high cash flow* samples. This extension became possible through doubling the numbers recruited for the experiment and by pairing participants so that there would be equal numbers undergoing the experiment with the two basic cash flow levels. The *low cash flow* grouping of 66 participants will be compared with the *high cash flow* grouping of 66 participants for the purposes of H_{4A} . It does not matter that the *high flow* grouping consists of two levels of high cash flow because what is necessary to falsify the implied null hypothesis in this instance is simply showing that the *high flow* group as a whole behaves in a less risk-seeking manner than the *low flow* group as a whole. The salient issue is the existence and direction of any difference rather than the difference's absolute magnitude. Nevertheless, because the data is available, the hypothesis will be tested separately on the \$4200 (pseudo) data set (sample size of 30) and the \$3,500 (pseudo) data set (sample size of 36) so that any differences may be noted. The summarized size data for this extension may be found above in Table 7.1.

7.5.3.2 Summary Syndrome Extension

It was not known in advance of the experimental sessions how large the available sample would be for the testing of H_{5A} (introduced in Chapter Four, Subsection 4.5.2). This was because the subjects needed to be partitioned in terms of answers given to two of the additional ratings questions (explained in Subsection 4.5.2 and also mentioned in Chapter Six, Section 6.7) relating to use of data provided in the story-lines in Tax Cases X and Y. Recollect that the answers to both were structured in the following format:

- Not at all — I ignored it.
- Only in a slight fashion.
- It was significant in my thinking.
- It was quite important.
- Of major importance — it totally shaped my thinking.

⁸ Note that the terms *flow* and *float* are both used in describing this extension. These terms are interchangeable.

TABLE 7.2					
Summary Syndrome Extension Sample Sizes for Refund Tax Case Y.					
FULL SAMPLE (N = 129)					
GROUP	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
A	32	17	9	4	2
A+B	59	30	18	5	6
C+D+E	70	27	21	14	8
COMBINED HIGH CASH FLOW SUBSAMPLE ONLY (N = 63)					
GROUP	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
A	16	10	3	3	0
A+B	32	17	10	4	1
C+D+E	31	11	10	5	5
LOW CASH FLOW SUBSAMPLE ONLY (N = 66)					
GROUP	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
A	16	7	6	1	2
A+B	27	13	8	1	5
C+D+E	39	16	11	9	3

If boxes are given alphabetical labels ranging from **A** (I ignored it) to **E** (it totally shaped my thinking), the participants who ticked boxes **C**, **D**, and **E** may be used as the sample of subjects who made decisions in accordance with the training provided, and in accordance with the experimental intention. By contrast, participants who ticked box **A** constitute a sample of subjects who definitely did not. This second sample is enlargeable by the inclusion of participants who ticked box **B**. Both **A** alone and **A+B** will be compared with **C+D+E** in the testing of H_{5A} .⁹

On the assumption that the partitioning of the full sample of 129 available taxpayers¹⁰ into *high cash flow* and *low cash flow* subsamples might have an influence on the outcome of the testing, it was envisaged that three different samples would be used to test the hypothesis with respect to each of the two tax cases.¹¹ These are shown in Table 7.2.

⁹ This argument was developed in Chapter Four, Subsection 4.5.2.1.

¹⁰ Three people, who failed to answer the additional rating question on Tax Case Y, were removed from the full sample (reducing it to 129) and from the high cash float subsample (reducing it to 63).

¹¹ When no significant shift in levels of risk between *high pay* Tax Case X and *refund* Tax Case Y was found in terms of the combined high cash float subsample (or its constituent subsamples), little point was seen in testing H_{5A} in terms of this subsample. The reason for this decision lay in the purpose of the Summary Syndrome extension, which was to investigate

TABLE 7.3					
Summary Syndrome Extension Sample Sizes for High Pay Tax Case X.					
FULL SAMPLE (N = 131)					
GROUP	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
A	32	14	10	4	4
A+B	56	28	17	5	6
C+D+E	75	31	22	14	8
COMBINED HIGH CASH FLOW SUBSAMPLE (N = 65)					
GROUP	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
A	15	10	3	2	0
A+B	27	17	7	2	1
C+D+E	38	13	13	7	5
LOW CASH FLOW SUBSAMPLE (N = 66)					
GROUP	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
A	17	4	7	2	4
A+B	29	11	10	3	5
C+D+E	37	18	9	7	3

With respect to testing H_{5A} in terms of the *high pay* tax case, Case X, the sample sizes set out in Table 7.3 applied.¹²

7.5.3.3 Risk Profile Stability Extension

Investigation of the matched data sets required in the assessment of H_{6A} (introduced and explained in Chapter Four, Subsection 4.5.3) is restricted to the experiment's first 83 participants, who belonged in the Church and the Kindergarten subsamples only. The size data for these subsamples is as shown in Table 7.4:

TABLE 7.4			
Sample Size for Risk Profile Stability Extension.			
SAMPLE	TOTAL	CHURCH	KINDERGARTEN
FULL	83	59	24
LOW FLOW	33	29	4
HIGH FLOW	50	30	20

whether a significant shift in risk levels could be shown to exist in the absence of awareness or cognizance of the information about tax withholdings provided in the scenario story-lines.

¹² One person, who failed to answer this additional rating question, was removed from the full sample (reducing it to 131), and the same person was removed from the high cash flow subsample (reducing it to 65).

TABLE 7.5					
Sample Size for Value Function Extension.					
SAMPLE	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
FULL	131	59	39	19	14
LOW FLOW	66	29	19	10	8
HIGH FLOW	65	30	20	9	6

7.5.3.4 Value Function Extension

Hypotheses H_{7A} and H_{8A} , as explained in Chapter Four, Subsection 4.5.4, involve analyses of binomial distributions. Since there is no comparison of one data set with another, these hypotheses can be tested on a sample of 131 individuals.¹³ While it is unlikely that cash flow considerations would have an impact on the smoothness of the value function, H_{7A} and H_{8A} will also be tested on the two cash flow subsamples. The size data for the extension is summarized in Table 7.5.

In the next chapter, Chapter Eight, the results of tests of the three replication hypotheses are presented; while Chapter Nine contains the results of investigations concerning the nature of the sample, its various subsamples, and the questionnaire. Chapter Ten completes the tabling of findings by providing results of tests made of the six hypotheses formulated with respect to the study's extensions.

¹³ This represents the entire sample of participants minus one school subsample recruit who did not provide the required information. This person had a high cash float; therefore the high cash float subsample is reduced by one to 65 with respect to examination of this extension.

8. CONCEPTUAL REPLICATION RESULTS

8.1 INTRODUCTION

This chapter discusses the results obtained from the testing of hypotheses, H_{1A} , H_{2A} and H_{3A} ,¹ which were the three hypotheses investigated in a United States context by Dusenbury (1994). With this restriction of focus, the information of interest is limited to the risk levels recorded for the initial five scenarios only.

The salient feature of H_{1A} , H_{2A} and H_{3A} , is that each hypothesis entails a comparison of responses made to a pair of decision problems offering choices amongst risky prospects. The most immediately accessible parameter for making this comparison is the mean of the risk-related choices recorded in response to the two problems. Two computations of this mean were possible. Since every decision problem contained five prospects (labelled Options A to E) laid out in order of increasing riskiness, it was possible to convert them into a five-point ordinal scale in which A equated with 1 and E equated with 5. This enabled numerical means to be calculated such that a mean of 2.5 could be interpreted as representing a higher average level of risk willingness in the minds of the participants than a mean of 2.3. This approach is used in most instances in which means are computed in Chapters Eight, Nine and Ten.

The second approach to computing means involved summing the chosen risk levels as decimal fractions and dividing by the number of decision makers. This produced a mean level of preferred risk, which could be read as a defined probability or converted into a percentage level of chance. This approach is used in Tables 8.1, 8.2 and 8.3 in Section 8.2.

On the other hand, testing for statistical significance involved the use of the variance parameter. The vehicle for most of the statistical analysis undertaken in the study was a SAS

¹ These hypotheses were explained in Chapter Four, Subsection 4.2.1 in some detail with respect to Dusenbury's use of them; and were repeated in summary form in Subsection 4.4.1 of the same chapter.

repeated measures General Linear Model (GLM) analysis of variance procedure.² This is explained in Section 8.3 and in Appendices C and D. The *F*-statistics furnished by this procedure indicate whether or not there are significant differences in the risk levels chosen in response to the contrasted decision problems. Since the procedure does not report means, these have been furnished by an independent procedure run on the same data set. The data used by the GLM analysis of variance procedure was the numerical data ranging from 1 to 5 on an ordinal scale, as described in the previous paragraph.

The chapter is laid out as follows. Section 8.2 deals with evidence discernible in a cursory inspection of the means and standard deviations obtained from the collated New Zealand data. Section 8.3 then discusses computational aspects of the in-depth statistical analysis; and Section 8.4 contains the results of the validation process. The next section, Section 8.5 contains results pertaining to H_{1A} ; and following on from this, Section 8.6 looks at the findings associated with H_{2A} . Finally, the results related to H_{3A} are laid out in Section 8.7.

8.2 INITIAL INSPECTION OF SUMMARY INFORMATION

An initial superficial analysis of the risk levels adopted by participants in response to the five replication scenarios yielded the summary information reported in Table 8.1. These results were derived from the full sample of 132 subjects irrespective of the impact attributable to the cash float variable. The table also contains Dusenbury's (1994) results for the purpose of comparison.

The means and standard deviations, obtained on samples in different countries and several years apart, are remarkably similar. They differ at most by two percentage points. In Panel A, the New Zealand subjects showed a mean propensity for risk roughly one percent greater than Dusenbury's subjects in the *refund* and high-pay tax cases. This is reflected in Panel B by disclosure of a slightly lower level of the uncertain tax item by New Zealanders in these two

² SAS was used in the computation of all of the validation statistics used in this study.

TABLE 8.1				
Subjects' Choices: Descriptive Statistics.				
PANEL A: The Mean Risk Level Selected				
Sample size	CURRENT STUDY N=132		DUSENBURY (1994) N=65	
Cases	Mean	St. Deviation	Mean	St. Deviation
<i>Refund Case (Tax Case Y)</i>	0.16	0.12	0.15	0.12
<i>High pay Case (Tax Case X)</i>	0.21	0.12	0.20	0.13
<i>Low pay Case (Tax Case Z)</i>	0.15	0.13	0.15	0.13
Medical Insurance Case	0.17	0.14	0.15	0.15
Gamble Case	0.26	0.14	0.27	0.14
PANEL B: Portion of the Uncertain Amount Not Excluded as Taxable Income				
Sample size	CURRENT STUDY N=132		DUSENBURY (1994) N=65	
Cases	Mean	St. Deviation	Mean	St. Deviation
<i>Refund Case (Tax Case Y)</i>	0.68	0.27	0.69	0.28
<i>High pay Case (Tax Case X)</i>	0.55	0.31	0.57	0.33
<i>Low pay Case (Tax Case Z)</i>	0.70	0.28	0.70	0.29
PANEL C: Preliminary Support for the Replication Hypotheses				
Sample Size	CURRENT STUDY N=132		DUSENBURY (1994) N=65	
Hypothesis (And Constituent Cases)	Mean Risk Shift	Support for Rejection of H_0	Mean Risk Shift	Support for Rejection of H_0
H_{1A} (Tax Cases Y and X)	+ 5%	Yes	+ 5%	Yes
H_{2A} (Tax Case X and Medical Case)	+ 4%	Yes	+ 5%	Yes
H_{3A} (Tax Case Z and Gamble Case)	+ 11%	Yes	+ 12%	Yes

cases. On the other hand, the North American taxpayers exhibited a two percent lower mean riskiness in the medical insurance case and, on average, were one percent more risky with respect to the gamble (Panel A). The means and standard deviations reported with respect to the low-pay tax case turned out to be virtually identical. These details lend support for rejection of the null form of all three of the Dusenbury replication hypotheses — H_{1A} , H_{2A} and H_{3A} .

This apparent support for the rejection of the null forms of all three replication hypotheses is summarised in terms of the risk preference shifts derived from Panel A and recorded in Panel C of Table 8.1. A mean risk preference shift is defined here as the difference between the means of the risk levels chosen in the pair of decision problems examined when testing each hypothesis. It must be noted, however, that this summary information is preliminary evidence

TABLE 8.2				
Subjects' Choices Partitioned into Low and High Cash Float Subsamples.				
PANEL A: The Mean Risk Level Selected				
Sample Size	LOW: CASH = \$2,100 N = 66		HIGH: CASH ≥ \$3,500 N = 66	
Cases	Mean	St. Deviation	Mean	St. Deviation
<i>Refund</i> Case (Tax Case Y)	0.15	0.11	0.17	0.12
<i>High pay</i> Case (Tax Case X)	0.23	0.11	0.19	0.13
<i>Low pay</i> Case (Tax Case Z)	0.13	0.14	0.17	0.13
Medical Insurance Case	0.17	0.12	0.16	0.13
Gamble Case	0.26	0.14	0.25	0.14
PANEL B: Portion of the Uncertain Amount Not Excluded as Taxable Income				
Sample Size	LOW: CASH = \$2,100 N = 66		HIGH: CASH ≥ \$3,500 N = 66	
Cases	Mean	St. Deviation	Mean	St. Deviation
<i>Refund</i> Case (Tax Case Y)	0.70	0.25	0.65	0.29
<i>High pay</i> Case (Tax Case X)	0.51	0.30	0.65	0.30
<i>Low pay</i> Case (Tax Case Z)	0.74	0.26	0.70	0.29
PANEL C: Preliminary Support for the Replication Hypotheses				
Sample Size	LOW: CASH = \$2,100 N = 66		HIGH: CASH ≥ \$3,500 N = 66	
Hypothesis (And Constituent Cases)	Mean Risk Shift	Support for Rejection of H₀	Mean Risk Shift	Support for Rejection of H₀
H _{1A} (Tax Cases Y and X)	+ 8%	Yes	+ 2%	Yes
H _{2A} (Tax Case X and Medical Case)	+ 6%	Yes	+ 3%	Yes
H _{3A} (Tax Case Z and Gamble Case)	+ 13%	Yes	+ 8%	Yes

only, which, as yet, has not been subjected to tests for statistical significance. (These tests are reported in Sections 8.5, 8.6 and 8.7.)

However, this information does not take into account the impact of the cash float variable. Table 8.2 contains the means and standard deviations obtained when the New Zealand sample was partitioned by cash float level into a high cash float group and a low cash float group.

Table 8.2, like Table 8.1 before it, provides apparent support for rejection of the null hypothesis with respect to H_{1A}, H_{2A} and H_{3A}. This is shown in terms of the mean risk preferences for each decision problem disclosed in Panel A, and the risk shift apparent in each pair of hypothesis-related decision problems reported in Panel C. Of further interest in Panel A of Table 8.2 is an apparent amplification of the change from risk aversion to risk willingness on the part of the low cash float participants between the *refund* and *high pay* tax cases. The low cash float

group have a two percent lower mean risk level for the *refund* case (15 percent) than does the high float subsample (17 percent), and a four percent higher degree of risk willingness with respect to the *high pay* tax case (23 percent versus 19 percent). This preliminary evidence would seem to support rejection of the null form of the cash flow extension hypothesis, H_{4A} .

A similar amplification is discernible for the switch in preferences between the *low pay* tax case and the gamble. However, in terms of the contrast between the *high pay* tax case and the medical insurance case laid out in H_{2A} , the low cash float subsample appear to be uniformly slightly more risk willing than their high float counterparts.

However, the partitioning of the sample into a high cash float subsample and a low cash float subsample still fails to provide a completely accurate overview with respect to cash float influences. But when the high cash float group is partitioned into a \$3,500 stream and a \$4,200 stream, as reported in Table 8.3, the clarity of the cash float variable's influence on participants' risk profiles is muddied.

The \$4,200 subsample appears to be more risk seeking with respect to both the *refund* and *high pay* tax scenarios than the \$3,500 subsample, which apparently lends no support for H_{4A} . Furthermore, while H_{1A} appears to be supported by the pattern of the \$4,200 group means, this support disappears with respect to the mean risk profiles of the \$3,500 group. The participants in this subsample adopted a mean risk level of 17 percent for both the *refund* and *high pay* tax cases.³

With respect to H_{2A} , both groups exhibit greater risk aversion in the medical case than in the *high pay* tax case, thereby lending apparent support for rejection of H_{2A} 's null form; and again, the \$4,200 subsample exhibits uniformly greater risk willingness in the two cases. However, in terms of the *repeated measures* analysis of variance tests, the results of which are explained in

³ When a *repeated measures* analysis of variance (ANOVA) procedure was used in the analysis of the high cash float subsample as a single grouping and in terms of the \$4,200 and \$3,500 cash float subsamples, it furnished results which did not support rejection of the null hypothesis. This information is contained in Section 8.5 later in the chapter.

TABLE 8.3

Subjects' Choices: High Cash Float Subsample Partitioned into \$3,500 and \$4,200 Subsamples.

PANEL A: The Mean Risk Level Selected				
Sample Size	CASH = \$3,500 N = 36		CASH = \$4,200 N = 30	
Cases	Mean	St. Deviation	Mean	St. Deviation
<i>Refund</i> Case (Tax Case Y)	0.17	0.12	0.18	0.12
<i>High pay</i> Case (Tax Case X)	0.17	0.12	0.21	0.13
<i>Low pay</i> Case (Tax Case Z)	0.18	0.11	0.15	0.15
Medical Insurance Case	0.15	0.13	0.18	0.13
Gamble Case	0.27	0.11	0.23	0.17

PANEL B: Portion of the Uncertain Amount Not Excluded as Taxable Income				
Sample Size	CASH = \$3,500 N = 36		CASH = \$4,200 N = 30	
Cases	Mean	St. Deviation	Mean	St. Deviation
<i>Refund</i> Case (Tax Case Y)	0.65	0.29	0.64	0.28
<i>High pay</i> Case (Tax Case X)	0.64	0.30	0.54	0.33
<i>Low pay</i> Case (Tax Case Z)	0.63	0.26	0.68	0.34

PANEL C: Preliminary Support for the Replication Hypotheses				
Sample Size	CASH = \$3,500 N = 36		CASH = \$4,200 N = 30	
Hypothesis (And Constituent Cases)	Mean Risk Shift	Support for Rejection of H_0	Mean Risk Shift	Support for Rejection of H_0
H_{1A} (Tax Cases Y and X)	0%	No	+ 3%	Yes
H_{2A} (Tax Case X and Medical Case)	+ 2%	Yes	+ 3%	Yes
H_{3A} (Tax Case Z and Gamble Case)	+ 12%	Yes	+ 5%	Yes

Section 8.6 of this chapter, rejection of the null hypothesis could not be supported with respect to the full sample or any of the cash float subsamples.

With respect to H_{3A} , both high cash float streams produced case-specific means which apparently support rejection of the null hypothesis; but, paradoxically, in these two cases, the \$4,200 stream's mean risk levels were lower than those of the \$3,500 subsample. When the \$4,200 subsample was subjected to a *repeated measures* analysis of variance test, the apparent grounds for rejection of the null hypothesis were not supported. This is explained in Section 8.7, later in the chapter.

At this point we move on from cursory inspection of sample means to more rigorous statistical analysis.

8.3 TESTING FOR VALIDATION

8.3.1 Introduction

This section provides a description of the techniques used in the testing of the data. Considerable time was put into searching for group, session and case-order-related influences on the response data collected with respect to the five scenarios replicating Dusenbury (1994). Initially, since much of the data was nominal or ordinal in nature, a categorical data modelling procedure was employed. This is described in Subsection 8.3.2. However, since Dusenbury used a *repeated measures* ANOVA procedure to process his results, it was necessary to emulate that procedure.

One of the salient differences between Dusenbury's experimental design and the design in this study lay in the number of case orderings. Dusenbury restricted his case orderings to a total of four, which effectively was also the number of subsamples in his analysis step; and so he had a balanced design suitable for a *repeated measures* ANOVA. The current study, in contrast, employed a random spread of case orderings for the replication scenarios (with the exception that *refund* Case Y, as in Dusenbury's experiment, always came first). Because only 19 of the 24 possible permutations were actually used, and also because this number was quite incongruent with the number of subsamples (four as determined by recruitment) and the number of sessions, the current study had an unbalanced design. The downside of this was that it was not possible to use Dusenbury's exact computational procedure in the analysis of the data.

But there was an up-side too. The experimental conditions more closely mimicked the random ordering of actual decision problems in daily life. Furthermore, powerful computational tools are available for the handling analysis of variance for unbalanced designs; and in particular, this study made use of two versions of a General Linear Model (GLM) analysis of variance procedure, one of which, like Dusenbury's (1994) ANOVA tool, involved a *repeated measures* function. Details of these are laid out in Subsection 8.3.3.

8.3.2 Categorical Data Modelling Procedure

In depth investigation of the data began with tests for influence emanating from the participants' attendance at a particular session, or membership of their particular group, and the order in which they were required to work through the scenarios. Given that session, group and case-order were all categorical variables (by nature, nominal), the initial testing utilised the categorical data modelling procedure, CATMOD. CATMOD is a tool provided by the software package, SAS, to test combinations of these variables on a dependent variable. The dependent variable was the separate set of responses to each of the decision problems; hence, what was actually a logistic regression procedure was run separately for each case. The sets of response observations were segregated for this purpose by the size of the participants' cash float. No significant session, group or case-order effects were found with respect to any of the five replication cases.⁴ Because of the piecemeal nature of this testing, the results did not present themselves in a form compact enough to be readily presented here. Instead, the summary results obtained via SAS's General Linear Model (GLM) procedure for analyses of variance involving classification variables and unbalanced data are reported. Both a *repeated measures* GLM procedure and a GLM procedure without repeated measures (for convenience here termed the *basic model*) were used.

8.3.3 The General Linear Model Procedure

The *repeated measures* General Linear Model procedure for analysis of variance is similar to the procedure used by Dusenbury (1994). Dusenbury reported the *F*-statistics from *repeated measures* ANOVAs to test for differences in the means of his responses. He justified this on the ground that the residuals obtained from the procedure, when tested for normality in terms of the Shapiro-Wilks'⁵ *W* statistic and the Hartley test for homogeneity of variance, were shown to be normally distributed and homoscedastic.⁶

⁴ See Dusenbury, R., (1994) "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-filing Options", p. 10, Footnote 9. Dusenbury notes in this footnote that he too investigated his data in terms of ordinal measures and non-parametric tests.

⁵ Dusenbury (1994), op. cit. n. 4, p. 10, Footnote 10. Dusenbury refers to the Shapiro-Wilkes *W* statistic. The SAS Procedures Guide for Personal Computers, on the other hand, refers to the Shapiro-Wilks' *W* statistic, which, although spelt differently,

In the current study, the residuals obtained from the *basic model* were tested, by means of SAS's UNIVARIATE procedure, for normality in terms of the Shapiro-Wilks' W statistic. These residuals were then plotted against each independent variable used in the GLM procedure for a visual diagnosis of any potential heteroscedasticity. The residuals were indeed normally distributed, and no noteworthy departures from homoscedasticity were found.

The two forms of GLM procedure were usable on the ground that, while information pertaining to group membership, session attendance and case ordering was coded as nominal data only, responses to the decision problems were viewable as data measured on an interval scale since the possible outcomes associated with consecutive options posed as alternatives within a decision problem changed in steps of \$100. This enabled the study to avoid sole reliance on the CATMOD procedure.

While the *basic model* GLM procedure produced an ANOVA table in which the covariances of sundry categorical variables with respect to one dependent variable were reported, the *repeated measures* procedure produced four sets of useful results.

In the first instance, the *repeated measures* GLM procedure provided a univariate ANOVA table for each dependent variable similar to the one produced in the *basic model*.⁷ The second item of output was a set of MANOVA tables furnishing statistical output with respect to every possible main and interaction effect detectable in the model with respect to differences between the dependent variables or produced by the categorical variables. The third item was a univariate ANOVA table of *between-subjects* main and interaction effects generated by the categorical variables, in which the ANOVA's dependent measure was a procedure-generated measure of the difference between the dependent variables. The fourth item was a univariate *within-subject* ANOVA table in which the main effect was the level of intrinsic difference

clearly fills the same function. Since the data examined in the current study was processed by means of SAS, the SAS nomenclature has been adopted.

⁶ Dusenbury (1994), op. cit. n. 4, p. 10, Footnote 10.

⁷ However, differences in data grooming carried across into differences in output between the two types of GLM analysis of variance procedures. The *basic model* incorporated a categorical variable, CASENO, not used in the *repeated measures* procedure. The reason for this is presented in Appendix D.

between the dependent variables, and the interaction effects were the effects of the categorical variable multiplied with the main effect.

The final two items of output (the univariate *between-subjects* and *within-subjects* ANOVA tables) were the most useful; and have been cited for every *repeated measures* GLM procedure used in the study. Because these tables contained information which essentially paralleled the information provided in the MANOVA output, but in a more presentable format, the MANOVA results for the many *repeated measures* GLM procedures have generally not been reported.

An explanation of how the two GLM procedures function is provided in Appendix C; and a description of the data grooming required by the procedures is contained in Appendix D. Appendix D also contains a description of the variables representing the responses made to the six decision problems investigated in the experiment. (For brevity's sake, these were labelled C1 to C6.)

8.4 TESTS FOR MAIN AND INTERACTION EFFECTS ON THE MODEL

At present the validation procedure has been performed strictly in terms of the five decision-problem scenarios posed in the replication of the United States experiment. The *repeated measures* GLM procedure was applied to both the full sample and to the various subsamples as determined by levels of the cash float variable. The significance level (the maximum probability of a Type I error allowable if the null form of a hypothesis is to be rejected⁸), which was used in the study was the generally recognised five percent benchmark. This was also used by Dusenbury (1994).

The reason for running validation tests on more than just the full sample alone was to check for extraneous influences emanating from variations in the level of the cash float. The test results

⁸ A Type I error involves rejecting the null hypothesis when the null hypothesis is actually correct.

associated with the low cash float subsample are reported here because it was the largest pure sample available, since the high float grouping was further divisible into \$3,500 and \$4,200 subsamples. The sheer volume of output provides an excellent reason for not reporting validation statistics for subsamples representing every permutation of cash float level. The high and low cash float participants were fairly closely paired in every subsample by recruitment origin;⁹ and the corollary of this fact is, if the low cash float subsample can be shown to be a valid sample, then the same must be true of the high cash float grouping.

The *repeated measures* GLM statistical output provides strong validation of the experimental conditions and sample selection arrangements. Panel A of Table 8.4 contains the *F*-statistics for the *between subject* main effects brought about by the influence of group membership or order of scenario presentation and their mutual interaction with respect to the full sample. The insignificance of the *F*-statistics in the table indicates that no significant *between-subjects* main or interaction effects were found.

The *within-subject* main and interaction effects are reported in Panel B of Table 8.4. In this table, the *repeated measures* GLM procedure has created the variable, CASE; and the *F*-statistic associated with it is a measure of difference between the mean of variable C1 observations (*refund* Tax Case Y) and the means of observations for the other four replication scenarios. CASE, alone, is statistically significant ($F = 9.04$, $\text{Pr} > F = 0.0001$), indicating that the responses to the four scenarios were significantly different from the Tax Case Y responses. No *within-subject* interaction effects come close to being significant at the five percent benchmark. These findings are corroborated in terms of the Greenhouse-Geisser and Huynh-Feldt Epsilon statistics cited in Panel C. The equivalent MANOVA result is printed in full in Appendix E, Section E.1.

When the *repeated measures* procedure was applied to the low cash subsample, the validation of CASE-related differences in means was weaker than those provided by CASENO in the

⁹ The actual disposition of floats was as follows: 30 high cash float and 29 low cash float participants in the Church subsample; 20 high and 19 low cash float participants in the Kindergarten subsample; and 6 high and 8 low cash float participants in the Lincoln University sample. With respect to the School group, exactly equal numbers received each float type. These minor variations were caused by the rejection of two people from the sample, which was explained in Chapter Seven, Section 7.4.

TABLE 8.4

**Tests for Main and Interaction Effects on the Five Replication Scenarios.
Full Sample (N = 132)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	10.2395	3.4132	1.36	0.2600
CASEORD	18	29.6157	1.6453	0.66	0.8421
GROUP*CASEORD	29	52.6364	1.8150	0.72	0.8340
Error	81	202.8093	2.503818		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	4	49.6754	12.4188	9.04	0.0001
CASE*GROUP	12	15.9283	1.3274	0.97	0.4815
CASE*CASEORD	72	97.2423	1.3506	0.98	0.5221
CASE*GROUP*CASEORD	116	155.4639	1.3402	0.98	0.5556
Error(CASE)	324	445.3014	1.3744		

PANEL C

Greenhouse-Geisser Epsilon and the Huynh-Feldt Epsilon

Source	G - G	H - F
CASE	0.0001	0.0001
CASE*GROUP	0.4775	0.4815
CASE*CASEORD	0.5196	0.5221
CASE*GROUP*CASEORD	0.5522	0.5556
Whole model	0.9113	1.5486

basic model procedure reported in Appendix E, Section E.2. The *between-subjects* effects are reported in Panel A of Table 8.5.

It is clear in Panel A of Table 8.5 there are no *between-subjects* main or interaction effects; and in Panel B, CASE is statistically significant at about the $Pr > F = 0.0001$ level of error. However, two weak *within-subject* interaction effects are reported in this panel (CASE*GROUP)¹⁰ and (CASE*CASEORD); and one effect is statistically significant within

¹⁰ GROUP is a categorical variable denoting which of the four organisations provided the participant for the experiment. CASEORD is a variable which denotes the order in which the participant received (and worked through) the five replication

TABLE 8.5

**Tests for Main and Interaction Effects on the Five Replication Scenarios.
Low Cash Float Subsample (N = 66)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	4.5919	1.5306	0.52	0.6723
CASEORD	15	33.1567	2.2104	0.75	0.7162
GROUP*CASEORD	19	39.2538	2.0660	0.70	0.7873
Error	28	82.5000	2.9464		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	4	35.7636	8.9409	9.02	0.0001
CASE*GROUP	12	19.9697	1.6641	1.68	0.0807
CASE*CASEORD	60	84.3710	1.4062	1.42	0.0561
CASE*GROUP*CASEORD	76	106.9374	1.4071	1.42	0.0454
Source: Error(CASE)	112	111.0000	0.9911		

PANEL C

Greenhouse-Geisser Epsilon and the Huynh-Feldt Epsilon

Source	G - G	H - F
CASE	0.0001	0.0001
CASE*GROUP	0.0878	0.0807
CASE*CASEORD	0.0632	0.0561
CASE* GROUP*CASEORD	0.0519	0.0454
Whole Model	0.9245	2.4901

the five percent benchmark (CASE*GROUP*CASEORD). When this result was investigated in terms of main and interaction effects on each of the five scenario response variables separately, no significant between-subject effects were found.¹¹ In the course of analysing the study's results it was found that the low cash float subsample was the most responsive to the intrinsic differences among the scenarios interpreted as decision frame differences. The noise effects detected in Table 8.5 may be connected with the low cash float subsample's failure to

scenarios. This is also explained, along with other categorical variables used in the significance testing procedure, in Appendix D, Sections D.1 and D.2.

¹¹ A table containing the repeated measures GLM procedure univariate results for each separate response (dependent) variable may be found in Appendix E, Section E.3.

make a statistically significant risk shift between *high pay* Tax Case X and the medical insurance case. This is explored in Section 8.6.

A stronger validation of both the full sample and the low cash float subsample was provided by the *basic model* GLM analysis of variance procedure. Both configurations of the sample returned results in which the intrinsic difference among the five replication cases was stated to be significant at the level of probable error, $\text{Pr} > F = 0.0001$; and in both instances no significant main or interaction effects were detected emanating from the categorical variables, GROUP and CASEORD. The full results obtained from these two *basic model* procedures may be found in Appendix E, Section E.2.

8.5 HYPOTHESIS H_{1A}

8.5.1 Introduction

This hypothesis, stated in its alternative form as it was by Dusenbury (1994), is repeated here for convenience¹²:

H_{1A} : Participants will choose riskier options in the *high pay* case than in the *refund* case.

Because it was possible that the level of cash float given to participants might have a significant influence on the risk profiles under investigation in this hypothesis, the hypothesis was tested on every cash float subsample in addition to the full sample of 132 participants. Indeed, a significant distinction was found between subjects who received the low float and those who received the high one. It was not possible to reject the null form of the hypothesis with respect to all subsamples.

As in the more general validation procedure, GROUP and CASEORD (and in the \$4,200 cash float subsample instance, SESSION) were included in the analysis so that any concomitant

¹² Dusenbury (1994), op. cit. n. 4, p. 6, and the current study, Chapter 4, Subsection 4.4.1.

TABLE 8.6					
Refund Tax Case Y and High Pay Tax Case X.					
Full Sample (N = 132)					
PANEL A					
General Linear Models Procedure Repeated Measures Analysis of Variance:					
Univariate Tests of Hypotheses for Within Subject Effects					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	7.1045	7.1045	6.86	0.0105
CASE*GROUP	3	1.2286	0.4095	0.40	0.756
CASE*CASEORD	18	16.6940	0.9274	0.90	0.5856
CASE*GROUP*CASEORD	29	17.7431	0.6118	0.59	0.9440
Error (CASE)	81	83.9071	1.0359		
PANEL B					
Response Variable	N	Mean	Std Dev		
<i>Refund</i> Tax Case Y (C1)	132	2.2955	1.0753		
<i>High pay</i> Tax Case X (C2)	132	2.7879	1.2358		

interaction effects could be detected. The change with respect to the \$4,200 cash float subsample was made because this group was drawn from the church group alone, and to test for GROUP-related effects would have been redundant.

The H_{1A} test results are delivered in subsections partitioned by subsample. The first of these subsections contains those pertaining to the full sample.

8.5.2 The Full Sample and H_{1A}

Panel B of Table 8.6 contains the summary information relevant to the full sample. The higher mean recorded for the *high pay* case is shown to be statistically significant by the levels of error shown for the *F*-statistics in Panel A, providing support for rejecting the null form of the hypothesis. Only the variable CASE is significant, and is so sufficiently close to the one percent level of error. This implies there has been a framing effect-induced shift in risk preferences between the *refund* and *high pay* tax cases in question.

8.5.3 The Low Cash Float Subsample and H_{1A}

When the sample was split into low and high cash float groups (the high cash flow group containing the \$3,500 and \$4,200 subsamples undifferentiated), two divergent trends became

TABLE 8.7

**Refund Tax Case Y and High Pay Tax Case X.
Low Cash Float Subsample (N = 66)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	11.4107	11.4107	14.18	0.0008
CASE*GROUP	3	1.0875	0.3625	0.45	0.7190
CASE*CASEORD	15	14.2383	0.9492	1.18	0.3412
CASE*GROUP*CASEORD	19	13.5601	0.7137	0.89	0.6004
Source: Error(CASE)	28	22.5333	0.8048		

PANEL B

Response Variable	N	Mean	Std Dev
<i>Refund</i> Tax Case Y (C1)	66	2.1818	0.9909
<i>High pay</i> Tax Case X (C2)	66	2.9545	1.2081

evident. The low cash float subsample proved itself to be similar to the full sample in that it supported a rejection of H_{1A} 's null form; but the evidence produced in the processing of the high cash float subsample displayed the opposite. The low cash float subsample is evaluated in Table 8.7, and the undifferentiated high cash float subsample is discussed in the next subsection.

Panel A of Table 8.7 provides evidence of a very clear difference in the means of risk levels chosen in response to the two scenarios attributable to framing effects induced by the scenarios themselves. The F -statistic for CASE is significant with a probability of error, $Pr > F = 0.0008$; and there are no significant interaction effects. The null hypothesis is easily rejected with respect to this subsample.

8.5.4 The Various High Cash Float Subsamples and H_{1A}

In the first instance, the combined high cash float subsample containing both the \$3,500 and \$4,200 cash float recipients was examined. Although the mean risk level chosen by these participants in response to *high pay* Tax Case X (recorded in Panel B of Table 8.8) was slightly higher than their mean riskiness of choices with respect to *refund* Tax Case Y, the difference was not significant. The F -statistic associated with CASE is not only the smallest in the set,

TABLE 8.8

**Refund Tax Case Y and High Pay Tax Case X.
Combined High Cash Float Subsample (N = 66)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.2532	0.2532	0.19	0.6622
CASE*GROUP	3	1.6945	0.5648	0.43	0.7298
CASE*CASEORD	17	14.1104	0.8300	0.64	0.8317
CASE*GROUP*CASEORD	17	10.4636	0.6155	0.47	0.9449
Error(CASE)	28	36.3750	1.2991		

PANEL B

Response Variable	N	Mean	Std Dev
<i>Refund</i> Tax Case Y (C1)	66	2.4091	1.1499
<i>High pay</i> Tax Case X (C2)	66	2.6212	1.2498

but has a high probability of error, indicating little evidence of the existence of any scenario-related framing effect. In terms of this subsample, the null form of the hypothesis cannot be rejected.

When the high cash float subsample was split into its two components — the \$3,500 subsample and the \$4,200 cash float subsample, the negative findings for the high cash float group as a whole were reinforced. In terms of the \$3,500 cash float participants, the reason may be found in the close proximity of the means of the responses to the two scenarios, which are reported in Panel B of Table 8.9.

Table 8.9 shows quite emphatically that there is no significant difference between the means occurring in *high pay* Tax Case X and *refund* Tax Case Y with respect to the \$3,500 cash float subsample. Again in this instance, the null form of H_{1A} cannot be rejected.

A similar story unfolds with respect to the \$4,200 cash float participants isolated as a subsample. On the off-chance that the result might be distorted by the dropping of the variable, GROUP (since there was only one group for this subsample), the GLM procedures were run with the variable, SESSION included as an independent variable — since there were four discrete sessions in which participants were given a \$4,200 cash float.

TABLE 8.9

**Refund Tax Case Y and High Pay Tax Case X.
\$3,500 Cash Float Subsample (N = 36)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.0057	0.0057	0.00	0.9524
CASE*GROUP	2	1.3432	0.6716	0.43	0.6565
CASE*CASEORD	11	11.3911	1.0356	0.67	0.7435
CASE*GROUP*CASEORD	9	3.7930	0.4214	0.27	0.9713
Error(CASE)	13	20.0833	1.5449		

PANEL B

Response Variable	N	Mean	Std Dev
<i>Refund Tax Case Y (C1)</i>	36	2.3889	1.1778
<i>High pay Tax Case X (C2)</i>	36	2.4444	1.1819

Both with and without it, the result was uniformly conclusive; so only the results of the *repeated measures* procedures incorporating SESSION are reported. The CASE *F*-statistic reported in Table 8.10 makes it clear that H_{1A} 's null hypothesis cannot be rejected with respect to the \$4,200 cash float subsample.

8.5.5 Investigation Incorporating Compliance Attitude Variables

Because the failure to reject H_{1A} in terms of the three possible high cash float subsamples could possibly be attributed to prevailing attitudes to non-compliance, the full sample (upon which H_{1A} 's null form could be rejected) and the \$3,500 and \$4,200 cash float subsamples were investigated in terms of the participants' responses to the three questions in the questionnaire designed to elicit their attitudes towards non-compliant tax filing. These questions were¹³:

¹³ These questions were coded on a Likert scale. The first box was coded 1, the second, 2 ... and the fifth, 5. The use of Likert scales was mentioned in conjunction with the questions from the questionnaire in Chapter Six, Subsection 6.9.1; and these questions were discussed in Subsection 6.9.3. The full questionnaire is reprinted in Appendix B, Section B.4.

TABLE 8.10
Refund Tax Case Y and High Pay Tax Case X.
\$4,200 Cash Float Subsample (N = 30)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	2.1276	2.1275	1.61	0.2368
CASE*SESSION	3	0.5741	0.1914	0.14	0.9306
CASE*CASEORD	14	10.9246	0.7803	0.59	0.8191
CASE*SESSION*CASEORD	3	3.6747	1.2249	0.93	0.4673
Error (CASE)	9	11.9167	1.3241		

PANEL B

Response Variable	N	Mean	Std Dev
<i>Refund</i> Tax Case Y (C1)	30	2.4333	1.1351
<i>High pay</i> Tax Case X (C2)	30	2.8333	1.3153

Q6: Attitude to non-compliance #1:

Given the tax laws that are in place in New Zealand at the present time, would you consider it acceptable for a taxpayer earning over \$35,000 to under-report income that he/she would have the power to conceal?

- never
 seldom
 on average
 mostly
 always

Q7: Attitude to non-compliance #2:

Given the tax laws that are in place in New Zealand at the present time, would you consider it acceptable for a taxpayer earning less than (say) \$18,000 to under-report income that he/she would have the power to conceal?

- never
 seldom
 on average
 mostly
 always

Q8: Feeling about actual taxpaying:

When you have found yourself in the position of owing more tax to the IRD, which single box of the following would come closest to most accurately describing what you have experienced:

- I am very unhappy: This is a loss to me and an imposition
 This is something I feel moderately unhappy about; but it is necessary
 No strong feeling
 I'm moderately happy to do the right thing
 I'm happy to fulfil this requirement: it is my contribution to society

TABLE 8.11

**Responses to Compliance Questions with Respect to
Refund Tax Case Y and High Pay Tax Case X.
Full Sample (N = 132)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q6	4	4.2977	1.0744	0.62	0.6514
Q7	4	1.2193	0.3048	0.18	0.9509
Q8	4	4.6761	1.1690	0.67	0.6132
Error	119	207.2720	1.7418		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	4.8586	4.8586	5.25	0.0237
CASE*Q6	4	4.0738	1.0184	1.10	0.3596
CASE*Q7	4	3.8571	0.9643	1.04	0.3886
CASE*Q8	4	1.8977	0.4744	0.51	0.7265
Error(CASE)	119	110.1195	0.9254		

PANEL C

Variable	N	Mean	Std Dev
<i>Refund Tax Case Y (C1)</i>	132	2.2955	1.0753
<i>High pay Tax Case X (C2)</i>	132	2.7879	1.2358
Q6	132	1.9849	0.9961
Q7	132	2.4394	1.3265
Q8	132	2.7424	1.2578

Table 8.11 contains results generated from the full sample of participants. The first two panels make it clear that the responses to the three compliance questions produced no *between-subjects* or *within-subject* main or interaction effects with respect to choices made by participants to *refund Tax case Y* and *high pay Tax Case X*.

No divergence from the overall pattern reported in Table 8.11 were discernible in terms of the \$3,500 and \$4,200 cash float subsamples.

Table 8.12 contains the results obtained from the \$3,500 group. Contemplation of Panel C in both Tables 8.11 and 8.12 provides an interesting insight into the attitudes of the \$3,500 cash float subsample in contrast with those of the full sample. Panel C of Table 8.12 indicates that

TABLE 8.12
Responses to Compliance Questions with Respect to
Refund Tax Case Y and High Pay Tax Case X.
\$3,500 Subsample (N = 36)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q6	4	4.0205	1.0051	0.47	0.7601
Q7	4	3.8526	0.9632	0.45	0.7739
Q8	4	3.8821	0.9705	0.45	0.7715
Error	23	49.6290	2.1578		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.4724	0.4724	0.51	0.4841
CASE*Q6	4	4.4081	1.1020	1.18	0.3457
CASE*Q7	4	3.7595	0.9399	1.01	0.4244
CASE*Q8	4	0.7814	0.1953	0.21	0.9307
Error(CASE)	23	21.4756	0.9337		

PANEL C

Variable	N	Mean	Std Dev
<i>Refund</i> Tax case Y (C1)	36	2.3889	1.1778
<i>High pay</i> Tax Case X (C2)	36	2.4444	1.1819
Q6	36	2.1389	1.0462
Q7	36	2.6389	1.2907
Q8	36	2.8056	1.1909

the \$3,500 cash float participants were, on average, slightly more accepting of non-compliance for both high and low income earners than was the case for the full body of participants (Panel C of Table 8.11). At the same time, however, the \$3,500 subjects were also happier than the full sample to pay terminal taxes owing. No significant trend could be extracted from this ambivalence.

The means contained in Panel C of Table 8.13 indicate that the \$4,200 cash float participants were, on average more strongly disapproving of non-compliance for both low and high income earners than was the case for the full sample; and, unlike any other grouping of participants, they actually claimed to have positive feelings about making their cash contribution to the New Zealand Inland Revenue Department's tax collection (Question 8). Nevertheless, no

TABLE 8.13

**Responses to Compliance Questions with Respect to
Refund Tax Case Y and High Pay Tax Case X.
\$4,200 Subsample (N = 30)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q6	2	0.9956	0.4978	0.32	0.7333
Q7	3	4.7048	1.5683	0.99	0.4171
Q8	4	9.8747	2.4687	1.56	0.2245
Error	19	29.9919	1.5785		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.1224	0.1224	0.11	0.7448
CASE*Q6	2	0.3233	0.1617	0.14	0.8668
CASE*Q7	3	1.6372	0.5457	0.49	0.6958
CASE*Q8	4	1.6540	0.4135	0.37	0.8281
Error(CASE)	19	21.3238	1.1223		

PANEL C

Variable	N	Mean	Std Dev
<i>Refund</i> Tax Case Y (C1)	30	2.4333	1.1351
<i>High pay</i> Tax case X (C2)	30	2.8333	1.3153
Q6	30	1.6667	0.9223
Q7	30	1.8000	1.0635
Q8	30	3.3667	1.2994

statistically significant impact was found with respect to their Case Y and Case X decision problem choices.

8.5.6 H_{1A} : The Complete Picture

The null form of H_{1A} (that participants will choose riskier options in the *high pay* case than in the *refund* case) may be rejected on the evidence provided by the full sample when the influence of the cash float variable was ignored, and on the evidence provided by the low cash float participants. This finding accords with Dusenbury's rejection of the null form with respect to H_{1A} in his study.

However, the high cash float subsample, examined as a whole and in terms of its two nested subsamples (determined by dollar size), produced evidence on which the null form of the hypothesis could not be rejected. This phenomenon could not be attributed to the participants' stated attitudes towards non-compliance; but it could possibly be attributed to a remuneration expectation set up by knowledge of the funding arrangements associated with the study.

The phenomenon may well indicate that participants with high cash floats may have approached the decision problems with less fear of encountering a cash flow crisis than did the low cash float subjects. This effect may have been induced by all participants knowing there were two cash float levels (a low and a high one), and knowing which level float they personally were given. The fact that the low float participants provided a clearer shift towards risk seeking, when faced by *high pay* Tax Case X, from each individual's benchmark risk preference level established in the *refund* Tax Case Y, may indicate they took the decision problems more seriously. Using their common sense, they would be aware that a lean float in a two-float experiment meant they could be expected to need to exercise greater care to avoid cash flow difficulties. If this reasoning is accurate, then it was a flaw in the current study that participants were allowed to know there were two levels of cash float, and which of these floats applied to them personally.¹⁴ It is to be noted that the mean of the high float participants' responses moved in the same direction as those of the low group; but that the movements remained statistically insignificant. This information was disclosed in Subsection 8.5.4 above.

8.6 HYPOTHESIS H_{2A}

8.6.1 The Full Sample and H_{2A}

The second replication hypothesis, as formulated in its alternative form by Dusenbury (1994) and disclosed in Chapter Four Subsection 4.4.1 of this study, states¹⁵:

¹⁴ The participants in the two high cash float subsamples were not aware of the shift from \$4,200 to \$3,500, which was made as an economy measure in the course of running the experimental sessions.

¹⁵ Dusenbury op. cit. n. 4, p. 11 and Chapter Four, Subsection 4.4.1.

TABLE 8.14

**High Pay Tax Case X versus Medical Insurance Case.
Full Sample (N = 132)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	1.0115	1.0115	0.69	0.4095
CASE*GROUP	3	4.4095	1.4698	1.00	0.3977
CASE*CASEORD	18	29.1335	1.6185	1.10	0.3677
CASE*GROUP*CASEORD	29	48.8122	1.6832	1.14	0.3120
Error(CASE)	81	119.1810	1.4714		

PANEL B

Response Variable	N	Mean	Std Dev
<i>High pay</i> Tax Case X (C2)	132	2.7879	1.2358
Medical Case (C3)	132	2.4242	1.2848

H_{2A}: Participants will choose less risky options in the health insurance case than in the *high pay* tax case.

In every instance, participants sorted by cash float grouping adopted higher mean risk levels in response to the *high pay* tax scenario than to the medical scenario. This is consistent with Dusenbury's (1994) findings in Florida.

However, none of the differences in the means of the risk levels chosen by participants sorted by cash float category were statistically significant until a further partitioning of the sample was undertaken. When the sample and the low cash float subsample were both divided into subsamples according to the participants' attitude towards the holding of private health insurance, a reasonable explanation for the New Zealand study's divergence from Dusenbury's (1994) result emerged.¹⁶ Table 8.14 provides the statistical details with respect to the full sample without partitioning of any sort.

¹⁶ In accordance with the treatment of H_{1A}, the *repeated measures* GLM univariate analysis of variance results are reported in this subsection; and the *basic model* equivalents may be found in Appendix F, Section F.2.

In terms of the 132 participants undifferentiated by cash float category or by attitude to the holding of health insurance, CASE is the least significant variable both in terms of its F -statistic and in terms of that statistic's likelihood of error. There is no evidence of a scenario-related framing effect here; and none of the *within-subjects* interaction effects possess the statistical significance needed to explain the difference between the reported means.

However, when the full sample was partitioned into two subsamples according to the nature of the participants' responses to Question 4 in the end-of-session questionnaire, soliciting the participants' attitudes regarding the need for private health insurance, the outcome of the *repeated measures* GLM procedure was markedly different. Question 4, repeated here for convenience, has the number of responses for each box recorded in brackets.¹⁷

Q4: Attitude to health insurance:

Given the current state of the public health system in New Zealand, how worthwhile do you consider a personal health insurance scheme to be for you or other members of your family?:

Please tick the box which most closely approximates your judgement of it:

- | | |
|---|------|
| <input type="checkbox"/> It is ridiculous | (4) |
| <input type="checkbox"/> It is not very necessary | (15) |
| <input type="checkbox"/> I am neither for nor against | (9) |
| <input type="checkbox"/> I think it is a reasonable thing to have | (47) |
| <input type="checkbox"/> I would not be without it | (57) |

The sample was partitioned into a group of participants who favoured holding private medical insurance (the 104 people who ticked one of the final two boxes), and a smaller group who were either neutral or considered it unnecessary (the 28 people who ticked one of the first three boxes). The results for the subsample favouring private medical insurance are in Panel A of Table 8.15.

The effect of the partitioning is quite clear. Participants who believed there is a need for private medical insurance exhibit a statistically significant shift in risk levels, which is captured in Panel A by the variable CASE ($F = 4.79$, $\text{Pr} > F = 0.0324$). On the other hand, the

¹⁷ A Likert scale was used, ranging from 1 for the top box up to 5 for the bottom box. This use of Likert scales is in accordance with the treatment of all the tick-in-box questions used in this study. See Footnote 12 of this chapter.

TABLE 8.15

**High Pay Tax case X versus the Medical Insurance Case
Partitioned by Q4 (Attitude to Health Insurance).
Full Sample (N = 132)**

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

PANEL A:

Participants who Support Private Medical Insurance (N = 104)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	7.2183	7.2183	4.79	0.0324
CASE*GROUP	3	7.7834	2.5945	1.72	0.1717
CASE*CASEORD	17	31.4592	1.8505	1.23	0.2708
CASE*GROUP*CASEORD	21	37.8187	1.8009	1.20	0.2871
Error(CASE)	62	93.4298	1.5069		

PANEL B

Participants indifferent to, or against Private Medical Insurance (N = 28)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	1.1307	1.1307	2.19	0.1991
CASE*GROUP	3	3.9700	1.3233	2.56	0.1682
CASE*CASEORD	12	18.1209	1.5101	2.92	0.1224
CASE*GROUP*CASEORD	7	6.4971	0.9282	1.80	0.2687
Error(CASE)	5	2.5833	0.5167		

participants who dismissed the need for private medical insurance fail to provide evidence of a shift in risk level in Panel B. There were no significant interaction effects apparent in either subsample.¹⁸

The diagnosis provided by Table 8.15 clearly corroborates the behaviour of the means of the responses recorded by the participants in the two health insurance attitude subsamples. This information is contained in Table 8.16.

The 104 participants who approved of private medical insurance adopted a higher mean risk level with respect to *high pay* Tax Case X than in the Medical Insurance Case; but this position was actually reversed by the 28 participants who did not see the need for holding such insurance in contemporary New Zealand.

¹⁸ The between-subjects test results for these two private medical insurance-partitioned subsamples may be found in Appendix F, Section F.2.

TABLE 8.16		
Means obtained when the Full Sample was partitioned by Attitude to Personal Health Insurance (Q4)		
Attitude to Personal Health Insurance Subsample Size (N)	In Favour N = 104	Not in Favour N = 28
<i>High Pay Tax Case X (C2)</i>	2.7981	2.7500
Medical Insurance Case (C3)	2.5000	3.0714

While the information contained in Tables 8.15 and 8.16 does not allow the null form of H_{2A} to be rejected without reservation, it does explain why the New Zealand study's participants appeared, in the first instance, to behave differently from the North American participants in Dusenbury's (1994) study.

The participants who have no faith in the public health system in New Zealand have reacted in a similar fashion to their American counterparts; but the participants who do have faith in the New Zealand public health system's ability to meet their medical needs have translated their discounting of the necessity of private health insurance into a propensity for taking relatively higher risks in the Medical Insurance Case provided in this New Zealand tax compliance experiment.

8.6.2 The Low Cash Float Subsample and H_{2A}

The dearth of a significant scenario-based framing effect noted with respect to the full sample (left unpartitioned by attitude to health insurance) recurs when the sample is reduced to the low cash float subsample. This is indicated by the statistics disclosed in Panel A of Table 8.17.

In terms of the low cash subsample partitioned by health insurance attitude, the participants who favoured holding a private policy displayed a stronger shift in risk level between *high pay* Tax Case X and the Medical Insurance Case than occurred with respect to the full sample. This is confirmed by the statistic for CASE in Panel A ($F = 5.85$, $Pr > F = 0.0243$).

Panel C of Table 8.17, which furnishes the statistical output for the low cash float participants who did not value private medical insurance, contains a CASE F -statistic which is not significant. Neither Panel B nor Panel C contain any significant interaction effects.

TABLE 8.17

**High Pay Tax Case X versus Medical Insurance Case.
Low Cash Float Subsample (N = 66)**

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

PANEL A

Low Cash Float Subsample without Partitioning (N = 66)¹⁹

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.2711	0.2711	0.26	0.6159
CASE*GROUP	3	2.7516	0.9172	0.87	0.4680
CASE*CASEORD	15	24.7968	1.6531	1.57	0.1473
CASE*GROUP*CASEORD	19	42.1817	2.2201	2.11	0.0357
Error (CASE)	28	29.5000	1.0536		

PANEL B

Participants who support Private Medical Insurance (N = 50)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	6.2677	6.2677	5.85	0.0243
CASE*GROUP	3	3.3441	1.1147	1.04	0.3942
CASE*CASEORD	14	23.9462	1.7104	1.60	0.1580
CASE*GROUP*CASEORD	10	19.2105	1.9210	1.79	0.1218
Error(CASE)	22	23.5667	1.0712		

PANEL C

Participants indifferent to, or against Private Medical Insurance (N = 16)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.7104	0.7104	2.84	0.3409
CASE*GROUP	2	1.5694	0.7847	3.14	0.3707
CASE*CASEORD	9	14.0000	1.5556	6.22	0.3021
CASE*GROUP*CASEORD	2	4.3235	2.1618	8.65	0.2338
Source: Error(CASE)	1	0.2500	0.2500		

The behaviour of the risk levels noticed with respect to the 132 participants of the full sample recurs in terms of the low cash float subsample. This information is contained in Table 8.18.

The four means contained in Table 8.18 clearly show, in the first instance, that the low cash float recipients who valued private health insurance cover were more risk averse in the Medical

¹⁹ Panel A of Table 8.17 contains one result which is of interest. The three-way interaction effect, CASE*GROUP*CASEORD, in terms of the low cash float subsample, has become significant ($F = 2.11$, $Pr > F = 0.0357$). Since GROUP and CASEORD did not become significant when all five sets of replication scenario responses were tested for noise of this nature, it can be assumed that this low cash float response may be related to the influence of one or more confounding variables. However, this interaction effect did not recur when the subsample was partitioned with respect to attitude towards the holding of private medical insurance in Panels B and C.

Attitude to Personal Health Insurance Subsample Size (N)	In Favour N = 66	Not in Favour N = 16
<i>High Pay Tax Case X (C2)</i>	2.9200	2.0625
Medical Insurance Case (C3)	2.2400	3.3125

Insurance Case than in *high pay* Tax Case X. In the second instance, they show that the low cash float subsample participants who did not value this form of insurance were more risk seeking in the health context.

8.6.3 The Combined High Cash Float Subsample and H_{2A}

When the combined high cash float subsample was investigated in the absence of partitioning by the attitude towards private health insurance solicited by Question 4, the phenomenon noted with respect to the full sample and low cash float subsample repeated itself with only minor variation. The statistics for the unpartitioned combined high cash float subsample are supplied in Panel A of Table 8.19. The *F*-statistics associated with both CASE and each of the interaction effects are very small indeed.

However, partitioning the combined high cash float subsample by the participants' responses to Question 4 made little difference. The statistical insignificance of CASE in Panel B of Table 8.19 indicates that no shift in risk levels could be corroborated, although the mean of the 54 health insurance policy supporters' recorded responses was 2.6852 with respect to *high pay* Tax Case X, which was higher than their mean of 2.2593 for the Medical Case.

Because there is no longer any significant difference between the combined high cash float subsample in its partitioned and unpartitioned forms, tables disclosing the *repeated measures* GLM analysis of variance results for each of the \$3,500 and \$4,200 cash floats have been relegated to, Appendix F, Section F.2, Table 2. In all instances involving the high cash float subsample, it was not possible to reject the null form of H_{2A}.

TABLE 8.19

**High Pay Tax Case X versus Medical Insurance Case
Effect of Partitioning by Q4 (Attitude to Health Insurance).
Combined High Cash Float Subsample (N = 66)²⁰**

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

PANEL A

Combined High Cash Float Subsample without Partitioning (N = 66)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	1.4286	1.4286	0.86	0.3615
CASE*GROUP	3	4.3158	1.4386	0.87	0.4699
CASE*CASEORD	17	27.0735	1.5926	0.96	0.5232
CASE*GROUP*CASEORD	17	26.5215	1.5601	0.94	0.5418
Error (CASE)	28	46.4750	1.6598		

PANEL B

Participants who Support Private Medical Insurance (N = 54)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	3.1898	3.1898	1.77	0.1958
CASE*GROUP	3	0.9632	0.3211	0.18	0.9098
CASE*CASEORD	15	20.5119	1.3675	0.76	0.7037
CASE*GROUP*CASEORD	12	23.7415	1.9785	1.10	0.4043
Error(CASE)	23	41.3333	1.7971		

8.6.4 H_{2A}: The Complete Picture

The null form of H_{2A} cannot be rejected on the strength of the evidence provided by the testing which has been done. There is a very good reason. The sample of participants under investigation could be divided into ideologically heterogeneous subsamples. The results of the testing depended on the side of the ideological divide from which the sample was selected; and partitioning by the responses made to Question 4 (*attitude to personal health insurance*) provided subsamples from both sides.

The failure to reject the null form of H_{2A} cleanly is traceable to changes occurring in New Zealand society. As late as the mid 1980s, New Zealanders could safely assume that state-

²⁰ Because there were only twelve participants in the combined high cash float subsample who did not profess to value private medical insurance, the *repeated measures* GLM analysis of variance procedure, in this instance, had insufficient degrees of freedom to produce results for that category of participant.

funded services could and would meet their medical needs. This was one of the widely understood functions of what was known as the Welfare State. The restructuring of the New Zealand public health system in the last decade has lessened citizens' confidence in its adequacy; but even so, many New Zealanders maintain an underlying belief that when the 'chips are down', the state will provide the necessary care, funded out of the tax dollar.

This explanation, however, does not adequately account for the behaviour of the high cash float participants. Perhaps the phenomenon suggested as an explanation for the lack of significance of risk preference shifts with respect to H_{1A} in Subsection 8.5.5 also applies with respect to the high cash float participants' responses to the two H_{2A} decision problems: these people might have been made complacent in their decision making by the knowledge they had been given the higher of the two stated cash floats.

8.7 HYPOTHESIS H_{3A}

8.7.1 The Full Sample and H_{3A}

Dusenbury's third hypothesis was²¹:

H_{3A} : Participants will select riskier options in the gamble case than in the *low pay* tax case.

Dusenbury's repeated measures ANOVA procedure on his sample of 65 North Americans with respect to these two cases produced strong support for rejecting the hypothesis' null form. The *F*-statistic he obtained, confirming a significant difference between the means of the two cases, was 31.68 with an associated probability of 0.0001 of a Type I error.

²¹ Dusenbury op. cit. n. 4, p. 11 and Chapter Four, Subsection 4.4.1.

TABLE 8.20

**Low Pay Tax Case Z versus Gamble Case.
Full Sample (N = 132)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	36.7463	36.7463	27.22	0.0001
CASE*GROUP	3	0.7758	0.2586	0.19	0.9019
CASE*CASEORD	18	22.3156	1.2398	0.92	0.5594
CASE*GROUP*CASEORD	29	30.8380	1.0634	0.79	0.7621
Error (CASE)	81	109.3667	1.3502		

PANEL B

Response Variable	N	Mean	Std Dev
Low pay Tax Case Z (C4)	132	2.2197	1.1278
Gamble Case (C5)	132	3.3485	1.4302

TABLE 8.21

**Low Pay Tax Case Z versus the Gamble Case
Low Cash Float Subsample (N = 66)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	22.4902	22.4902	18.85	0.0002
CASE*GROUP	3	2.7521	0.9174	0.77	0.5210
CASE*CASEORD	15	16.2907	1.0860	0.91	0.5627
CASE*GROUP*CASEORD	19	22.2763	1.1724	0.98	0.5054
Error (CASE)	28	33.4000	1.1929		

PANEL B

Response Variable	N	Mean	Std Dev
Low pay Tax Case Z (C4)	66	2.0455	1.0441
Gamble Case (C5)	66	3.4242	1.4257

In four subsets out of the five generated in the New Zealand replication, this result was quite firmly corroborated. When a *repeated measures* GLM procedure was run on the full sample, the resulting *F*-statistic was 27.22 with a level of error ($\text{Pr} > F = 0.0001$) as small as that obtained in the North American study; and as in Dusenbury's test, no *within-subject* interaction effects were found to be significant. These figures are laid out in Panel A of Table 8.20.

TABLE 8.22

**Low Pay Tax Case Z versus the Gamble Case.
Combined High Cash Float Subsample (N = 66)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	27.8263	27.8263	15.98	0.0004
CASE*GROUP	3	5.1760	1.7253	0.99	0.4115
CASE*CASEORD	17	23.5127	1.3831	0.79	0.6855
CASE*GROUP*CASEORD	17	10.7438	0.6320	0.36	0.9839
Error (CASE)	28	48.7667	1.7417		

PANEL B

Response Variable	N	Mean	Std Dev
Low pay Tax Case Z (C4)	66	2.3939	1.1881
Gamble Case (C5)	66	3.2727	1.4416

8.7.2 The Low Cash Float Subsample and H_{3A}

When the hypothesis was tested on the low cash float subsample, the *repeated measures* procedure produced similar results; but the value of the *F*-statistic fell to 18.85 and the probability of a Type I error doubled ($\text{Pr} > F = 0.0002$). Nevertheless, the figures disclosed in Table 8.21 still firmly support rejection of H_{3A} 's null form.

8.7.3 The Various High Cash Float Subsamples and H_{3A}

The pattern continued with the results obtained from a *repeated measures* GLM procedure run on the combined high cash float subsample. In this instance, the *F*-statistic signifying the difference between the means of responses to the two cases dropped to a still strongly significant 15.98, while the probability of a Type I error doubled again to a still minuscule 0.0004. Furthermore, in keeping with the two result-sets already obtained with respect to H_{3A} , all *within-subject* interaction effects (and for that matter, the *between-subjects* main and interaction effects not tabled here) were found to be statistically dismissable. This information is laid out in Table 8.22.

TABLE 8.23

**Low Pay Tax Case Z versus the Gamble Case.
\$3,500 Cash Float Subsample (N = 36)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	15.9298	15.9298	8.05	0.0140
CASE*GROUP	2	1.1231	0.5616	0.28	0.7575
CASE*CASEORD	11	8.8626	0.8057	0.41	0.9280
CASE*GROUP*CASEORD	9	3.1733	0.3526	0.18	0.9931
Error (CASE)	13	25.7250	1.9788		

PANEL B

Response Variable	N	Mean	Std Dev
<i>Low pay</i> Tax Case Z (C4)	36	2.4722	1.0278
Gamble Case (C5)	36	3.3611	1.1748

However, when the high cash float subsample was divided into its component \$3,500 and \$4,200 sub-subsamples, a slightly different situation emerged. The results for the \$3,500 cash float subsample are disclosed in Table 8.23. They follow the diminishing trend established with respect to the previous three sets of H_{3A} results.

The difference in chosen risk levels is still significant; but the F -statistic signifying it has dropped to 8.05, and the probability of a Type I error has risen ($Pr > F = 0.0140$). Nevertheless, the *within-subject* interaction effects (and undisclosed *between-subject* effects) remain insignificant. H_{3A} 's null form is still refutable on this evidence; but this is not true of the evidence laid out in Table 8.24 with respect to the \$4,200 cash float subsample.

It is clear that the trend of diminishing CASE-related F -statistics is strongly confirmed in this instance, with a minuscule $F = 3.83$ with an associated 0.0821 probability of error. Possible causes for this trend are considered in the next subsection.

8.7.4 Investigation Incorporating the Attitude to Gambling Variable

Because the means of the risk levels chosen in *low pay* Tax Case Z and the Gamble Case were statistically indistinguishable at the 0.05 level of probable error with respect to the \$4,200 cash

TABLE 8.24

**Low Pay Tax Case Z versus Gamble.
\$4,200 Cash Float Subsample (N = 30)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	6.0261	6.0261	3.83	0.0821
CASE*SESSION	3	4.2582	1.4194	0.90	0.4774
CASE*CASEORD	14	18.8962	1.3497	0.86	0.6158
CASE*SESSION*CASEORD	3	3.8656	1.2885	0.82	0.5155
Error (CASE)	9	14.1667	1.5741		

PANEL B

Response Variable	N	Mean	Std Dev
<i>Low pay</i> Tax Case Z (C4)	30	2.3000	1.3684
Gamble Case (C5)	30	3.1667	1.7237

float subsample, which was recruited solely from the Anglican congregation, a *repeated measures* GLM procedure was run on this subsample's Case Z and Gamble Case responses with their answers to Question 11 from the Questionnaire as the independent variable. Question 11 was worded as follows.²²

Q11: Feelings about Gambling:

This question is about how you view the rights and wrongs of gambling. Your attitude to spending say \$5 a week on a Lotto ticket is most closely approximated by which single box:

- I would never buy a ticket because gambling is always wrong
- I am happy to support a charity that funds good causes; but I consider regular ticket-buying to be not right for me
- I do not have an opinion either way
- I am quite happy to have a regular flutter of this sort; but if Lotto was banned, I would not miss it.
- I feel good about playing Lotto and would resent it if it were banned

²² The boxes in this question corresponded, from top box to bottom, to a Likert scale running from 1 to 5. See Footnote 12 of this chapter.

TABLE 8.25

**Low Pay Tax Case Z versus the Gamble Case with Respect to Attitudes to Gambling.
\$4,200 Cash Float Subsample (N = 30)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q11	2	16.7304	8.3652	4.47	0.0561
SESSION	3	5.5878	1.8626	1.00	0.4490
CASEORD	14	52.2779	3.7341	1.99	0.1814
SESSION*CASEORD	2	6.4168	3.2084	1.71	0.2478
Error	7	13.1029	1.8718		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	2.7571	2.7571	1.40	0.2760
CASE*Q11	2	0.3382	0.1691	0.09	0.9189
CASE*SESSION	3	4.1190	1.3730	0.70	0.5838
CASE*CASEORD	14	15.6078	1.1148	0.56	0.8285
CASE*SESSION*CASEORD	2	3.0412	1.5206	0.77	0.4987
Error (CASE)	7	13.8284	1.9755		

PANEL C

Variable	N	Mean	Std Dev	Minimum	Maximum
Q11	30	2.5667	1.0063	1.000	4.000

Of immediate interest is the subsample's mean response to Question 11, disclosed in Panel C of Table 8.25. The mean of 2.57 implies that this group, on average, viewed gambling in a negative light.²³ Also of note is the fact that the maximum observation recorded in Panel C is only 4. Nobody in the subsample indicated strong, unreserved support for gambling (as proxied by Lotto). However the main effect recorded when these participants' Question 11 responses were subjected to a *repeated measures* GLM procedure was statistically significant only at the 0.0561 probability of a Type I error (see Panel A); and the interaction effect recorded in Panel B was quite insignificant.²⁴ Nevertheless, the mere inclusion of an 'attitude

²³ This assumption may be made on the ground that Lotto is generally considered to be a relatively innocuous form of gambling which raises money for voluntary organisations. Disapproval of this mild form of gambling may be taken as a proxy for disapproval of making bets and wagers for money in general.

²⁴ The equivalent *basic model* GLM procedure, with the Gamble Case variable as the sole dependent variable and Q11 as the sole categorical variable, was run on both the full sample of 132 participants and on the \$4,200 cash float subsample. While Q11 produced a statistically insignificant *F*-statistic for the full sample, when the Gamble Case responses were restricted to those

to gambling' variable in the analysis of variance caused the difference between Tax Case Z's and the Gamble Case's means to become thoroughly insignificant. This is shown in Panel B of Table 8.25.

When a *repeated measures* GLM procedure of a similar nature was run on the full sample, any impact originating from Q11 was shown to be statistically insignificant.²⁵

The anomalous \$4,200 cash float subsample was also investigated for evidence of a significant impact emanating from attitudes held about compliance in terms of the three compliance questions (Q6, Q7 and Q8) asked in the questionnaire (and repeated in full in Subsection 8.5.5 above). When these three questions were used as independent variables in a *repeated measures* GLM analysis of variance procedure with Tax Case Z and the Gamble Case as the two dependent variables, no significant tax-related main or interaction effects were found. The results generated by this procedure are reported in Table 8.26.

Nevertheless, the means of Questions 6 and 7 (disclosed in Panel C of Table 8.26) indicate that the 30 respondents of the \$4,200 cash float group considered non-compliance to be generally unacceptable, while the mean of Question 8 for the subsample suggests that these people have positive feelings about the process of paying their own taxes.

8.7.5 H_{3A}: The Complete Picture

The results of the New Zealand replication largely agree with Dusenbury's finding in his North American context. Broadly speaking, the null form of H_{3A} can safely be rejected on the evidence presented in all but one of the last few subsections. The New Zealand participants, when analysed in terms of the full sample, exhibited a significantly greater degree of risk willingness in the Gamble Case than in the *low pay* tax case, Case Z; and this implies the

of the \$4,200 recipients only, the associated *F*-statistic was associated with a .0765 level of error. See Appendix F, Section F.3, Table F. 7.

²⁵ The results of the *repeated measures* GLM procedure run on the full sample may be found in Panel A of Table F.7 in Appendix F, Section F.3. Corroborative evidence is also provided in this appendix in the form of a table disclosing the results of a *basic model* GLM procedure on both the full sample and the \$4,200 subsample. While Q11 had an *F*-statistic which was significant at the 8 percent level of possible error for the \$4,200 subsample, the level of error shot up to 26 percent for the full sample. The variable was clearly unimportant in this context.

TABLE 8.26

**Low Pay Tax Case Z versus the Gamble Case with Respect to Disclosed Attitudes to Compliance.
\$4,200 Cash float subsample (N = 30)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q6	2	3.4352	1.7176	0.57	0.5735
Q7	3	8.2143	2.7381	0.91	0.4534
Q8	4	6.7842	1.6960	0.57	0.6907
Error	19	56.9936	2.9997		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	1.7118	1.7118	0.93	0.3479
CASE*Q6	2	2.0923	1.0461	0.57	0.5770
CASE*Q7	3	2.4110	0.8037	0.43	0.7305
CASE*Q8	4	3.8654	0.9663	0.52	0.7201
Error(CASE)	19	35.1124	1.8480		

PANEL C

Variable	N	Mean	Std Dev
Tax Case Z	30	2.3000	1.3684
Gamble Case	30	3.1667	1.7237
Q6	30	1.6667	0.9223
Q7	30	1.8000	1.0635
Q8	30	3.3667	1.2994

existence of a scenario-based decision frame effect since the ranges of expected values in the two scenarios were identical.

However, H_{3A} 's null form cannot be rejected on the evidence of the \$4,200 cash float subsample of 30 Anglican parishioners. But neither their relatively negative perception of gambling, nor their relatively positive perception of compliance (both in themselves or in others), provided a reason for the lack of a perceptible shift in risk profile statistically significant within the five percent level of significance.

Attention will now focus on the characteristics of the participants. These are discussed in Chapter Nine.

9. CHARACTERISTICS NOTED FROM THE QUESTIONNAIRE

9.1 INTRODUCTION

Two issues connected with data obtained from the end-of-session questionnaire, which every participant filled out, are explored in this chapter. The first of these is the nature of the sample's participants; and the second is, do any of the variables, on which data was gathered in the questionnaire, have an impact on the choice of risk levels participants made when working through the decision problems set out in the experiment?

To date the sample has been treated as a body of taxpayers assumed to be homogeneous in all respects except for the level of the cash float provided at the start of the experiment. However, at this point it would be profitable to have a more thorough demographic understanding of the participants, given that they were recruited from four very dissimilar organisations. Information about the participants will also be of value in determining the representational faithfulness of the sample with respect to the New Zealand taxpaying public at large. It will also be of assistance in demonstrating quantitatively that the New Zealand participants were valid equivalents of Dusenbury's 65 North American subjects.

Furthermore, it would be useful to check the characteristics disclosed by the participants in their questionnaire responses against the risk profiles they exhibited in the experiment. This latter exploration will be restricted to consideration of potential impacts on risk levels chosen in the three replication tax decision contexts, *refund* Tax Case **Y**, *high pay* Tax Case **X** and *low pay* Tax Case **Z**. The imposition of this restriction affords a closer comparison with Dusenbury (1994), since this is the way in which Dusenbury utilised and reported his questionnaire data.¹

¹ Dusenbury mentioned his questionnaire only very briefly. He disclosed that he had gathered data on *years of tax filing experience*, *self-reported non-compliance*, *attitudes towards taxpaying*, and *frequency of refund and payment-due tax filings*. Of these sets of observations, Dusenbury found that only *tax-filing experience* had a main effect on risk levels chosen; but not a *within-subject* interaction effect. See Dusenbury, R., (1994), "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options", p.13.

The chapter is organised as follows: Section 9.2 provides summary information concerning the participants in terms of the sample and also (where appropriate) in terms of their recruitment origin. Section 9.3 then utilises variables coded from responses to the questionnaire to determine whether any of these generate statistically significant main or interaction effects in a *repeated measures* GLM analysis of variance procedure in which the three replication tax case variables are jointly the dependent variables.

9.2 NATURE OF THE SAMPLE

9.2.1 Introduction

Three questions immediately spring to mind regarding the volunteers recruited to participate in a social science experiment in which random sample selection was not feasible. Who were the participants? How did they compare with the subjects of Dusenbury's (1994) non-random sample? And, how reliable a representation of the taxpaying public of New Zealand were they? These issues are addressed in the next two subsections. Subsection 9.2.2 deals with the specifications required of participants by Dusenbury; then Subsection 9.2.3 broadens the focus to include further specifications which make the current study's sample appropriate for the purpose to which it is put. Subsection 9.2.4 then closes this area of investigation with some evidence and reasoning which suggest that the sample, nevertheless, fails to achieve the representational faithfulness a randomly selected sample might be claimed to provide.

9.2.2 Comparison with Dusenbury's Requirements

Dusenbury (1994) required his participants to be at least 30 years old or be fully employed for more than two years. The minimum number of tax return filings reported by his participants was four (by two participants); and 78 percent of the sample had had ten years, and upwards, of filing experience.² In contrast, the New Zealand sample was recruited on the basis that participants must have had *some* experience in filing tax returns. Since most people who, as

² Dusenbury (1994), op. cit. n. 1, p. 9.

TABLE 9.1
Breakdown by Age and Recruitment Origin (N = 132)

RAW COUNT	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
18 to 20 years	5	5	0	0	0
21 to 25 years	9	1	1	0	7
26 to 30 years	11	3	3	4	1
31 - 40 years	51	5	28	13	5
41 - 60 years	43	33	7	2	1
More than 60 years	13	12	0	1	0
Sum	132	59	39	20	14
PERCENTAGES	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
18 to 20 years	3.79	8.47	0.00	0.00	0.00
21 to 25 years	6.82	1.69	2.56	0.00	50.00
26 to 30 years	8.33	5.08	7.69	20.00	7.14
31 - 40 years	38.64	8.47	71.79	65.00	35.71
41 - 60 years	32.58	55.93	17.95	10.00	7.14
More than 60 years	9.85	20.34	0.00	5.00	0.00

pay period taxpayers, earn \$20,000 and less are currently not obliged to file a return under New Zealand tax law,³ the assumption was made that filing experience equates with some form of employment experience, and that filing with a low income may imply expectations of a positive tax adjustment after the end of the income year.

9.2.2.1 Gender and Age

With respect to age, Table 9.1 provides a breakdown of the participants' details.⁴ From Table 9.1 it can be seen that the bulk of the participants (94 people or 71 percent) were more than thirty years of age, but not more than sixty years old.

³ Subpart IC1(2) of the Income Tax Act 1994 (NZ) states that a pay-period taxpayer need not furnish a tax return unless required to do so by the Commissioner of Inland Revenue; and Subpart OB4(1) defines a pay-period taxpayer as a natural person who receives not more than \$20,000 per annum from the sum of employment income and income in the nature of interest. However, this interest income may not exceed \$1,500; and taxation is required to have been levied on all income via the pay-as-you-earn system (PAYE), collected by employers, and resident withholding tax on dividends and interest (RWT), collected by financial institutions and companies. All other income earners are required to furnish an annual tax return by Section 33 of the Tax Administration Act 1994 (NZ).

⁴ The percentages shown in this and the following tables are rounded to two decimal places. This gives rise to a rounding error of 0.01 percent.

When the sample was divided into its four subsamples by recruitment origin, the church congregation furnished the widest age spread, accounting for all subjects under twenty and all but one of the 13 subjects over sixty. Forty five (76 percent) of the 59 recruits in the modified congregation subsample were aged forty or more.

By contrast, the bulk of the 39 kindergarten recruits (28 people or 72 percent) were in their thirties; as were 65 percent of the school subsample (13 of the 20 recruits). The thirties age group also accounted for 36 percent (5 of the 14 recruits) of the Lincoln graduate student subsample; but in this case, 50 percent (7 students) were in the twenty-one to twenty-five age bracket.

While Dusenbury did not disclose the gender of his subjects, gender was recorded in the current study. The gender balance was quite uneven. Forty nine of the participants were male (37 percent) and 83 were female (63 percent). This inequality was detected in all samples by recruitment origin, and was most pronounced in the case of the school subsample, in which there were 19 females (95 percent) and one male (5 percent).⁵

9.2.2.2 Filing Experience

With regard to the annual filing of tax returns, most of the participants (100 people or 76 percent) had more than ten years' filing experience. Return-filing summary data is set out in Table 9.2.

When this data was broken down by subsample, 71 percent of the church congregation recruits (42 people) had had more than 20 years of filing experience; while for the kindergarten subsample there was a wider spread with 33 percent (13 people) recording between 16 and 20 years, 28 percent (11 people) recording between 11 and 15 years, and 18 percent (7 people) recording more than 20 years experience.

⁵ The impact of gender-related effects on H_{1A} is investigated later in this chapter. See Subsection 9.3.2.1.

RAW COUNT	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
1 to 5 years	17	6	4	1	6
6 to 10 years	15	4	4	1	6
11 to 15 years	19	2	11	6	0
16 to 20 years	25	5	13	6	1
More than 20 years	56	42	7	6	1
Sum	132	59	39	20	14
PERCENTAGES	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
1 to 5 years	12.88	10.17	10.26	5.00	42.86
6 to 10 years	11.36	6.78	10.26	5.00	42.86
11 to 15 years	14.39	3.39	28.21	30.00	0.00
16 to 20 years	18.94	8.47	33.33	30.00	7.14
More than 20 years	42.42	71.19	17.95	30.00	7.14

The school subsample spread over the three same ranges was spectacularly uniform, with 30 percent (6 people) in each range. By contrast, the Lincoln graduate students had less filing experience, which is not surprising given their younger age group. In this subsample, 43 percent had filed between one and five annual returns (6 people); and another 43 percent (6 people again) disclosed six to ten years of filing experience.

9.2.3 Further Salient Features of the New Zealand Sample

9.2.3.1 History of End-of-Year Taxes Payable

It made sense to poll the participants on their experience of terminal tax adjustments with New Zealand's Inland Revenue Department and other tax authorities since the prime focus of this study is on changes in risk profile with regard to fulfilment of such terminal tax obligations. The summary data is contained in Table 9.3.

It is clear from Table 9.3 that approximately 70 percent of the 132 participants had been required to pay an end-of-year tax bill at some time or another. Given this percentage, it is argued that the sample composition is appropriate for the purpose to which it is being put in the study, which is to investigate the influence of a prepayment-based decision frame on taxpayers' risk profiles with respect to their tax return-filing behaviour.

TABLE 9.3
Breakdown by Frequency of Payment-Due Tax Filings (N = 132)

RAW COUNT	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
Never	28	15	8	2	3
20% of years filed	57	25	12	12	8
40% of years filed	19	8	8	2	1
60% of years filed	6	2	3	1	0
80% of years filed	8	3	2	2	1
100% of years filed	3	3	0	0	0
Don't know	11	3	6	1	1
Sum	132	59	39	20	14
PERCENTAGES	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
Never	21.21	25.42	20.51	10.00	21.43
20% of years filed	43.18	42.37	30.77	60.00	57.14
40% of years filed	14.39	13.56	20.51	10.00	7.14
60% of years filed	4.55	3.39	7.69	5.00	0.00
80% of years filed	6.06	5.08	5.13	10.00	7.14
100% of years filed	2.27	5.08	0.00	0.00	0.00
Don't know	8.33	5.08	15.38	5.00	7.14

9.2.3.2 Income-Earning Experience by Type

The income-earning experience of participants was of particular importance to the design of the study because the prediction of diversity of income type among participants was one of the reasons for rewriting the decision problem scenarios.⁶ The nature of income-earning experience disclosed by the participants is summarised in Table 9.4.

This information is set out at this point because the experimental scenarios were customised to cover a range of easily understandable tax situations. It was not known in advance what the exact nature of the participants in the sample would be; and this information became available only once the subjects had completed the experiment and filled in the closing questionnaire. Since the four organisations were told only that adult taxpayers with some tax return filing experience were required, it was reasonable to expect a range of employment types.

From Table 9.4 it is clear the largest grouping in every subsample is employment in the private sector, followed in all but the school subsample by employment in the state sector.

⁶ This was covered in Chapter Six, Section 6.2.

TABLE 9.4
Breakdown by Nature of Income Earning Experience (N=132)

RAW COUNT	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
State sector employee	41	22	12	3	4
Private sector employee	56	31	11	8	6
Self-employed	6	1	2	2	1
Employ others	10	1	6	3	0
Home manager	12	1	7	4	0
Other	7	3	1	0	3
Sum	132	59	39	20	14
PERCENTAGES	TOTAL	CHURCH	KINDERGARTEN	SCHOOL	LINCOLN
State sector employee	31.06	37.29	30.77	15.00	28.57
Private sector employee	42.42	52.54	28.21	40.00	42.86
Self-employed	4.55	1.69	5.13	10.00	7.14
Employ others	7.58	1.69	15.38	15.00	0.00
Home manager	9.09	1.69	17.95	20.00	0.00
Other	5.30	5.08	2.56	0.00	21.43

Collectively these two sectors account for 73 percent of the sample while taxpayers in self employment (including employers of others) account for 12 percent. The final two categories, collectively accounting for 14 percent of the sample, have been included on the basis that household managers and people who have described themselves as ‘other’ have also disclosed histories of filing tax returns. The category, ‘other’, turned out to contain students whose main work experience was possibly holiday employment.

9.2.3.3 Prior Knowledge of Prospect Theory and Expected Utility Theory

Also of note at this point is the level of prior knowledge, held by subjects, of *Prospect Theory* and *Expected Utility Theory*. Since at the time of running the experiment, an unknown proportion of the participants would be university students or graduates, it could be expected that at least some of them would have an awareness of theories of decision making under risk. It turned out that while 50 participants held university degrees, only one person claimed to have studied *Prospect Theory* and *Expected Utility Theory*. Aside from this individual, 127 of the participants professed to having been totally unaware of *Prospect Theory*, while the remaining four had heard of the theory but were unaware of its details. With respect to *Expected Utility Theory*, 123 claimed no prior knowledge; and of the remaining eight (excluding the individual

already mentioned), seven said they had heard of it (but did not know any details) and one professed a moderate awareness of the basic concepts.

The person claiming to have studied the two theories was not removed from the sample on two related grounds. The first of these was that her selection was as random as the recruitment of any of the other 131 taxpayer participants since she responded to a call for volunteers made by her church. The second was that in any taxpaying population there will be people with special knowledge. It could be argued that removing such individuals would reduce bias in the sample. However, this same act could also be interpreted as introducing extra bias into the sample in accordance with a preconception of debatable validity as to the level of knowledge that is representative of New Zealand taxpayers in general.

9.2.4 Representational Faithfulness of the Sample

In this subsection, attention is turned to characteristics which either allow the sample to be considered a faithful representation of the New Zealand taxpaying public, or disqualify it from claiming this status. The most recent demographic information available with respect to New Zealand as a nation was gathered in the Census of Populations and Dwellings, which was held on the 5th of March 1991.⁷ It is recognised that any information published in the census will, by virtue of its age, be only a rough approximation of the actual, but unknown demographic facts of the present time. However, if the values of several salient demographic variables can be shown to be similar, then the claim may be made that the sample provided a relatively (as distinct from absolutely) faithful representation. Census data for both New Zealand as a whole and for the city of Christchurch are considered.

⁷ These are held at five-year intervals. The next one is due to be held during March 1996.

TABLE 9.5			
Income Earning Experience: Sample versus Census.			
Full Sample (N = 132)			
RAW COUNT	SAMPLE	1991 CENSUS	CHRISTCHURCH
Employee	97	1,095,903	100,080
Self-employed	6	159,393	9555
Employ others	10	108,261	7407
Not in workforce	19	1,026,111	97440
PERCENTAGES	SAMPLE	1991 CENSUS[†]	CHRISTCHURCH[†]
Employee	73.48	42.30	43.31
Self-employed	4.55	6.15	4.14
Employ others	7.58	4.18	3.21
Not in workforce	14.39	39.61	42.17
Other Census Category	NA	7.76	7.17
† Other Census Category includes residents registered as unemployed, unpaid workers in family businesses and residents who did not disclose their employment status.			

9.2.4.1 The Nature of Income-Earning Experience (Sample versus Census)

With respect to the nature of participants' income earning experience, the raw count and percentages recorded by the sample are compared with the national figures, and also the Census figures available for the city of Christchurch alone, in Table 9.5.⁸

It is immediately clear that the proportions of self-employed⁹ (4.55%) and self-employed employers of others (7.58%) are fairly similar to the proportions of both the national population and the population of Christchurch aged fifteen years and over which fell into these two categories four years earlier. The sample contains a much higher percentage of employees (public and private sector combined); but then it was a recruitment requirement of the study that participants must be taxpayers with tax return-filing experience. This requirement evidently increased the proportion of employees and decreased the proportion of non-workers

⁸ The national data used in this table was computed from Table 28 of the 1991 Census of Population and Dwellings, National Summary, p. 41. The Christchurch data was computed from Table 11 of the summary for Canterbury in the Census' Regional Reports Volume 2. Table 9.5 is capable of providing only an imprecise comparison in that if apples and oranges are not being compared, at the very least the table presents two quite dissimilar varieties of apple. The participants were asked in Question 14 to tick the category which best described their major income-earning experience. This experience may be of a different nature from the income status enjoyed by the subjects at the time they participated in the study. The 1991 Census figures, on the other hand, recorded the formal income earning status of New Zealand residents applicable at the time the census was taken.

⁹ The self employed in this category do not employ others.

relative to the national levels. If the employee category is measured as a percentage of New Zealand residents, aged fifteen years and over and receiving an income from work, then this category's national figure rises to 69.42 percent. This is within five percent of the proportion of the sample which taxpayer-participants of this kind constitute. The category *not in the workforce* comprises household managers and participants such as students, who did not fit into any other category, but who, nevertheless, had some experience in filing tax returns. The much larger national percentage will have included New Zealand residents who lacked this experience.

Table 9.5 also contains another interesting item of information. The population and percentage figures in the third and fourth columns provide evidence that the taxpayers of Christchurch, the city where the study took place, possessed a profile quite close to that of the taxpayers of New Zealand as a whole.

9.2.4.2 Size of Household Income (Sample versus Census)

Table 9.6 contrasts the proportions of the sample by household income size with the corresponding proportions of the population resident in New Zealand aged fifteen years and over¹⁰.

The immediate trend apparent in Table 9.6 is that the participants in the study were, on average, wealthier than the population aged fifteen and over, at large. The trend remains consistent with respect to both the nation-wide figures in the third column, and to the Census data for Christchurch alone, which are listed in the fourth column.¹¹ The discrepancy between the income profile of the participants and the Christchurch and national profiles remained acute even when the income figures disclosed in Question 13 of the questionnaire were reduced either one or two classes in accordance with participants' answers to Question 12, which

¹⁰ The national data was computed from Table 19, "Total Income and Sex by Age Group for Population Resident in New Zealand Aged 15 Years and Over", published in the 1991 Census of Population and Dwellings, National Summary, p.33. The Christchurch data was computed from Table 7 of the summary for Canterbury in the Census' Regional Reports Volume 2.

¹¹ It is clear in Table 9.6 that the taxpayers of Christchurch do not differ to any great degree from the nation as a whole in the spread of their income sizes.

TABLE 9.6
Income Estimates (Full Sample N = 131)[†]

RAW COUNT	SAMPLE	ADUSTED[‡]	1991 CENSUS	CHRISTCHURCH
Up to \$20,000	23	33	1,573,746	146,940
\$20,001 - \$30,000	21	30	441,930	39,270
\$30,001 - \$40,000	29	25	222,540	18,300
\$40,001 - \$50,000	20	32	106,095	8,934
More than \$50,000	38	11	109,629	8,103
Total	131	131	2,590,287	231,057
PERCENTAGES	SAMPLE	ADUSTED[‡]	1991 CENSUS	CHRISTCHURCH
Up to \$20,000	17.56	25.19	60.76	63.59
\$20,001 - \$30,000	16.03	22.90	17.06	17.00
\$30,001 - \$40,000	22.14	19.08	8.59	7.92
\$40,001 - \$50,000	15.26	24.43	4.10	3.87
More than \$50,000	29.01	8.40	4.23	3.51
[†] The sample was reduced from 132 by the failure of one person to answer Questions 12 and 13. [‡] The response to Q13 (size of household income) was reduced by one or more income bracket(s) where the answer to Q12 (number of contributors) indicated that the Q13 response was a pooled figure.				

solicited the number of contributors to their household's income.¹² The most likely source of this discrepancy is, once more, the fact that recruits were required to be experienced tax return filers. The high proportion of '\$20,000 and under' residents in the 1991 national figures will have included the unemployed (and most of the nation's students) at a time when the country was in a deep recession. The New Zealand economy was recognised as enjoying a boom in 1995.

9.2.4.3 Religion and Altruism

There were several further differences between the sample and the taxpayers of the nation as a whole. Forty-five percent of the sample was recruited from the congregation of an Anglican church. Anglicans comprised 25.88 percent of the population of Christchurch and 21.70 percent of the total New Zealand population, although many of those who stated this affiliation in the 1991 Census doubtless were not sufficiently active in their faith to have necessarily had

¹² It is recognised that this measurement is purely done by rule of thumb, and as such, is of doubtful accuracy. Nevertheless, it does provide an unsophisticated insight into the difference between the sample and the New Zealand population (without controlling for demographic changes relating to improvement in the economy and the simple passage of time).

the same characteristics as the Anglicans recruited through their congregation membership for participation in this study.¹³

A related variable is even more imponderable. A common factor uniting all 132 volunteers was altruism. Every subject agreed to participate upon being approached with a fund-raising proposition that would benefit an organisation to which he or she belonged; and nobody went home personally enriched. It would have been interesting to find out if a sample of compulsorily drafted non-altruists would have recorded risk preferences of the same nature. However, it would be very difficult to measure with any accuracy what proportion of the taxpaying populace could be considered altruistic, and what proportion non-altruistic.

In the light of these considerations, the sample used in this study may be considered only approximately representative of New Zealand taxpayers. It continues to lack the validity conferred by recruitment via a random selection process.

9.3 THE IMPACT OF THE QUESTIONNAIRE VARIABLES

9.3.1 Introduction

In this Section the data obtained from the end-of-session questionnaire is used in a series of *repeated measures* GLM analysis of variance procedures in order to investigate any impact on the risk level selection process not traceable to the decision frame effect tested for in Dusenbury's (1994) hypotheses. As stated in the introduction to this chapter, the focus of interest is now narrowed to what influences participants' choice of risk level in the tax decision problems alone. Consequently, the dependent variables in the *repeated measures* procedures are primarily the three replicated tax scenarios, *refund* Tax Case **Y**, *high pay* Tax Case **X**, and *low pay* Tax Case **Z**. However, whenever a promising line of enquiry presents itself in the

¹³ The national totals for Anglicans were: 335,403 men and 396,642 women. The total population on the day of the Census was 3,373,926.

results from these procedures, this material is inserted into the *repeated measures* procedure used for testing H_{1A} .¹⁴

9.3.2 Category One: Demographic Variables

The demographic variables available to this study were elicited by the eleven factual questions discussed previously in Chapter Six, Subsection 6.9.2.¹⁵ It was possible to execute a *repeated measures* GLM analysis of variance procedure on the full sample incorporating all eleven question responses as categorical variables. However, an ANOVA table reporting all the possible interaction combinations of the eleven would be vast; so it was necessary to abridge the procedure by selecting only interaction effects which might plausibly occur. The most likely combinations were those listed in the following four categories:

1. Permutations of the three variables, *gender (Q1)*, *age (Q2)* and *education (Q15)*.
2. Permutations of the three variables *tax filing experience (Q3)*, *frequency of payment-due tax filings (Q9)* and *frequency of professional tax-filing assistance (Q17)*.
3. Permutations of the pair of variables, *household income (Q13)* and *nature of income earning experience (Q14)*.
4. *Household size (Q12)*¹⁶, *knowledge of Expect Utility Theory (Q18)* and *knowledge of Prospect Theory (Q19)* entered singly.

The *between-subjects* effects are reported in Table 9.7.

Of all of the main and interaction effects reported in Table 9.7, only the main effect associated with the variable, *gender (Q1)* achieves statistical significance. Located in the top row of the table, it has an *F*-statistic of 5.17 with an associated 0.0342 probability of a Type I error.

¹⁴ At this point the variable from the questionnaire was used as a partitioning device; and the *repeated measures* GLM procedure incorporated *refund* Tax Case Y responses (C1) and *high pay* Tax Case X responses (C2) as the dependent variables, and GROUP and CASEORD as the categorical variables.

¹⁵ The format of these questions is presented in Appendix B, Section B.4.

¹⁶ This variable becomes important only if a statistically significant effect associated with responses to Q13, *household income* is detected.

TABLE 9.7
The Three Tax Cases and the Eleven Demographic Variables.
Full Sample (N = 130)[†]

**General Linear Models Procedure Repeated Measures Analysis of Variance:
 Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q1	1	10.7894	10.7894	5.17	0.0342
Q2	5	4.5035	0.9007	0.43	0.8213
Q1*Q2	5	9.5636	1.9127	0.92	0.4905
Q15	4	5.0940	1.2735	0.61	0.6601
Q1*Q15	4	10.2412	2.5603	1.23	0.3311
Q2*Q15	10	12.4458	1.2445	0.60	0.7986
Q1*Q2*Q15	3	5.9187	1.9729	0.95	0.4376
Q3	4	8.3723	2.0931	1.00	0.4294
Q9	6	7.0002	1.1667	0.56	0.7576
Q3*Q9	12	25.4865	2.1238	1.02	0.4696
Q17	3	1.2971	0.4324	0.21	0.8903
Q3*Q17	5	8.9317	1.7863	0.86	0.5272
Q9*Q17	6	24.1418	4.0236	1.93	0.1256
Q3*Q9*Q17	2	3.7440	1.8720	0.90	0.4237
Q12	3	2.4303	0.8101	0.39	0.7628
Q13	4	9.7204	2.4301	1.16	0.3562
Q14	5	6.1760	1.2352	0.59	0.7064
Q13*Q14	11	12.6809	1.1528	0.55	0.8440
Q18	2	4.0104	2.0052	0.96	0.3996
Q19	1	3.5438	3.5438	1.70	0.2074
Error	20	41.7494	2.0875		

[†] The size of the sample was reduced from 132 to 130 by deletion of one participant who failed to supply an answer to Q12 and Q13, and one participant who did not answer Q17.

Table 9.8 reports the within-subject effects associated with the eleven factual variables with respect to the full sample. The first set of interaction effects, the permutations of *gender* (Q1), *age* (Q2) and *education* (Q15), had no significant *within-subject* effect on participants' choices of tax problem risk levels. This is conveyed by the low *F*-statistics in the first panel of Table 9.8.

Combinations of the tax-related factual variables reported in the second band of Table 9.8 were similarly insignificant. These three variables were *tax-filing experience in years* (Q3), *frequency of payment-due tax filings* (Q9) and *frequency of use of professional tax-filing assistance* (Q17).

TABLE 9.8:
The Three Tax Cases and the Eleven Demographic Variables.
Full Sample (N = 130)[†]

**General Linear Models Procedure Repeated Measures Analysis of Variance:
 Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	0.7490	0.3745	0.35	0.7034
CASE*Q1	2	1.6087	0.8043	0.76	0.4732
CASE*Q2	10	9.0591	0.9059	0.86	0.5776
CASE*Q1*Q2	10	10.2616	1.0262	0.97	0.4816
CASE*Q15	8	7.2312	0.9039	0.86	0.5600
CASE*Q1*Q15	8	11.9957	1.4995	1.42	0.2176
CASE*Q2*Q15	20	19.3914	0.9696	0.92	0.5683
CASE*Q1*Q2*Q15	6	5.3056	0.8843	0.84	0.5480
CASE*Q3	8	8.8989	1.1124	1.05	0.4135
CASE*Q9	12	18.0883	1.5074	1.43	0.1934
CASE*Q3*Q9	24	25.0092	1.0420	0.99	0.5012
CASE*Q17	6	6.4101	1.0684	1.01	0.4309
CASE*Q3*Q17	10	9.3540	0.9354	0.89	0.5534
CASE*Q9*Q17	12	5.2693	0.4391	0.42	0.9482
CASE*Q3*Q9*Q17	4	0.6103	0.1526	0.14	0.9643
CASE*Q12	6	4.1190	0.6865	0.65	0.6893
CASE*Q13	8	19.5502	2.4438	2.32	0.0382
CASE*Q14	10	19.0110	1.9011	1.80	0.0918
CASE*Q13*Q14	22	37.0639	1.6847	1.60	0.0975
CASE*Q18	4	5.9677	1.4919	1.41	0.2469
CASE*Q19	2	2.9979	1.4990	1.42	0.2535
Error(CASE)	40	42.2024	1.0551		

[†] The size of the sample was reduced from 132 to 130 by deletion of one participant who failed to supply an answer to Q12 and Q13, and one participant who did not answer Q17.

The next combination consisted of *size of household income (Q13)* and *nature of income-earning experience (Q14)* — in particular whether a participant was self-employed or a private sector or state sector employee. While the *earning experience* variable registered a *within-subject* interaction effect (CASE*Q14) at the 9 percent level of significance, the *income-size* variable produced a stronger, if small interaction impact (CASE*Q13), with a 3.82 level of significance, falling within the five percent benchmark. In terms of the Greenhouse-Geisser Epsilon, the probability of a Type I error associated with the CASE*Q13 interaction effect rose

to 0.0448. Nevertheless, the F -statistic reported for this impact is a very small 2.32.¹⁷ The three-way interaction effect associated with these variables was only weakly significant at the 9.75 percent level of significance.

There were no significant *within-subject* effects associated with the *household size* variable (Q12) or the two *risk theory knowledge* variables (Q18 and Q19).

A summary of the closer scrutiny afforded each of the combinations of factual variables in turn is given in the following subsections.

9.3.2.1 Gender (Q1), Age (Q2) and Education (Q15)

When just *gender*, *age*, and *educational level* alone were utilised as categorical variables in a *repeated measures* GLM procedure on the full sample, two effects were noticed. In the first instance, the *within-subject* main effect, CASE, became significant with an F -statistic of 3.64 supported with a 0.0282 probability of error; and a CASE*Q2*Q15 interaction effect was observed with an F -statistic of 1.64 associated with a probability of error of 0.0360 (rising to 0.0391 in terms of the Greenhouse-Geisser Epsilon). In the second instance, *gender* furnished a significant *between subjects* main effect with an F -statistic of 4.58 supported at the 3.5 percent level of significance.

When the three variables, *gender*, *age* and *education* were run on the low cash float subsample, the result was similar. *Gender* (Q1) provided a between-subjects main effect ($F = 4.89$, $\text{Pr} > F = 0.0334$); CASE was strongly significant, with an F -statistic of 7.48 with a 0.001 probability of a Type I error; and a CASE*Q2 (*age*) *within-subject* interaction effect became significant with an F -statistic of 2.66 with a 0.008 probability of error. It would appear that both *gender* and *age* have a potentially measurable impact on participants' choices of tax scenario risk level.

¹⁷ Information relating to this variable is printed in bold in the table.

However, when the three high cash float subsamples¹⁸ were investigated in terms of *gender*, *age* and *education*, no significant *within-subject* or *between-subject* effects were detected with probabilities of error within the $Pr > F = 0.05$ benchmark level. This finding mirrors the similarly insignificant findings with respect to these subsamples of participants reported in Chapter Eight.

With the above findings on the full sample and low cash float grouping in mind, the full sample was partitioned into male and female subsamples, so that the *repeated measures* GLM analysis of variance procedure used in the testing of H_{1A} in Chapter Eight, Subsection 8.5.2, could be re-run to determine if *gender* caused any difference in outcome.¹⁹ This partition yielded a male subsample of 49 participants and a female subsample containing 83 subjects. When the procedure was run on the male subsample, no statistically significant main or interaction effects were found; and CASE itself yielded a low *F*-statistic of 1.45 with a high probability of a Type I error, 0.2412. The null form of H_{1A} could not be rejected on the choices made by the men alone. However the female subsample provided evidence supporting the power of framing effects in that CASE alone was statistically significant, supported by an *F*-statistic of 6.55, with a 0.0143 probability of error.²⁰

When the procedure was applied to the low cash float subsample partitioned by *gender*, subsamples of 19 men and 47 women were obtained. In the instance of the low cash float male subsample, there was a *between-subjects* main effect emanating from GROUP with an *F*-statistic of 6.01 with a 0.0703 probability of a Type I error, which may have been a result of the composition of the school parent group.²¹ The low float male subsample furnished little evidence of an intrinsic framing effect. The *within-subjects* CASE variable in this instance had an *F*-statistic of 3.64 with an associated 0.1291 probability of error. On this evidence, H_{1A} could not be rejected.

¹⁸ \$3,500, \$4,200 and combined high cash float subsamples.

¹⁹ This involved partitioning the sample by *gender* and employing C1 and C2 as the dependent variables, and GROUP and CASEORD as the categorical variables.

²⁰ See Appendix G for the full tables of results for these tests.

²¹ This group contained 19 females and one male.

On the other hand, the low cash float female subsample provided evidence which strongly supported rejection of H_{1A} 's null form. The lone significant F -statistic belonging to CASE rose this time to 11.93, supported with a 0.0027 probability of error.

A possible explanation for the existence of a gender-based difference runs as follows. There were many more female than male volunteers, yet three of the four organisations supplying recruits were more likely to be supported by men and women in relationships (married or otherwise) than by single persons. This implies that it was the woman rather than the man of a household who responded to the plea for help in fundraising. This, in turn, implies that women, on average, (even when the men did turn up), treated more seriously the issue of earning a dollar return for their organisation. And this, in turn, suggests that the women put more thought into the decisions they made in the course of the experiment. If this argument is correct, the difference between the men and the women should not be viewed as a matter of basic differences in perceptions of risk, but as a difference in perceptions regarding the personal importance of fundraising. In other words, the women were more engaged in the experiment than the men. However, this argument is promoted as a possible explanation rather than as a certainty; but it is supported by the circumstantial evidence, observed in experimental sessions, of some women (but few men) working beyond the call of duty.

The full sample and low cash float subsample were also arbitrarily partitioned by *age* and submitted to the *repeated measures* GLM procedure used to test H_{1A} . Four *age* subsamples were created by cutting the full sample into an above-thirty subsample and a thirty-and-below group, and then an above-forty subsample and a forty-and-below group.²² The low cash float subsample was partitioned in the same fashion. When the H_{1A} *repeated measures* GLM procedure was run on these eight new subsamples, four of the eight provided evidence on which the null form of H_{1A} could be rejected; and in seven instances, either CASE was significant or no variable at all was statistically significant. In the eighth case, the subsample

²² A minor typographic error in Question 2 of the Questionnaire meant that the age 40 was included in two boxes. This did not have a significant impact on the results, as can be seen in the relative homogeneity of the results reported in the table.

AGE	FULL SAMPLE (N = 132)			LOW CASH FLOAT SUBSAMPLE (N = 66)		
	N	F-stat	Pr > F	N	F-stat	Pr > F
> 30	107	5.85	0.0182	56	8.64	0.0070
> 40	56	3.11	0.0873	28	4.56	0.0652
≤ 30	25	1.99	0.2179	10	Not able to be computed	
≤ 40	76	5.29	0.0279	38	21.85	0.0007

was too small for F -statistics to be computed. The F -statistics and their associated probabilities of error are laid out in Table 9.9.

From Table 9.9 it is apparent that the thirty-and-under age group were not strongly motivated by any decision frame effect, since their shift in risk level was either statistically insignificant or not computable. When the low age group subsamples were increased to encompass the forty-and-unders, H_{1A} could conclusively be rejected, especially in terms of the under-forties low cash float subsample. By contrast, the over-forties displayed a tendency to perform a risk shift; but it was not significant in either subsample within the five percent statistical benchmark. The over-thirties, like the forty- and-unders, provided evidence supporting the rejection of H_{1A} 's null form. It is noted from Subsection 9.2.3 that almost 39 percent of the 132 participants were within the 31-40 age range. The pattern discernible from the results in Table 9.9 is that the participants in the middle range of ages (of which 31 - 40 years could be considered the bulk) were susceptible to framing effects based on the scenarios, whereas the age groups at the extremes were more impervious.

9.3.2.2 Tax-Filing Experience (Q3), Frequency of Payment-Due Filings (Q9) and Frequency of Professional Assistance (Q17)

When a *repeated measures* GLM procedure was run on the three tax variables incorporating *tax-filing experience*, *frequency of payment-due filings*, and *frequency of professional assistance* as categorical variables, not one of the three categorical variables produced a statistically significant effect. This relationship held true for the procedure run on all cash float-related configurations of the sample of 132 participants.

With respect to *tax-filing experience* (Q3), this result differs from the finding made by Dusenbury (1994). Dusenbury found a main effect ($\text{Pr} > F = 0.0069$),²³ indicative of the more experienced taxpayers willingness to take higher risks in both the *high pay* and *refund* tax cases. In order to investigate this divergence between the results of the two studies further, *tax-filing experience* (Q3) was employed as a partitioning device rather than as a categorical variable.

When the *repeated measures* GLM procedure employed in the analysis of H_{1A} was partitioned by *tax-filing experience* into separate runs on the three subsamples 1 - 5 years experience, 6 - 15 years experience, and more than 15 years experience, the 1 - 5 years subsample produced a strong CASE effect ($F = 10.73$, $\text{Pr} > F = 0.0466$) and the over 15 years subsample produced a weaker CASE effect with about the same degree of statistical significance ($F = 4.24$, $\text{Pr} > F = 0.0447$). The 1 - 5 years subsample, which contained 17 participants, recorded a *refund* case (C1) mean of 1.9412 versus 2.2346 recorded by the over 15 years subsample ($N = 81$). However the 1 - 5 years subsample recorded a much higher mean level of response to the *high pay* case (C2). This was 3.2353 versus 2.7037 recorded by the over 15 years subsample. In other words, the less experienced taxpayers were more extreme in their risk preference shifts than were the more experienced taxpayers. The subsample of 34 participants with 6 - 15 years experience recorded an insignificant CASE statistic, indicative of very little difference between their means for the two cases (C1 = 2.6276, C2 = 2.765).²⁴

9.3.2.3 Size of Household Income (Q13) and Nature of Income-Earning Experience (Q14) Considered Together

This area promised to be quite interesting on the ground that in Table 9.8, a statistically significant CASE*Q13 within-subject interaction effect was noted. When *household income size* (Q13) and *nature of income earning experience* (Q14) were utilised as the sole categorical variables in a *repeated measures* GLM procedure on the three joint dependent variables

²³ Dusenbury (1994), op. cit. n. 1, p. 13.

²⁴ This investigation was conducted in terms of partitioning by three categories of responses from Question 3 only. No further testing was conducted in terms of the double partitioning required if levels of the cash float variable were to be taken into account. However, there is scope for such testing to be done in the future.

provided by Tax Cases X, Y and Z, this interaction effect became insignificant for every cash-float configuration of the sample.

In its place, another interesting phenomenon occurred. Recall that in Chapter Eight,²⁵ it was shown that, with respect to the three configurations of the high cash float subsample, H_{IA} 's null form could not be rejected on the evidence provided. The current procedure incorporating *income size* and *earning experience* produced a significant CASE effect for the \$4,200 cash float subsample. Its *F*-statistic, albeit low at 3.54, was associated with 0.0386 probability of error.²⁶

On the basis of this \$4,200 cash float subsample CASE effect discovery, the *repeated measures* procedure for testing H_{IA} was re-run on this grouping with *household income size* (Q13), incorporated along with SESSION and CASEORD as categorical variables. However, this investigation was unproductive, resulting in the disappearance of statistical significance on the part of all main and interaction effects in the model.²⁷ This result indicates that *household income size* did not significantly affect the risk preferences of the \$4,200 cash float participants in *refund* Tax Case Y and *high pay* Tax Case X.

Another avenue was also investigated with respect to *household income size* (Q13) and *income earning experience* (Q14). The 132 participants of the full sample were partitioned into groupings according to the *number of contributors to household income* (Q12). The three levels were single contributor, two contributors and 'other'.²⁸ The partitioned sample was then subjected to a *repeated measures* GLM procedure in which the dependent variables were the responses to Tax Cases X, Y and Z.

²⁵ Chapter Eight, Subsection 8.5.4.

²⁶ The combined high cash float subsample and the \$3,500 group continued, however, to return insignificant *F*-statistics.

²⁷ The *F*-statistic for CASE in this instance was 3.01 with a 12.63 percent probability of error.

²⁸ Generally speaking, the 20 participants who recorded 'other' as their response, were students and other young adults living in flats; and their behaviour was noticeably more akin to that of the study's single income earners than that of participants from households earning double incomes.

The variables, Q13 and Q14 did not furnish any statistically significant effects in terms of the 52 participants from single income households or the 59 participants from double income households; but the subsample of 20 participants from multiple income households produced a weak *between-subjects* Q13 effect ($F = 3.36$, $\text{Pr} > F = 0.0552$) and a strong *within-subject* CASE*Q14 effect ($F = 3.17$, $\text{Pr} > F = 0.0136$). However, it was not possible to pursue this investigation further in terms of the rather interesting \$4,200 cash float subsample (mentioned in the last paragraph) on the ground that the intersection between the multiple income household group and the \$4,200 cash float recipients was too small to be useful.²⁹

9.3.2.4 Nature of Income Earning Experience (Q14) Considered Alone

It was proposed in Chapter Six, Subsection 6.9.2, that self-employed taxpayers might possibly develop a greater propensity for risk taking as a result of their self-employment status. In order to test for any underlying difference in risk preferences existing between self-employed participants and participants who were employees, the full sample was also partitioned into three subsamples in terms of the responses participants made to Question 14. This also facilitated a probe for differences between public and private sector employees. Neither the self-employed nor the private sector employee participants recorded a significant movement in their risk preferences between *refund* Tax Case Y and *high pay* Tax Case X; but the public sector employees did furnish evidence of a significant shift. The *within-subject* results for all three subsamples are contained in Table 9.10.³⁰

An examination of the means of *refund* Tax Case Y and *high pay* Tax Case X revealed that the three types of participants did not materially differ in their risk preferences. This is shown in terms of the two sets of insignificant ANOVA F -statistics and Kruskal-Wallis Chi-Square approximations reported in Table 9.11.

²⁹ There were three eligible participants.

³⁰ None of the *between subjects* effects furnished by the partitioned *repeated measures* GLM procedure used in the production of the results in this table were found to be statistically significant.

TABLE 9.10

**Refund Tax Case Y and High Pay Tax Case X:
Effect of Partitioning by Income Earning Experience.
Full Sample (N = 113)[†]**

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

PANEL A

Public Sector Employees Subsample (N = 41)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	8.6863	8.6863	8.22	0.0112
CASE*GROUP	3	0.4911	0.1637	0.15	0.9250
CASE*CASEORD	15	14.9449	0.9963	0.94	0.5428
CASE*GROUP*CASEORD	6	8.2538	1.3756	1.30	0.3115
Error(CASE)	16	16.9000	1.05625		

PANEL B

Private Sector Employees Subsample (N = 56)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.8723	0.8723	1.80	0.1903
CASE*GROUP	3	0.2841	0.0947	0.20	0.8985
CASE*CASEORD	14	8.0480	0.5749	1.19	0.3368
CASE*GROUP*CASEORD	10	4.2155	0.4216	0.87	0.5700
Error(CASE)	28	13.5583	0.4842		

PANEL C

Self Employed Subsample (N = 16)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	4.0289	4.0289	3.06	0.1407
CASE*GROUP	3	2.5172	0.8391	0.64	0.6227
CASE*CASEORD	6	7.4256	1.2376	0.94	0.5381
CASE*GROUP*CASEORD	1	0.5714	0.5714	0.43	0.5392
Error(CASE)	5	6.5833	1.3167		

[†] The 19 participants who disclosed that they were household managers, or that they fell into the category, 'other', were excluded from this analysis.

The conclusion, which may be reached from these findings, is that the nature of the participants' income earning experience did not have an effect on their risk preferences. Hence the difference in compliance between self-employed taxpayers and waged (or salaried)

TABLE 9.11

**Effect of Source of Income on Individual Case Responses.
Full Sample (N = 113)[†]**

ANOVA for Refund Tax Case Y (C1) classified by Q14

NPARIWAY Procedure for Analysis of Variance [‡]

Q14	N	Mean	Among MS	Within MS
			0.3611	1.2747
Private Sector	56	2.3036		
Public Sector	41	2.1951	F Value	Prob > F
Self-Employed	16	2.4375	0.283	0.7538

NPARIWAY Procedure for Kruskal-Wallis Test [‡]

Kruskal-Wallis Test Chi-Square Approximation

CHISQ = 0.2184 DF = 2 Prob > CHISQ = 0.8965

ANOVA for High Pay Tax Case X (C2) Classified by Variable Q14

NPARIWAY Procedure for Analysis of Variance [‡]

Q14	N	Mean	Among MS	Within MS
			1.5260	1.5441
Private Sector	56	2.6250		
Public Sector	41	2.9756	F Value	Prob > F
Self-Employed	16	2.8750	0.988	0.3755

NPARIWAY Procedure for Kruskal-Wallis Test [‡]

Kruskal-Wallis Test Chi-Square Approximation

CHISQ = 1.8885 DF = 2 Prob > CHISQ = 0.3890

[†] The 19 participants who stated that they were home managers or 'other' have been excluded.

[‡] In all instances, average scores were used for ties.

taxpayers may be ascribed only to the opportunity levels commented on by Clotfelter (1983)³¹ and Wallschutzky (1984),³² and not to any employment status-related ideological change.

³¹ Clotfelter, C. T., (1983), "Tax Evasion and Tax Rates: An Analysis of Individual Returns", p. 368, also cited in Chapter Six, Subsection 6.9.2.

³² Wallschutzky, I. G., (1984), "Possible Causes of Tax Evasion", p. 381, also cited in Chapter Six, Subsection 6.9.2.

9.3.2.5 Knowledge of Expected Utility Theory (Q18) and Knowledge of Prospect Theory (Q19)

These were not investigated further on the ground they were shown to be insignificant in Table 9.8.³³

9.3.3 Category Two: Attitude Variables

Three questions elicited responses which were catalogued in Chapter Six, Subsection 6.9.3 as attitude variables. They divide into two categories — health issue attitudes, and attitude towards gambling. The two categories are investigated separately.

9.3.3.1 Health Issues Attitude: Questions 4 and 5

The participants' attitudes towards the holding of private medical insurance were shown in Chapter Eight, Section 8.6, to make the difference between rejection or non-rejection of the null form of H_{2A} when these attitudes were employed as a sample partitioning device. In the current subsection, attention is turned to the impact of the *medical insurance attitude* variable (Q4) and the *tax fairness attitude* variable (Q5) on all three tax cases.

When Q4 and Q5 were submitted jointly as classification variables in a *repeated measures* GLM analysis of variance procedure, in which the dependent variables were *high pay* Tax Case X, *refund* Tax Case Y and *low pay* Tax Case Z, only Q4 was found to be influential. While there were no *between-subjects* effects of statistical significance, a *within-subject* CASE*Q4 interaction effect, possessing an *F*-statistic of 2.27, supported with a 0.0236 probability of a Type I error, is reported in Panel B of Table 9.12.

This finding accords with the significant role of participants' attitudes towards private health insurance in determining their risk preferences with respect to the Medical Insurance Case; but since the dependent variables in this instance were the three sets of tax case responses alone,

³³ See Subsection 9.2.3.3. Only one participant had knowledge of the two theories.

TABLE 9.12

**The Three Tax Cases and Q4 and Q5.
Full Sample (N = 131)[†]**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q4	4	2.0709	0.5177	0.30	0.8757
Q5	4	11.4169	2.8542	1.67	0.1624
Q4*Q5	11	20.5417	1.8674	1.09	0.3747
Error	111	189.9130	1.7109		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	16.0913	8.0456	7.69	0.0006
CASE*Q4	8	19.0127	2.3766	2.27	0.0236
CASE*Q5	8	6.7758	0.8470	0.81	0.5948
CASE*Q4*Q5	22	28.5862	1.2994	1.24	0.2148
Error(CASE)	222	232.2815	1.0463		

Source	Greenhouse - Geisser Epsilon	Huynh - Feldt Epsilon
CASE	0.0007	0.0006
CASE*Q4	0.0247	0.0236
CASE*Q5	0.5923	0.5948
CASE*Q4*Q5	0.2170	0.2148

PANEL C

Variable	N	Mean	Std Dev	Min	Max
Q4	132	4.0455	1.1111	1	5
Q5	131	2.2672	1.2330	1	5

[†] The sample was reduced from 132 to 131 by the failure of one participant to answer Q5.

there is no intuitive explanation which immediately springs to mind to account for the interaction effect.

When the identical *repeated measures* GLM procedure was applied to the low cash float subsample, the result was similar. The CASE*Q4 *within-subject* interaction effect, this time, had an *F*-statistic of 2.32 with a 2.55 percent likelihood of error, which increased to 3.11 percent in terms of the Greenhouse-Geisser Epsilon.

However, both the low cash float subsample and the full sample failed to yield statistically significant results concerning the *tax fairness attitude* (with respect to public health sector changes) variable, (Q5).

9.3.3.2 Gambling Attitude: Question 11

The impact of the gambling attitude variable was investigated in Chapter Eight, Subsection 8.7.4 in terms of a *repeated measures* GLM analysis of variance procedure in which the dependent variables were *low pay* Tax Case Z and the Gamble Case. When the gambling attitude variable was used as a classification variable in a *repeated measures* procedure in which the dependent variables were Tax Cases X, Y and Z, the results obtained are set out in Table 9.13. It is clear in the table that the gambling attitude variable has no statistically significant *between-subjects* or *within-subject* effects. It is also interesting that the mean of the responses to Question 11 for the full sample was exactly three, which indicates strict neutrality, which may be interpreted as either having no fixed opinion, or being indifferent to the issue.

When the procedure was run on the low cash float sample, there was a similar absence of any statistically significant gambling attitude effect; but when the combined high cash float subsample was processed, a CASE*Q11 *within-subject* interaction effect with an *F*-statistic of 1.77 significant at the 9.01 percent level of significance occurred (9.54 in terms of the Greenhouse-Geisser Epsilon). In this instance, the mean of the responses to Question 11 dropped to 2.8939, which indicated that this body of 66 participants regarded gambling (as proxied by LOTTO) in a slightly negative light. However, this result remained outside the standard five percent benchmark for tolerance of possible error.

When the combined high cash float grouping was separated into its constituent \$3,500 and \$4,200 subsamples, the \$3,500 participants regarded gambling in a slightly favourable light (with a mean of 3.1667) and furnished no statistically significant gambling attitude variable effects, while the \$4,200 subjects produced a relatively more heavily anti-gambling mean of 2.5667. Even so, the gambling attitude variable was unable to produce statistically significant effects in terms of this subsample.

TABLE 9.13

**The Three Tax Cases and Q11.
Full Sample (N = 132)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q11	4	2.5080	0.6270	0.35	0.8421
Error	127	226.1182	1.7805		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	6.0650	3.0325	2.76	0.0653
CASE*Q11	8	10.3674	1.2959	1.18	0.3120
Error(CASE)	254	279.1730	1.0991		

Source	Greenhouse - Geisser Epsilon	Huynh - Feldt Epsilon
CASE	0.0673	0.0653
CASE*Q11	0.3131	0.3120

PANEL C

Variable	N	Mean	Std Dev	Min	Max
Q11	132	3.0000	1.0842	1	5

In terms of the full sample and each of its four cash float subsamples, the gambling attitude variable produced from responses to Question 11 was irrelevant in the selection of risk levels in the three replication tax decision problems.

9.3.3.3 Compliance Attitude: Questions 6, 7, 8 and 11

An analysis involving the three questions relating to participants' attitudes to tax compliance was undertaken in Chapter Eight, Subsection 8.5.5, in which the anomalous behaviour of the high cash float subsamples with respect to H_{1A} was investigated.³⁴ Later in the same chapter, the responses collected from these questions were used in a further *repeated measures* GLM procedure to throw light on why the \$4,200 cash float subsample provided evidence which was

³⁴ Questions 6 and 7 were also discussed in Chapter Six, Subsection 6.9.3.

TABLE 9.14					
The Three Tax Cases and Q6, Q7 and Q8.					
Full Sample (N = 132)					
PANEL A					
General Linear Models Procedure Repeated Measures Analysis of Variance:					
Tests of Hypotheses for Between Subjects Effects					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q6	4	3.3672	0.8418	0.49	0.7460
Q7	4	2.8165	0.7041	0.41	0.8036
Error	119	206.0997	1.7319		
PANEL B					
General Linear Models Procedure Repeated Measures Analysis of Variance:					
Univariate Tests of Hypotheses for Within Subject Effects					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	11.6590	5.8295	5.41	0.0050
CASE*Q6	8	10.4225	1.3028	1.21	0.2941
CASE*Q7	8	8.5260	1.0658	0.99	0.4450
CASE*Q8	8	8.3630	1.0454	0.97	0.4598
Error(CASE)	238	256.3935	1.0773		
Source	Greenhouse - Geisser Epsilon		Huynh - Feldt Epsilon		
CASE	0.0054		0.0050		
CASE*Q6	0.2950		0.2941		
CASE*Q7	0.4441		0.4450		
CASE*Q8	0.4588		0.4598		
PANEL C					
Response Variable	N	Mean	Std Dev		
Refund Tax Case Y (C1)	132	2.2955	1.0753		
High pay Tax Case X (C2)	132	2.7879	1.2358		
Low pay Tax Case Z (C4)	132	2.2197	1.1278		
Q6	132	1.9848	0.9961		
Q7	132	2.4394	1.3265		
Q8	132	2.7424	1.2578		

too insubstantial to reject the null form of H_{3A} .³⁵ In the current chapter, the analysis is extended to cover all three of the tax replication decision problems.

Table 9.14 contains the statistical output recorded from the full sample. There are no statistically significant *between-subjects* or *within-subject* effects relating to these compliance attitude variables.

³⁵ This was in Chapter Eight, Subsection 8.2.6.4.

The absence of statistically significant output in the Table 9.14 provides immediate answers to two of the three questions asked in Chapter Six, Subsection 6.9.3:

1. Is there a connection between the implied fairness variable contained in Question 6 and propensity for risk willingness over the tax scenarios?

From the table, the answer is no.

2. Is there a connection between the implied fairness variable contained in Question 7 and propensity for risk willingness over the tax scenarios?

Again, from the table, the answer is no. Nevertheless, the means reported in Panel C of Table 9.14 indicate that the sample, as a whole, had a more strongly negative attitude towards non-compliant taxpaying behaviour on the part of high income earners (Q6) than on the part of low income earners (Q7), while feeling just a little bit negative about paying its own tax bills (Q8). The slight negativity detected in terms of the responses to Question 8 in Panel C of Table 9.14 and the insignificance of any Q8 effects in Panels A and B deny any importance to the confounding variable, *patriotic duty*, posited in Chapter Six, Subsection 6.9.3.

When the procedure was run on the low cash float subsample, a similar pattern was observed. As for the full sample, the CASE variable *within-subject* main effect (with an *F*-statistic slightly higher at 5.57) was significant at the half of one percent level of significance, while none of the compliance attitude variables offered a significant explanation.

When the procedure was run on the combined high cash float subsample and its constituent \$3,500 and \$4,200 groupings, none of the variables produced significant *F*-statistics at all.

However, a further, more complex research question concerning the influence of the implied fairness variables in Questions 6 and 7 was asked in Chapter Six, Subsection 6.9.3:

3. Do people who choose the *never* box in both Question 6 and Question 7 exhibit total risk aversion with respect to the tax scenarios? If so, what inference may be drawn from this? Is there any clear connection between these people and those who have indicated that they disapprove of gambling in their response to Question 11?

TABLE 9.15

**Means of Tax Cases X, Y and Z:
Subsamples Determined by Strongly Disapproving Responses to Q6, Q7 and Q11.**

PANEL A

Subsample of respondents who chose $Q6 = Q7 = 1$.

Response Variable	N	Mean
<i>Refund</i> Tax Case Y (C1)	43	2.1628
<i>High Pay</i> Tax Case X (C2)	43	2.4419
<i>Low Pay</i> Tax Case Z (C4)	43	1.9302

PANEL B

Subsample of respondents who chose $Q11 \leq 2$.

Response Variable	N	Mean
<i>Refund</i> Tax Case Y (C1)	63	2.2063
<i>High Pay</i> Tax Case X (C2)	63	3.0000
<i>Low Pay</i> Tax Case Z (C4)	63	2.1746

The answer to the first part of this question is contained in the left-hand columns of Table 9.15. It is clear from the divergence of the means of the Tax Cases X, Y and Z from unity and from each other, that total risk aversion has not been exhibited.

The second part of this question involved examining responses to Question 11. Sixty-three people indicated, by ticking this question's first or second box, that they personally would not indulge in gambling.³⁶ Their means, also displayed in Table 9.15, are higher than the means for the $Q6 = Q7 = 1$ subsample. Simple inspection of these means alone indicates no relationship of any interest can be detected between the risk profiles of the people who chose to make either $Q6 = Q7 = 1$ responses and the risks taken by the participants who made $Q11 \leq 2$ responses.³⁷

³⁶ This question was couched in terms of whether or not the participants would buy tickets in the national weekly Lotto draw. It is noted that these people were, nevertheless, prepared to complete the *Choose a Gamble* decision problem in this study.

³⁷ It was possible, but not profitable, to investigate these relationships exhaustively. However, it was interesting to note that there were only two participants in the subsample determined by the $Q6 = Q7 = Q11 = 1$ response; and their risk levels with respect to the three cases were: both 15% for *refund* Tax Case Y, 33% and 40% in *high pay* Tax Case X, and no risk and 40% in *low pay* Tax Case Z. Quite plainly there was no uniformity in these responses.

TABLE 9.16
Three Tax Cases and Q10 and Q16.
Full Sample (N = 128)[†]

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
 Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q10	16	18.0200	1.1262	0.62	0.8621
Q16	4	8.9286	2.2321	1.23	0.3035
Error	107	194.5438	1.8182		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
 Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	3.0102	1.5051	1.31	0.2726
CASE*Q10	32	29.6845	0.9276	0.81	0.7627
CASE*Q16	8	4.9613	0.6202	0.54	0.8265
Error(CASE)	214	246.3222	1.1510		

Source	Greenhouse - Geisser Epsilon	Huynh - Feldt Epsilon
CASE	0.2723	0.2726
CASE*Q10	0.7584	0.7627
CASE*Q16	0.8211	0.8265

PANEL C

Variable	N	Mean	Std Dev	Min	Max
Q10	128	5.6484	4.3735	1	21 [‡]
Q16	132	2.9015	0.7702	1	5

[†] The sample was reduced from 132 to 128 by the removal of 4 participants who did not provide responses to Q10.

[‡] This question entailed a 21-point Likert scale, each point representing a 5% increase in audit probability.

Consequently, it may be concluded that the participants' declared attitudes to compliance had no significant influence on the risk levels they chose in the three decision problems, Tax Cases X, Y and Z.

9.3.4 Category Three: Perception Variables relating to Taxpayer and IRD Behaviour

This subsection provides results relating to the two questions soliciting information on participants' assumptions as taxpayers, explained in Subsection 6.9.4 of Chapter Six.

In terms of the full sample, neither of these assumption variables had a statistically significant effect on the participants' choices of risk level over the three tax cases, as reported in Table 9.16.

With minor changes to the means of the responses reported for the questions, the result obtained when the procedure was re-run on the low cash float subsample was similar. The two assumption variables produced neither *between-subjects* nor *within-subject* effects of any note. However a sole *within-subjects* interaction effect was detected when the procedure was repeated on the combined high cash float subsample. In this instance, the CASE*Q10 interaction impact yielded an *F*-statistic of 1.46, significant at the 0.0916 probability of a Type I error, which, adjusted in terms of the Greenhouse-Geisser Epsilon rose only to 0.0976.³⁸

Conversely, when the procedure was used on the \$3,500 cash float subsample, it was Q16 (general compliance assumption) which produced a main effect close to gaining significant status at the five percent level of significance. This *between-subjects* effect had an *F*-statistic of 3.02 with a 0.0552 probability of error. Meanwhile, the Q10 *within-subjects* effect observed in the case of the combined high float grouping disappeared with respect to this subsample. However, both of these anomalies, while providing further insight into why the high cash float subsamples furnished results for H_{1A} and H_{3A} contrary to Dusenbury's (1994) findings and to those for the full and low float groupings of this current study, may nevertheless be dismissed at the five percent error-level cut-off for statistical significance. The two assumption variables furnished no statistically significant effects with respect to the \$4,200 cash float subsample.

Effectively these two variables, *belief in taxpayer compliance* (Q16) and *assumed IRD audit rate* (Q10) had no impact on the participants' behaviour. With respect to Q10, this result is in keeping with the results of Warneryd and Walerud (1982), who found no significant connection between non-compliance and the perceived rate of detection.³⁹ This Q10 finding also offers little support for Allingham and Sandmo's (1972) claim that perceived detection rate is a major

³⁸ There was no change at all in terms of the Huynh-Feldt adjustment.

³⁹ See Chapter Six, Subsection 6.9.4.

tool in the hands of tax authorities for ensuring taxpayer compliance.⁴⁰ However, it is noted that the mean risk preference levels detected in this study were always lower than the mean perception of audit frequency disclosed by the participants.

In the next chapter, Chapter Ten, the extension hypotheses for this study are investigated.

⁴⁰ See Chapter Six, Subsection 6.9.4 and Chapter Three, Section 3.2.

10. RESULTS OF EXTENSION HYPOTHESES

10.1 INTRODUCTION

We now move on from the results associated with the conceptual replication of Dusenbury (1994) to cover the findings made with respect to the four extension hypotheses. In its layout, this chapter follows the order in which these extensions were discussed in Chapter Four, Section 4.5. The first to be considered is the Cash Flow Extension in Section 10.2; after which, attention is turned to the Summary Syndrome Extension in Section 10.3. The findings associated with the Risk Profile Stability Extension are then laid out in Section 10.4, and finally, in Section 10.5, the results associated with the Value Function Extension are reported.

10.2 CASH FLOW EXTENSION

This extension was undertaken on the strength of a comment Dusenbury (1994) made about the availability to participants of liquid cash reserves in his study (\$3,000 per person in \$100-denominated stock cards). He noted that risk taking may have been diminished as a consequence of his participants' funds always remaining positive.¹ The aim in the New Zealand experiment was to issue two levels of cash float, one of which might cause subjects to know they had a tougher task than their fellow participants to end their participation with a positive residue. The associated research question was, would differential cash floats make any difference to the subjects' propensities for risk?

The evidence accumulated during the testing of the first three hypotheses suggests that the level of cash float had a singularly noticeable impact on participants' decisions. This was detected when the level of cash float was used as a sample partitioning device. You may recall that H_{1A} 's null form could be rejected in terms of the full sample and the low cash float subsample,

¹ Dusenbury, R., (1994) "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-filing Options", p. 13. (This point was cited in Dusenbury's words at the start of Chapter Four, Subsection 4.5.1.)

but not in terms of any configuration of the high cash float subsample. In this extension, the issue of cash flow influence is formally investigated with respect to the full sample unpartitioned. The wording of the cash flow hypothesis, H_{4A} , is repeated here from Chapter Four:

H_{4A} : Participants with a potentially inadequate cash flow will choose riskier options in the *high pay* case than chosen by participants with high cash flow.

A procedural adjustment during the running of the experimental sessions brought the number of float levels to three, of which participants were aware only of two — a high and a low float.² All three levels of the cash flow variable were taken into account in the analysis.

The method for testing H_{4A} was determined by several factors. The first of these was that the observations on the *high pay* Tax Case X variable, C2, were arguably ordinal in nature; and the second was that the testing involved a comparison of independent samples of different sizes, in which a normality and homoscedasticity could not be assumed. The Wilcoxon rank sum test (also known as the Mann-Whitney U test)³ provided an appropriate nonparametric tool for making this comparison.⁴ This test was available, along with the Kruskal-Wallis test in the SAS software programme's NPAR1WAY procedure. This procedure also produced an analysis of variance table.⁵ The NPAR1WAY results are printed in full for the first comparison; but thereafter, only the salient statistics are cited in a tabular summary.

When the NPAR1WAY procedure was performed with respect to the impact of CASH on the Tax Case X risk choices of participants with either the low (\$2,100) cash float or the \$3,500 cash float, a statistically significant difference was found between the two sets of Case X responses. The low cash float participants adopted a higher mean risk level, as shown in Panel

² This issue was discussed in Chapter Four in footnote 35.

³ SAS/STAT User's Guide (Version Six, Fourth Edition, 1990) Volume 1, p. 127.

⁴ Emory, C. W. and Cooper, D. R., (1991) *Business Research Methods* (Fourth Edition), p.574. Emory and Cooper state that the Mann-Whitney U test is an alternative to the *t* test without the *t* test's limiting assumptions, for instance, the normal distribution requirement.

⁵ SAS/STAT User's Guide (Version Six, Fourth Edition, 1990) Volume 2, p.1202.

TABLE 10.1

**Cash Flow Effect on Responses to High Pay Tax Case X:
Low Cash Float and \$3,500 Cash Float.
(NPARIWAY PROCEDURE)**

PANEL A

Analysis of Variance for High Pay Tax Case X Classified by CASH⁶:

CASH	N	Mean	Among MS	Within MS
\$2,100	66	2.9545	6.0612	1.4375
\$3,500	36	2.4444	F Value	Prob > F
			4.216	0.0426

PANEL B

Wilcoxon Scores (Rank Sums) for Tax Case X Classified by Variable CASH⁷:

CASH	Sum of N	Expected Scores	Std Dev Under H0	Mean Under H0	Score
\$2,100	66	3686.0	3399.0	137.8695	55.8485
\$3,500	36	1567.0	1854.0	137.8695	43.5278

Wilcoxon 2-Sample Test (Normal Approximation)
(with Continuity Correction of .5)

$$S = 1567.00 \quad Z = -2.07805 \quad \text{Prob} > |Z| = 0.0377$$

$$\text{T-Test approx. Significance} = 0.0402$$

Kruskal-Wallis Test (Chi-Square Approximation)

$$\text{CHISQ} = 4.3334 \quad \text{DF} = 1 \quad \text{Prob} > \text{CHISQ} = 0.0374$$

A of Table 10.1. The fact this mean risk level was significantly different from the \$3,500 cash float subsample's mean risk level is made clear by the F-statistic of 4.216 supported with a 0.0426 probability of a Type I error. This result is also supported in Panel B by the Wilcoxon Z-score of -2.078 with an associated 0.0377 probability of error; and the Kruskal-Wallis Chi-square approximation of 4.333 with a 0.0374 probability of error.

On the basis of the difference between the low cash float group (\$2,100) and \$3,500 cash float group responses, the null form of H_{4A} may be rejected. The opposite is true in terms of every

⁶ Average scores were used for ties.

⁷ Ibid.

TABLE 10.2					
Summary of NPAR1WAY Procedure Results for all other Comparisons.					
ANOVA		WILCOXON		KRUSKAL-WALLIS	
Low (\$2,100) and Combined High (\geq \$3,500) Cash Float Subsamples (N=66, N=66)					
F-Statistic	Pr > F	Z-Statistic	Pr > Z 	Chi-Square	Pr > Chi-Square
2.427	0.1217	-1.6145	0.1060	2.6205	0.1055
Low (\$2,100) and \$4,200 Cash Float Subsamples (N=66, N=30)					
F-Statistic	Pr > F	Z-Statistic	Pr > Z 	Chi-Square	Pr > Chi-Square
0.196	0.6587	-0.46205	0.6441	0.21726	0.6411
\$3,500 and \$4,200 Cash Float Subsamples (N=36, N=30)					
F-Statistic	Pr > F	Z-Statistic	Pr > Z 	Chi-Square	Pr > Chi-Square
1.599	0.2106	1.22358	0.2211	1.5135	0.2186
All Independent Groupings: \$2,100, \$3,500 and \$4,200 Cash float Subsamples					
F-statistic	Pr > F	Z-Statistic	Pr > Z 	Chi-Square	Pr > Chi-Square
2.043	0.1339	The NPAR1WAY procedure did not compute this statistic for a three-way comparison.		4.2645	0.1186

other configuration of the sample, as is apparent in the total dearth of statistically significant results in Summary Table 10.2.

In the light of this finding, a further set of NPAR1WAY procedures were executed, incorporating, as a further option, the asymptotic Kolmogorov-Smirnov two-sample test. The statistics produced by the Kolmogorov-Smirnov test indicated that any differences in the choice of risk brought about by the level of cash provided at the start of an experimental session were entirely insignificant. The two-sample Kolmogorov-Smirnov statistics, the associated asymptotic distributions and the asymptotic probabilities of error are summarised in Table 10.3:

On the basis of this evidence, the null form of H_{4A} cannot be rejected. There are no significant differences amongst the mean risk levels chosen in response to *high pay* Tax Case X by the various subsamples of participants partitioned by cash float.⁸ Instead, it would seem that the

⁸ A similar set of tests were performed with respect to *refund* Tax Case Y. Again, all differences in risk levels were found to be statistically insignificant.

TABLE 10.3
Kolmogorov-Smirnov Two-Sample Test Results (Asymptotic).

PAIRS OF SUBSAMPLES TESTED	D	KSa	Pr > KSa
Low (\$2,100) and \$3,500 Cash Float Subsamples	0.189	0.914	0.374
Low (\$2,100) and \$4,200 Cash Float Subsamples	0.076	0.344	0.999
Low (\$2,100) and Combined High Cash Float Subsamples	0.121	0.693	0.717
\$3,500 and \$4,200 Cash Float Subsamples	0.150	0.607	0.855

most likely explanation of the cash flow variable's impact in the study was as proffered in Chapter Eight, Subsection 8.5.6.

This explanation hinged on the fact that every participant was aware he or she had one of two possible cash floats - a low one of \$2,100 or a high one of \$3,500;⁹ and that those who had the high float were perhaps a little more relaxed about their ability to perform in the experiment in such a way that would bring a positive dollar return to their organisation. On the other hand, the subjects who knew they had the low float may have felt they had to pay closer attention to what they were doing if they were to avoid reducing the return their organisation was potentially going to receive from its volunteers' cumulative efforts. This argument, however, does not necessarily contradict the findings of Kachelmeier and Shehata (1992),¹⁰ since their discovery of a significant impact on Chinese university students' risk preferences related to extremely high real cash pay-offs, while the rate of conversion of residual pseudo-dollar funds in the current study was set at a very modest NZ\$0.30 per \$100 (pseudo) for all participants.

10.3 SUMMARY SYNDROME EXTENSION

The rationale for the summary syndrome extension was given in Chapter Four, Subsection 4.5.2. In brief, it entailed a desire to investigate the salience of the scenario story-lines and the

⁹ This was kept at \$4,200 for the church subsample; and as the groups by recruitment origin attended separate sessions, the \$4,200 recipients would not have been aware of the existence of the \$3,500 high cash float afforded participants recruited from all other organisations.

¹⁰ Kachelmeier, S. J. and Shehata, M., (1992), "Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China", pp. 1120 - 1141. This study was cited in the more general discussion of issues concerning monetary incentives in Chapter Five, Subsection 5.4.2.

information they contained as a result of feedback received when the experimental instrument was pretested. The important issue at stake was that, if participants ignored the information provided in the scenario story-lines and made decisions based solely on the figures provided in the summary tables (with, perhaps, a passing glance at the scenario title in bold at the top of the page), then any impact attributable to the level of tax withholdings will have disappeared. This possibility arises because information on these withholdings appeared *only* within the text of the story-lines and *not* in the tables. Hence, if participants made decisions based on the summary tables and scenario titles alone, these might have given rise to evidence of a decision frame effect; but this effect would not have been based on a neutral reference point determined by prior tax monies paid. The hypothesis (repeated from Chapter Four) was worded:

H_{5A}: Participants who claim to have ignored withheld tax data in the *high pay* and *refund* tax cases will make filing decisions in which a non-withholdings-related differential decision frame effect is discernible.

Several approaches were used in investigating this hypothesis. In the first instance, a partitioned *repeated measures* GLM analysis of variance procedure was employed; and in the second, the variables measuring story-line influence were incorporated as categorical variables in both *basic model* and *repeated measures* GLM procedures.

10.3.1 Partitioning by Levels of Declared Influence.

This approach involved re-running the *repeated measures* GLM procedure used in testing **H_{1A}** with the data available set (after adjustments for missing observations), partitioned by the two variables representing participants' declared usage of story-line data. These variables were INFLY (the declared influence of withholding data contained in the Scenario Y story-line) and INFLX (the declared influence of withholding data from the Scenario X story-line).¹¹ Since there were five levels of observation for INFLX and another five for INFLY, five sets of *repeated measures* GLM procedure results were obtained with respect to each of these variables. As reported in Chapter Eight, this procedure incorporated GROUP and CASEORD

¹¹ These variables were explained in Chapter Four, Subsection 4.5.2.

as categorical variables, and used C1 (the *refund* Tax Case Y response variable) and C2 (the response variable from *high pay* Tax Case X) as its dependent variables.¹²

Partitioned *repeated measures* GLM procedures were run only on the full sample of 132 participants for two reasons. The full sample produced results favouring rejection of H_{1A} in Chapter Eight; furthermore it was a large enough grouping to sustain a five-way partitioning exercise. If participants claimed to have ignored withholding data from the scenarios (coded as INFLX = 1 and INFLY = 1), the production of a statistically significant CASE variable would be sufficient to show that H_{5A} 's null form should be rejected, since this result implies that these people undertook a risk-level shift independent of the decision frame-induced shift posited by Dusenbury. A weaker version of the same argument applies to participants who claimed to have been influenced in their thinking "only in a slight fashion", and whose responses were coded INFLX = 2, and INFLY = 2.

The summary results (restricted to CASE) of two partitioned *repeated measures* GLM procedures run on the full sample are summarised in Table 10.4. The statistical output obtained in the procedure partitioned by INFLX (reported in Panel A) was very clean. It provided evidence on which H_{5A} 's null form could be rejected. The statistical significance of CASE was greater for the partitions containing participants who ignored or cursorily used information given on tax withholdings. In fact, no significant difference was reported between responses to the two decision problems in every partition in which subjects claimed to have been influenced by the story-line information on withholdings.¹³

The clear picture obtained in Panel A did not carry over into Panel B, in which the partitioning was according to the variable, INFLY. In Panel B the results were inconclusive. Only one partition reported a significant CASE statistic; but it was associated with significant

¹² Note that CASEORD and GROUP are explained in Appendix D.

¹³ The procedure also produced no significant effects emanating from the categorical variables, GROUP and CASEORD.

TABLE 10.4			
Refund Tax Case Y and High Pay Tax Case X: CASE Statistics by Partition.			
(Repeated Measures General Linear Model Analysis of Variance Procedures)			
PANEL A			
Impact of Partitions by Case X Story-line Influence Variable (INFLX)			
Coding Level [†]	N	F-statistic	Pr > F
1	32	4.31	0.0676
2	24	11.98	0.0180
3	38	0.03	0.8552
4	27	2.19	0.1828
5	10	0.01	0.9177
PANEL B			
Impact on CASE Effect of Partitions by CaseY Story-line Influence Variable (INFLY)			
Coding Level	N	F-statistic	Pr > F
1	32	2.49	0.1404
2	27	21.15	0.0018
3	45	0.27	0.6117
4	19	2.76	0.1954
5	6	[Insufficient Degrees of Freedom]	
† While a CASE statistic was produced for this coding level, the partition was nevertheless too small to generate a full set of statistics.			

CASE*GROUP and CASE*CASEORD interaction effects. In all other INFLY partitions, no significant effects were detected. The reason for these INFLY-related results is not clear.

10.3.2 Declared Influence Incorporated as a Categorical Variable

The second approach to testing H_{5A} involved employing the participants' story-line tax-withholding cognisance as categorical variables in a series of *basic model* and *repeated measures* GLM analysis of variance procedures in which *refund* Tax Case Y responses (C1) and *high pay* Tax Case X responses (C2) were the dependent variables.

In order to simplify computational requirements and interpretation of the results, the five levels of INFLX and INFLY were condensed into two levels. The first level coalesced the participants who ignored the story-line withholding data with those who claimed to have heeded it "only in a slight fashion". The second level incorporated the three cognisance-paying levels of the variables ("significant", "quite important", and "totally shaped my thinking").

TABLE 10.5
Summary Table Syndrome Impact on Choice of Risk Levels:
Refund Tax Case Y and High Pay Tax Case X Considered Separately.
Full Model (N = 129)[†]

PANEL A

Basic Model GLM Analysis of Variance Procedure:[‡]
Dependent Variable: Refund Tax Case Y (C1)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	6.8797	2.2932	2.17	0.1051
INFLX	1	0.2449	0.2449	0.23	0.6326
GROUP*INFLX	1	0.0418	0.0418	0.04	0.8432
INFLY	1	0.8362	0.8362	0.79	0.3786
GROUP*INFLY	1	0.0048	0.0048	0.00	0.9468
INFLX*INFLY	1	0.8199	0.8199	0.78	0.3832
CASEORD	18	17.3021	0.9612	0.91	0.5715
GROUP*CASEORD	21	24.6868	1.1756	1.11	0.3716
INFLX*CASEORD	7	11.9994	1.7142	1.62	0.1544
GROUP*INFLX*CASEORD	1	0.6176	0.6176	0.58	0.4487
INFLY*CASEORD	4	2.8459	0.7115	0.67	0.6141
INFLX*INFLY*CASEORD	1	0.0429	0.0429	0.04	0.8413

PANEL B

Basic Model GLM Analysis of Variance Procedure:
Dependent Variable: High Pay Tax Case X (C2)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	12.0685	4.0228	2.60	0.0641
INFLX	1	1.8447	1.8447	1.19	0.2809
GROUP*INFLX	1	1.1478	1.1478	0.74	0.3938
INFLY	1	0.4092	0.4092	0.26	0.6097
GROUP*INFLY	1	0.0048	0.0048	0.00	0.9560
INFLX*INFLY	1	0.1054	0.1054	0.07	0.7954
CASEORD	18	23.4390	1.3022	0.84	0.6449
GROUP*CASEORD	21	24.2500	1.1548	0.75	0.7629
INFLX*CASEORD	7	15.6724	2.2389	1.45	0.2115
GROUP*INFLX*CASEORD	1	0.4048	0.4048	0.26	0.6116
INFLY*CASEORD	4	2.8036	0.7009	0.45	0.7697
INFLX*INFLY*CASEORD	1	5.8333	5.8333	3.77	0.0586

[†] Three participants were dropped from the sample because they failed to answer one of the two relevant additional ratings questions.

[‡] The procedure was also unable to compute statistics for the following three interaction effects as insufficient degrees of freedom were available: GROUP*INFLX*INFLY, GROUP*INFLY*CASEORD, GROUP*INFLX*INFLY*CASEORD

The first application of this approach involved running the *basic model* GLM procedure on the full sample with the response to the *refund* Tax Case Y decision problem as the dependent variable. The results are reported in Panel A of Table 10.5. The universal statistical insignificance of the main and interaction effects reported here implies that, whether or not participants paid attention to information given about tax withholdings, that level of attentiveness was irrelevant to the risk level chosen with respect to the *refund* Tax Case Y decision problem. This alone does not provide sufficient evidence for accepting or rejecting

H_{5A} ; but when it is combined with evidence furnished in the second run of the procedure, this time on *high pay* Tax Case **X**, sufficient evidence is amassed for a rejection of H_{5A} 's null form. Panel B of Table 10.5 contains the relevant output for this decision to reject H_{5A} .

In Panel B, only two effects achieve even weak statistical significance (GROUP and INFLX*INFLY*CASEORD). These probabilities remain, however, outside the five percent threshold. The conclusion to be drawn from the irrelevance of INFLX and INFLY to risk levels chosen in both decision problems is that the tax withholding information in the story-lines was redundant. This conclusion is further supported and clarified by the information in Table 10.6.

Table 10.6 presents the results of a *repeated measures* GLM procedure utilising both Tax Case **Y** and **X** response variables as dependent variables. In Panel B, the response-difference variable created in the procedure, CASE, has an *F*-statistic of 6.90, significant at slightly more than the one percent level of significance. No other computed *within-subject* effects come within an order of magnitude of the accepted five percent threshold for statistical significance. This implies there has been a significant shift in risk levels between the two decision problems, which in turn, suggests in *Prospect Theory* terms, the existence of a decision frame with a neutral reference point. It would seem that this reference point was not necessarily determined, in participants' minds, by the level of PAYE or provisional tax withholdings.

However, a strong *between-subjects* GROUP effect is also reported in Panel A of Table 10.6. The fact that this has appeared only in conjunction with the inclusion of INFLX and INFLY in the *repeated measures* GLM procedure possibly suggests that one group by recruitment origin, to a greater degree than the others, discounted more heavily the story-lines' prior withholdings information.

TABLE 10.6
Summary Syndrome Impact on Choice of Risk Levels:
Refund Tax Case Y and High Pay Tax Case X Considered Jointly.
Full Sample (N = 129)[†]

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects[‡]

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	17.6655	5.8885	4.31	0.0095
INFLX	1	1.7169	1.7169	1.26	0.2684
GROUP*INFLX	1	0.8139	0.8139	0.60	0.4444
INFLY	1	1.2076	1.2076	0.88	0.3523
GROUP*INFLY	1	0.0000	0.0000	0.00	1.0000
INFLX*INFLY	1	0.1687	0.1687	0.12	0.7270
CASEORD	18	24.9477	1.3860	1.01	0.4638
GROUP*CASEORD	21	34.2436	1.6306	1.19	0.3029
INFLX*CASEORD	7	24.5718	3.5103	2.57	0.0261
GROUP*INFLX*CASEORD	1	1.0112	1.0112	0.74	0.3943
INFLY*CASEORD	4	2.1626	0.5407	0.40	0.8107
INFLX*INFLY*CASEORD	1	3.4381	3.4381	2.52	0.1199
Error	44	60.1310	1.3666		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects^{††}

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	8.5438	8.5438	6.90	0.0118
CASE*GROUP	3	1.2827	0.4276	0.35	0.7926
CASE*INFLX	1	0.3727	0.3727	0.30	0.5860
CASE*GROUP*INFLX	1	0.3757	0.3757	0.30	0.5845
CASE*INFLY	1	0.0377	0.0377	0.03	0.8622
CASE*GROUP*INFLY	1	0.0095	0.0095	0.01	0.9305
CASE*INFLX*INFLY	1	0.7566	0.7566	0.61	0.4385
CASE*CASEORD	18	15.7934	0.8774	0.71	0.7837
CASE*GROUP*CASEORD	21	14.6932	0.6997	0.57	0.9204
CASE*INFLX*CASEORD	7	3.1000	0.4429	0.36	0.9216
CASE*GROUP*INFLX*CASEORD	1	0.0112	0.0112	0.01	0.9246
CASE*INFLY*CASEORD	4	3.4869	0.8717	0.70	0.5933
CASE*INFLX*INFLY*CASEORD	1	2.4381	2.4381	1.97	0.1675
Error(CASE)	44	54.4643	1.2378		

[†] The sample was reduced from 132 to 129 by the removal of three participants who did not provide answers to one of the two relevant additional ratings of judgment questions.

[‡] The procedure was unable to compute statistics for the following three *between-subjects* interaction effects as insufficient degrees of freedom were available: GROUP*INFLX*INFLY, GROUP*INFLY*CASEORD, GROUP*INFLX*INFLY*CASEORD

^{††} The procedure was unable to compute statistics for the following three *within-subject* interaction effects as insufficient degrees of freedom were available: CASE*GROUP*INFLX*INFLY, CASE*GROUP*INFLY*CASEORD, CASE*GROUP*INFLX*INFLY*CASEORD

When an identical repeated measures GLM procedure was run on the low cash float subsample, this *between-subjects* GROUP effect lapsed into statistical insignificance. In this instance, only one significant *F*-statistic was reported. This was for CASE, and it was $F = 10.89$, supported at the 0.0036 probability of error. This provides firm grounds for rejecting the null form of H_{5A} .

It is beyond doubt that there is a valid decision frame effect discernible in the study; but that there is more than one decision frame effect in the study, is also beyond doubt.¹⁴ Furthermore, this second frame is not anchored to a neutral reference point determined by prior withholdings. It has functioned, therefore, as a confounding variable in this study — and is likely to have done so in Dusenbury's (1994) study as well.

10.4 RISK PROFILE STABILITY EXTENSION

This extension was made possible by the fact that the sets of options available in *high pay* Tax Case X and *high pay* Tax Case W were identical in the experimental scripts used by the first 83 participants. The research question, for which an investigation was made possible by this turn of events, was, how stable are people's risk choices? Or, in similar situations, will they make choices in a consistent manner?

The fact that the participants were presented with the five scenarios associated with the replication study and the two extra scenarios (*refund* Tax Case V and *high pay* Tax Case W) all in a single package, meant that any variable associated with passage of time could not be incorporated; but the impact of influences from *other* decision problems could be examined. This opportunity exists since the participants made their choices with respect to the Tax Case W scenario after having processed almost every other case; and the randomisation of the ordering of the five replication study scenarios¹⁵ followed by three general questions and then

¹⁴ Participants' responses to *low pay* Tax Case Z (C4) and the Gamble Case (C5) were also assessed in terms of their INFLX level. In this instance an NPARIWAY procedure calculating ANOVA and Kruskal-Wallis statistics was employed. With respect to the participants' *low pay* Tax Case Z risk levels (C4), there was a weakly significant difference amongst the means of the (C4) subsamples as partitioned by levels of INFLX ($F = 2.24$, $\text{Prob} > F = 0.0683$; Kruskal-Wallis Chi-square approximation = 7.9716, $\text{Prob} > \text{Chi-sq} > 0.0926$). Interestingly, the subjects who claimed to have ignored the story-line information in *high pay* Tax Case X (INFLX = 1, INFLX = 2) recorded the two lowest Tax Case Z (C4) means.

INFLX	N	Mean
1	32	1.7813
2	24	2.2083
3	38	2.2368
4	27	2.6296
5	10	2.4000

This result suggests that reading the story-line in Tax Case Z had a moderating effect on risk preferences in this instance. However, this line of thought is predicated on the assumption that participants approached Tax Case Z in the same manner as they approached Tax Case X. When responses to the Gamble Case (C5) were partitioned in the same way, no significant differences were found in the five C5 means ($F = 1.04$, $\text{Prob} > F = 0.3876$; Kruskal-Wallis chi-square = 3.10, $\text{Prob} > \text{Chi-sq} = 0.5412$).

¹⁵ With the exception that *refund* Tax Case Y was always presented first.

by the final two scenarios (of which Tax Case **W** was one) divided into two orders, guaranteed a good mix from which inter-scenario influences could be detected. The results obtained from running procedures to test the two hypotheses tabled for this extension are discussed separately.

10.4.1 Results Relating to H_{6A}

This hypothesis, repeated from Chapter Four, Subsection 4.5.3 was:

H_{6A} : The risk preference exhibited by participants in the *high pay* case, Tax Case **X**, will be different from the risk preference exhibited by the same participants in the *high pay* case, Tax Case **W**, (First 83 scripts).

When a *repeated measures* GLM analysis of variance procedure was applied to the two high pay tax cases with respect to a subsample containing all 83 participants, no risk level instability was detected. This provides no grounds for rejecting the null form of H_{6A} . The results of the procedure are reported in full in Table 10.7.

The only statistic of note in the table is in Panel B, where the ordering of Case **W** and *refund* Tax Case **V** was shown to be weakly significant in terms of a *within-subject* CASE*ORDVW interaction, which had an F-statistic of 3.43 with an associated 0.0715 probability of a Type I error. This *within-subject* effect represents the impact of receiving the Tax Case **V** refund (after a string of payment filings for other decision problems) immediately before the Case **W** payment filing, or not having this infusion of pseudo-dollars before having to make the filing. It implies that the participants were aware of how much (or little) pseudo-dollar liquidity they still had; and that this did have a weak influence on their Case **W** choice.

To test the strength of this end-of-session liquidity awareness effect, the size of the participants' residual pseudo-dollar balances was added as a categorical variable (named HOARD) to the *repeated measures* GLM procedure. No statistically significant HOARD-

TABLE 10.7

**Stability of Risk Profiles in High Pay Tax Cases X and W.
All Available Participants (N = 83)**

PANEL A

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	1	4.0912	4.0912	1.83	0.1835
CASEORD	17	25.6558	1.5092	0.68	0.8067
GROUP*CASEORD	8	13.4836	1.6854	0.75	0.6436
ORDVW	1	0.0871	0.0871	0.04	0.8445
GROUP*ORDVW	1	2.2578	2.2578	1.01	0.3207
CASEORD*ORDVW	11	39.7622	3.6147	1.62	0.1305
GROUP*CASEORD*ORDVW	1	0.0078	0.0078	0.00	0.9531
Error	40	89.3155	2.2329		

PANEL B

**General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.2902	0.2902	0.41	0.5273
CASE*GROUP	1	0.1036	0.1036	0.15	0.7053
CASE*CASEORD	17	12.9258	0.7603	1.07	0.4173
CASE*GROUP*CASEORD	8	8.6107	1.0763	1.51	0.1852
CASE*ORDVW	1	2.4473	2.4473	3.43	0.0715
CASE*GROUP*ORDVW	1	1.3203	1.3203	1.85	0.1814
CASE*CASEORD*ORDVW	11	6.0415	0.5492	0.77	0.6674
CASE*GROUP*CASEORD*ORDVW	1	0.1953	0.1953	0.27	0.6038
Error(CASE)	40	28.5536	0.7138		

related impacts were found.¹⁶ Therefore the end-of-session liquidity awareness factor can be ignored.

H_{5A} was also investigated in terms of the 83 subjects divided into subsamples by the level of cash float. The low cash float subsample contained 33 participants, all but four of whom were members of the church group.¹⁷ The *repeated measures* GLM procedure in this instance

¹⁶ In terms of between-subjects effects, HOARD had an F -statistic of 2.6 with an associated error, $Pr > F = 0.2348$. Two reported interaction effects were as follows: $CASE*HOARD$ $F = 0.52$ $Pr > F = 0.8523$; $CASE*CASEORD*HOARD$ $F = 0.72$ $Pr > F = 0.6848$. Because there were 35 levels of observation on HOARD, there were insufficient degrees of freedom for the *repeated measures* GLM procedure to calculate results for the *within-subject* effects, $CASE*ORDVW*HOARD$ and $CASE*CASEORD*ORDVW*HOARD$; but it is assumed the results for these would have been similar if they had been computable. A *basic model* GLM procedure run on *high pay* Tax Case W (C7) with HOARD, CASEORD and ORDVW as categorical variables furnished a HOARD effect, $F = 1.13$, $Pr > F = 0.5411$.

¹⁷ A weakly significant between-subjects GROUP effect ($Pr > F = 0.0603$) may be ascribed to this distribution by recruitment origin. Note that information concerning the low cash float subsample is not reported in a table. The salient features are disclosed in the text alone in this instance.

repeated the finding made in terms of the 83-participant subsample. The F -statistic produced for CASE was a statistically insignificant 0.45 ($\text{Pr} > F = 0.5230$). In terms of the combined high cash float grouping (which contained 50 participants), a significant CASE effect was detected ($F = 4.86$, $\text{Pr} > F = 0.0424$); but in this instance, a number of *within-subject* interaction effects were also statistically significant.¹⁸ This finding was somewhat anomalous in that the \$3,500 cash float subsample (20 participants) furnished no figures of statistical significance when the identical procedure was performed upon it;¹⁹ and the \$4,200 subsample (30 participants) produced a strongly significant *within-subject* interaction effect, CASE*CASEORD*ORDVW ($F = 7.71$, $\text{Pr} > F = 0.0048$)²⁰ but no significant CASE result.

On the basis of the results described above, there was no noteworthy difference in the risk levels chosen in *high pay* Tax Case X and *high pay* Tax Case W. In all subsamples constructed from of the available pool of eligible participants (with the exception of the combined high cash float grouping), the repeated measures procedure induced variable, CASE, was insignificant; while the results from the anomalous subsample were noise-ridden and indeterminate. Hence the null from of H_{6A} cannot be rejected.

10.4.2 Results Relating to H_{7A}

This hypothesis allowed for a direct comparison to be made between *refund* Tax Case Y and *high pay* Tax Case W in order to test the assumption that the body of responses accumulated with respect to Tax Case W did indeed have the same properties as the responses accumulated with respect to *high pay* Tax Case X. The wording of the hypothesis was:

H_{7A} : Participants will choose riskier options in *high pay* Tax Case W than in *refund* Tax Case Y.

¹⁸ These *within-subject* interaction effects were: CASE*CASEORD ($F = 2.72$, $\text{Pr} > F = 0.0267$); CASE*GROUP*CASEORD ($F = 7.62$, $\text{Pr} > F = 0.0012$) and CASE*CASEORD*ORDVW ($F = 4.98$, $\text{Pr} > F = 0.0061$)

¹⁹ However, the *within-subjects* interactions, CASE*CASEORD ($F = 3.56$, $\text{Pr} > F = 0.0866$), and CASE*ORDVW ($F = 6.24$, $\text{Pr} > F = 0.0546$) were weakly significant in that they were close to, but above, the 5 percent benchmark for allowable error.

²⁰ The influence of the two case order variables was tested on the full subsample ($N = 83$) and found to be unimportant. Therefore it was not investigated for a second time.

TABLE 10.8

Stability of Risk Preference Shifts in Refund Tax Case Y and High Pay Tax Case W.
All Available Participants (N = 83)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	1	6.7002	6.7002	3.78	0.0588
CASEORD	17	14.0997	0.8294	0.47	0.9532
GROUP*CASEORD	8	6.9728	0.8716	0.49	0.8545
ORDVW	1	1.0154	1.0154	0.57	0.4534
GROUP*ORDVW	1	2.8203	2.8203	1.59	0.2143
CASEORD*ORDVW	11	11.6042	1.0549	0.60	0.8209
GROUP*CASEORD*ORDVW	1	0.3828	0.3828	0.22	0.6445
Error	40	70.8393	1.7710		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	9.3594	9.3594	12.64	0.0010
CASE*GROUP	1	0.0595	0.0595	0.08	0.7783
CASE*CASEORD	17	22.6007	1.3295	1.80	0.0644
CASE*GROUP*CASEORD	8	9.2117	1.1515	1.55	0.1696
CASE*ORDVW	1	0.7255	0.7255	0.98	0.3282
CASE*GROUP*ORDVW	1	0.9453	0.9453	1.28	0.2653
CASE*CASEORD*ORDVW	11	10.7523	0.9775	1.32	0.2495
CASE*GROUP*CASEORD*ORDVW	1	0.0078	0.0078	0.01	0.9187
Error(CASE)	40	29.6250	0.7406		

PANEL C

Response Variable	Mean
Refund Tax Case Y (C1)	2.3976
High Pay Tax Case W (C7)	2.8916

The fact that the relationship posited to exist in the alternative form of H_{7A} was indeed found to be in evidence may be read in the top line of data provided in Panel B of Table 10.8.²¹ The created variable, CASE, has an *F*-statistic of 12.64 which is very strongly supported by a probability of error of 0.001.

Two other results are of some interest in Table 10.8. A *between-subjects* influence emanating from GROUP and a *within-subject* CASE*CASEORD interaction effect are both weakly

²¹ This table reports the results of a *repeated measures* GLM procedure incorporating the responses to Cases Y and W as its dependent variables.

significant, albeit outside the five percent benchmark for tolerance of error. The case-order interaction effect is likely to be related to the end-of-session liquidity effect discussed in Subsection 10.4.1. The GROUP *between-subjects* effect is likely to be related to the unbalanced proportions of the participants by cash float in the sample. This was also touched upon in Subsection 10.4.1, and therefore will not be pursued a second time.

When the *repeated measures* GLM procedure was performed on the subsample containing the 50 eligible participants who had been given low cash floats, the influence of a decision frame-related difference existing between the two tax cases was again found to be relatively strong and statistically significant. In this instance, CASE had an F -statistic of 9.37 significant at the three quarters of one percent level of significance, providing grounds for a confident rejection of the null form of H_{7A} .²² It is clear then, that when situations are similar and the range of choices are identical, decision makers will be consistent in their choices; and that this consistency applies with respect to tax compliance.

10.5 VALUE FUNCTION EXTENSION

At the end of each of the first five decision problems the participants were required to disclose the order in which they ranked the options from first to last (or least favoured) choice. This extension involves observing these stated preference orders to ascertain if a smooth value function underlay the selection of a unique option by these people in each decision problem. It was explained in Chapter Four, Subsection 4.5.4, that a variable would be constructed from the participants' rankings which would contain two levels of observations of the smoothness of their value functions with respect to each scenario. This two-level variable will be called CURVE. A simple count of CURVE's levels revealed the information which is contained in Table 10.9.

²² Nevertheless, there was one minor anomaly. While the *between-subjects* GROUP effect noted with respect to the procedure run on the full sample disappeared, a weakly significant *within-subjects* CASE*GROUP interaction effect was detected ($F = 3.99$, $Pr > F = 0.0630$). The results generated from running the procedure on the high cash float subsamples are not reported on the ground that when H_{1A} was tested, these subsamples returned results which did not allow for rejection of the null hypothesis. As a consequence, nothing of interest would be gained from a repetition in terms of H_{7A} .

TABLE 10.9					
Smoothness Data for Tax Cases X, Y and Z, the Medical Case and the Gamble Case.					
Full Sample (N = 131)[†]					
PANEL A					
Smoothness Count (Variable = CURVE)					
Level	Case Y	Case X	Medical	Case Z	Gamble
1	122	117	116	117	93
2	9	14	15	14	38
Sum	131	131	131	131	131
PANEL B					
Frequency of Level (Variable = CURVE)					
Level	Case Y	Case X	Medical	Case Z	Gamble
1	93.13%	89.31%	88.55%	89.31%	70.99%
2	6.87%	10.69%	11.45%	10.69%	29.01%
Sum %	100.00%	100.00%	100.00%	100.00%	100.00%
[†] The sample was reduced from 132 to 131 because one participant did not disclose his order of preferences.					

While it is clear from Table 10.9 that not every participant operated on the basis of a smooth value function, the high frequency of level 1 observations makes it clear that such a function can be considered the norm. The three tax cases possess reasonably homogeneous frequencies of smoothness in the vicinity of 90 percent; but while the Medical Insurance Case produced results of a similar nature, there is a notable reduction in level 1 observations in the Gamble Case. This pattern is repeated in terms of the magnitude of the discontinuities, which are reported in terms of a magnitude variable called MAG in Table 10.10.

Not only has the Gamble Case provoked more departures from smoothness, but it is clear in Panel B that a larger proportion of the Gamble Case discontinuities are of the maximum size.

The size of the maximum observation level of MAG in the Gamble Case made it worth exploring this case further. In six of the eleven instances in which a level 4 discontinuity was observed, participants had adopted the preference ordering, **A - E - D - C - B**. These people apparently decided that the best course of action was to take no risk; but if risk was obligatory (and it was made so by the requirement to fill out the preference order question), then they might as well adopt an order which maximised the possible positive outcome given that the negative outcomes remained constant at a payment of \$500 (pseudo-dollars). This may be

TABLE 10.10					
Smoothness Data for Tax Cases X, Y and Z, the Medical Case and the Gamble Case.					
Full Sample (N = 131)[†]					
PANEL A					
Magnitude of Discontinuity (Variable = MAG)					
Level	Case Y	Case X	Medical	Case Z	Gamble
1	122	117	116	117	93
2	5	9	9	11	20
3	0	2	4	1	7
4	4	3	2	2	11
Sum	131	131	131	131	131
PANEL B					
Frequency of Magnitudes (Variable = MAG)					
Level	Case Y	Case X	Medical	Case Z	Gamble
1	93.13%	89.31%	88.55%	89.31%	70.99%
2	3.82%	6.87%	6.87%	8.40%	15.27%
3	0.00%	1.53%	3.05%	0.76%	5.34%
4	3.05%	2.29%	1.53%	1.53%	8.40%
Sum	100.00%	100.00%	100.00%	100.00%	100.00%
[†] The sample was reduced from 132 to 131 because one participant did not disclose his order of preferences.					

interpreted as a partial adoption of a smooth value function.²³ Two further participants of the remaining five jumped from **A** (certain payment of \$200) to **E** (40 percent chance of a payment of \$500 and a 60 percent chance of win of \$200), but followed inconsistent patterns after that initial jump; and the three other participants chose **E** as their first option and **A** as their second, which indicated a rather random approach to risk level selection.²⁴

At this point the focus is narrowed to contemplation of the first of the value function hypotheses, which was:

H_{8A}: Subjects will *not* display a consistent incidence of smooth value functions across all three Dusenbury-related tax-filing decision problems.

²³ These six participants returned a mixture of responses to Question 11 (Attitude towards Gambling). Three said they were quite happy to have a flutter in Lotto (but would not miss it if it were banned). Only one said that gambling was always wrong; and one other said that supporting a charity was acceptable, but regular ticket buying was not right for him. The sixth participant in the category opted to give no opinion either way with respect to gambling (as represented by the purchase of Lotto tickets).

²⁴ Four of this latter grouping of five participants said, in response to Question 11 (Feelings about Gambling), that they were quite happy to have a flutter (but would not miss it if it were banned); and the fifth said he considered that regular purchases of Lotto tickets were not something that was right for him, although he was happy to support a charity.

TABLE 10.11
Tax Cases X, Y, Z, the Medical Insurance Case and the Gamble:
Smoothness of Value Function.
Full Sample (N = 131)[†]

PANEL A

NPARIWAY Procedure Analysis of Variance for Variable CURVE
Classified by Variable CASENO

CASENO	N	Mean	Among MS	Within MS
1 (Y)	131	1.0687	0.0636	0.0856
2 (X)	131	1.1069	F Value	Prob > F
4 (Z)	131	1.1069	0.743	0.4763

[Average Scores were used for Ties]

PANEL B

NPARIWAY Procedure Wilcoxon Scores (Rank Sums) for Variable CURVE
Classified by Variable CASENO

CASENO	N	Sum of Scores	Expected Under H ₀	Std Dev Under H ₀	Mean Score
1 (Y)	131	25152.0000	25807.0000	536.9565	192.0000
2 (X)	131	26134.5000	25807.0000	536.9565	199.5000
4 (Z)	131	26134.5000	25807.0000	536.9565	199.5000

[Average Scores were used for Ties]

Kruskal-Wallis Test (Chi-Square Approximation)

CHISQ = 1.4880 DF = 2 Prob > CHISQ = 0.4752

[†] The sample was reduced from 132 to 131 by the removal of one participant who did not disclose his order of preferences.

When the SAS NPARIWAY analysis of variance procedure was applied to the observations on smoothness afforded by Tax Cases X, Y and Z, the ANOVA and Kruskal-Wallis test results reported in Table 10.11 were obtained. Both the *F*-statistic of 0.743 with its associated 0.4763 probability of error reported in Panel A and the Kruskal-Wallis Chi-square approximation reported in the bottom line of Panel B, strongly suggest that no evidence exists on which the null form of H_{8A} may be rejected.

Instead, the existence of consistently smooth value functions is clearly supported in terms of the three tax decision problem preference orderings. This accords with the value function concept espoused by *Prospect Theory* and explained in Chapter Two, Subsection 2.4.2.2. Further confirmation of this conclusion may be found in the means information contained in Panel A. All three means are closely bunched at a level insignificantly different from 1.

TABLE 10.12

**Tax Cases X, Y, Z, the Medical Insurance Case and the Gamble:
Smoothness of Value Function.
Full Sample (N = 131)[†]**

PANEL A

NPARIWAY Procedure: Analysis of Variance for Variable CURVE

Classified by Variable CASENO:

CASENO	N	Mean	Among MS	Within MS
			0.9962	0.1133
1 (Y)	131	1.0687		
2 (X)	131	1.1069	F Value	Prob > F
3 (Medical)	131	1.1145	8.792	0.0001
4 (Z)	131	1.1069		
5 (Gamble)	131	1.2901		

[Average Scores were used for Ties]

PANEL B

Wilcoxon Scores (Rank Sums) for Variable CURVE

Classified by Variable CASENO:

CASENO	N	Sum of Scores	Expected Under H ₀	Std Dev Under H ₀	Mean Score
1 (Y)	131	40020.5000	42968.0000	1155.1221	305.5000
2 (X)	131	41658.0000	42968.0000	1155.1221	318.0000
3 (Medical)	131	41985.5000	42968.0000	1155.1221	320.5000
4 (Z)	131	41658.0000	42968.0000	1155.1221	318.0000
5 (Gamble)	131	49518.0000	42968.0000	1155.1221	378.0000

[Average Scores were used for Ties]

Kruskal-Wallis Test (Chi-Square Approximation)

CHISQ = 33.568 DF = 4 Prob > CHISQ = 0.0001

† The sample was reduced from 132 to 131 by the removal of one participant who did not disclose his order of preferences.

When the focus is broadened to encompass the Medical Insurance Case and the Gamble Case as well as the three tax cases, quite a different picture emerges, as reported in Table 10.12.

In Table 10.12, both the ANOVA statistic ($F = 8.792$, $\text{Pr} > F = 0.0001$) and the Kruskal-Wallis Chi-square approximation ($CHISQ = 33.568$, $\text{Pr} > CHISQ = 0.0001$) provide grounds for firmly rejecting the null form of H_{0A} , which in its alternative form (repeated from Chapter Four, Subsection 4.5.4) was:

H_{0A} : Subjects will *not* display a consistent incidence of smooth value functions across all five Dusenbury-related decision problems.

This finding indicates that not all of the attributes of the *Prospect Theory* value function may be held to be universally and strictly applicable. This being so, there is scope for the development of a more all-embracing value function concept within the *Prospect Theory* paradigm. In addition, since the H_{9A} result also conflicts with the smoothness attributed to *Expected Utility Theory's* corresponding concave function, there is scope for such a development within decision theory in general, with respect to the making of risky choices.

The conclusions that have been drawn from the material tabled in this and the preceding two chapters are reported in Chapter Eleven. Chapter Eleven also contains an assessment of the study's limitations and a few suggestions concerning further useful extensions.

11. CONCLUSIONS, LIMITATIONS AND EXTENSIONS

11.1 INTRODUCTION

There are four sections in this chapter. The first of these, Section 11.2, contains the conclusions which may be drawn from the results posted in Chapters Eight, Nine and Ten, after which, Section 11.3 provides some insight into the study's limitations. The third section, Section 11.4, then traverses a number of possible extensions; and the final section, Section 11.5, comments on the implications for tax compliance policy and future tax research in New Zealand inherent in this study's findings.

11.2 CONCLUSIONS

To the extent that this study replicates Dusenbury (1994), strong evidence was produced showing that the New Zealand-participants behaved in a similar manner to their North American counterparts in Dusenbury's study. With respect to Dusenbury's first hypothesis, which was also H_{1A} in this study, the 132 New Zealand participants, when observed as an unpartitioned sample, chose risk levels which were relatively more risk averse when a tax refund was in the offing, and risk levels which were relatively more risk seeking when a tax payment was required. This shift towards greater risk willingness equates with more than a 19 percent drop in the portion of the uncertain tax item included in taxable income declared between *refund* Tax Case Y and *high pay* Tax Case X, which is comparable with a 21 percent drop from Dusenbury's figures.¹ The evidence allowed rejection of the null form of Hypothesis H_{1A} . Since H_{1A} codified the central research question in both Dusenbury (1994) and the current study, this finding indicates that the two studies are essentially congruent.

¹ Dusenbury, R., (1994), "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options", p. 11. This percentage was calculated from the figures recorded in Panel B of Table 2.

Investigation of all of the ensuing hypotheses (H_{2A} to H_{9A}) took the form of providing corroboration for, and greater understanding of, the answer obtained to this central research question. Each hypothesis is considered in turn, with respect to the full sample, in the following paragraphs, with the exception of H_{4A} . Only after this territory has been traversed is the complicating influence of the three discrete cash float levels discussed. H_{4A} is reviewed in conjunction with this discussion.

The study's 132 participants, when the sample was partitioned only in terms of their attitude towards the holding of private medical insurance, provided evidence allowing a qualified rejection of the null form of H_{2A} — that participants will be relatively more risk averse when choosing among health insurance options than when choosing among financially identical terminal tax payment options.

The reason for the qualification was an underlying cultural difference detected as a result of the subjects in the New Zealand study being asked to disclose their attitude towards health insurance. New Zealand, in the past, has maintained a public health system, which was believed, by the bulk of its population, to meet the medical needs of all New Zealanders. By contrast, the United States of America has never provided its citizens with universal access to high quality federal- or state-funded medical care. Consequently North Americans may be expected to take out medical insurance as a matter of course, whereas this has not been generally true of New Zealanders.

However, in the last ten years there has been a change. Today there is a general perception in New Zealand that the public health system is no longer capable of fulfilling its traditional role; and that citizens who can afford to do so, should plan ahead to cover the cost of their own health crises. However, the New Zealand psyche, is at present in a state of transition, with part of the population adopting an American way of thinking, and the rest still maintaining faith in the existing public health system's ability to cater for them if the need arises.²

² Further discussion of this complex issue, with its many economic and political ramifications, is beyond the scope of this study, as is any proper analysis of changes in the New Zealand public health system. The only aspect relevant to this study is the

The study found that participants who value private medical insurance behaved similarly to their North American counterparts in adopting a relatively greater level of risk aversion in their response to the Medical Insurance Case than in their response to *high pay* Tax Case X. However, those who rejected the need for private health insurance responded to the Medical Insurance Case by taking on a greater level of risk in this area. With this cultural difference taken into account, the current study concurs with the finding on H_{2A} of Dusenbury (1994).

In terms of the unpartitioned full sample of 132 subjects, the New Zealand study also concurred with Dusenbury (1994) in finding that the taxpayer-participants were significantly more risk willing in their responses to the Gamble Case than they were when responding to a tax case in which the risks and returns were identical, option for option. In both studies, the null form of H_{3A} , could therefore be rejected. This also provides further endorsement of Baldry (1986), who detected this systematic difference when comparing tax evasion and gambling in the *Expected Utility Theory* paradigm.³

On the surface, these findings (in terms of the undifferentiated full sample) apparently also support the conclusions reached by Dusenbury (1994). The most important of these was that the prepayment position stated in the tax scenarios was influential in determining the participants' risk preferences.⁴ However, this could not be confirmed in the current study. It was found that a minority of the New Zealand taxpayer-participants, who stated that they ignored the information given in the story-line of *high pay* Tax Case X, still recorded a significant swing from relative risk aversion in the face of prospective gains to greater risk willingness when making a decision involving choices amongst prospective losses. This occurred in spite of the failure of these participants, in the words of Hite, Jackson and Spicer (1988), "to put themselves in the role".⁵

public perception mentioned in the text, and solicited in Questions 4 and 5 of the questionnaire for the purpose of evaluating H_{2A} .

³ Baldry, J. C., (1986), "Tax Evasion is not a Gamble".

⁴ Dusenbury (1994), op. cit. n. 1, p. 13.

⁵ Hite, P. S., Jackson, B. R. and Spicer, M. W., (1988), "The Effects of Decision Framing on Taxpayer Compliance", p. 12. This was cited in Chapter Three, Subsection 3.3.3 and in Chapter Four, Subsection 4.5.2.

This finding was formalised in the rejection of the null form of H_{5A} , which, in its alternative form, stated that participants would act in accordance with a framing effect not ascribable to tax withholdings information. This concurs with Carroll (1992), who argued that numbers of decision frames could co-exist with respect to one decision problem.⁶

However, a further question arises. What decision frame did the participants who claimed to have ignored story-line information actually use? The answer must be that a simple gain or loss frame (as posited by *Prospect Theory*) was implicit in the figures disclosed in each decision problem's summary table alone. However, this alternative framing effect was not studied exhaustively across all of the available decision problems. The participants were asked to disclose their treatment of the information given about prior withholdings in *refund* Tax Case Y and in *high pay* Tax Case X only. The treatment of Tax Case Y information appeared to be irrelevant to participants' risk preferences; but dismissal of Tax Case X prior withholdings information was associated with a strengthening of the statistical significance of the risk preference shift detected between these two decision problems. This suggests that the information supplied about withholdings and other related matters (if diligently contemplated by the participants), dampened rather than exacerbated the risk preference shift which consideration of the raw figures in the summary table would otherwise have ordained.

On the other hand, the participants who ignored the Tax Case X withholding figure made responses to the Gamble Case which were not significantly different from those made by participants who heeded the Tax Case X withholding figure. This indicates that both types of participant were motivated to respond in a similar fashion in a putatively *contextless* situation. The mean risk level adopted in response to the Gamble Case by those who ignored the Tax Case X withholdings figure was, nevertheless, still almost eight percent higher than this subsample's mean Tax Case X risk level.⁷ Perhaps this difference may be accounted for in

⁶ Carroll, J. S., (1992), "How Taxpayers think About Their Taxes: Frames and Values", pp. 43 - 63.

⁷ This made it about one third of the percentage difference between the Tax Case X withholdings figure heeders's mean Tax Case X response and their Gamble Case mean response, which was a 23 percent rise.

terms of a residual awareness, from the case titles, that one case involved a tax situation, and the other, a straight gamble.

Therefore, to the extent that a tax-related difference exists between *refund* Tax Case Y and *high pay* Tax Case X in the comparisons made in both the current study and in Dusenbury (1994), the results of this study are in agreement with those of Dusenbury; but the interpretation is different. In the current study, the decision frame which is found to exist in terms of the H_{1A} and H_{5A} findings, may be *either* the gain or loss frames inherent in the summary tables *or* frames based firmly on perceptions of the prepayment position. The possibility of the former confounds, what was in Dusenbury (1994), a clear perception of the pivotal role of the latter.⁸ However, this clarity, which Dusenbury argued was cross-checked by his participants' responses to his additional ratings of judgment question is suspect, since he did not directly ask to what extent they paid heed to his story-line withholdings data.

Nevertheless, this finding does not negate the importance of tax withholdings in influencing taxpayers' compliance behaviour. All that is required in order to bring about a reduction in taxpayers' preferred levels of risk is for a gain frame to be made apparent when they prepare their tax returns for filing. This, in turn may be assured by the tax authority's collections of PAYE and provisional tax at a pay-period rate which is, at the minimum, in line with the rate levied on a taxpayer's income in its annualised form. However, it is acknowledged this is easier to achieve in terms of PAYE than provisional tax.

The study then went on to investigate several fairly uncontroversial aspects of taxpayer behaviour predicted by *Prospect Theory*. The first of these involved studying the stability of the taxpayer-participants' risk profiles in terms of a comparison of the responses made to two similar *high pay* tax scenarios. The evidence strongly supported acceptance of risk preference

⁸ The possibility of the former also gives rise to the question as to whether the nature of the participants gave rise to adoption of a particular decision frame. See Carroll, J. S., (1989), "A Cognitive-Process Analysis of Taxpayer Compliance", p. 254, which was cited in Chapter Three, Subsection 3.3.1. Carroll noted that there were a number of factors giving rise to framing effects in actual tax disclosure situations. He listed five possible taxpayer archetypes: (1) the honest taxpayer, (2) the utility maximiser, (3) the beaten taxpayer, (4) the equity seeker, and (5) the taxpayer who is mobilised by need. Of these archetypes, all but the fourth could well have applied in the formation of the decision frames acted upon in the current study.

stability across similar situations; but it is recognised that this stability, as posited in the null form of H_{6A} , could not be assessed for *within-subject* differences evolving over time.

Another extension involved determining whether the New-Zealand taxpayer participants made decisions underpinned by smooth value functions. It was found that most participants provided evidence of smoothly sequential preference orders with respect to the three replication tax cases, and that no statistically significant differences could be detected among the three observed sets of tax-related value functions. However, this consistency did not carry over to the full set of five Dusenbury-related decision problems; and the null form of H_{9A} was therefore rejected. The source of this inconsistency was the Gamble Case, in which a number of participants recorded value function discontinuities of the maximum possible size.

This finding is of interest, because one of the in-built assumptions underlying any laboratory experiment investigating decision frames within the *Prospect Theory* paradigm is that the underlying value function of participants is a smooth S-shape, concave for the gains and convex for losses, and which contains one neutral reference point through which this function is inflected. The finding implies that the smooth S-shape is not universally applicable. This accords with Schneider and Lopes (1986),⁹ who found that *Prospect Theory* did not adequately explain decision makers' risk preferences over the range of lottery types examined in their study. In addition, the finding implies that the smooth concave curve, posited to exist by *Expected Utility Theory*, is also not universally appropriate. The study of value functions, however, is a large field of endeavour, which is beyond the scope of the current study.

The area of greatest interest and concern in the New Zealand replication and extension of Dusenbury (1994) must be the incorporation of a cash flow dimension. The findings in this area do not negate what has already been established in terms of the New Zealand-North American cross-cultural comparison; but they do throw light on the nature of laboratory experiments in the area of decision making under risk.

⁹ Schneider, S. L. and Lopes, L. L. (1986), "Reflection in Preference under Risk: Who and When May Suggest Why". This study was cited in Chapter Two, Subsection 2.4.4.

The participants were divided initially into a subsample which received a high cash float in pseudo-dollars at the start of the experiment, and a subsample of the same size which received a substantially lower level of cash float.

The low cash float subsample, which received a uniform \$2,100, returned results with respect to H_{1A} which were in accordance with Dusenbury (1994). These results also were in keeping with the results obtained with respect to H_{1A} from analysis of the full sample of 132 participants. It is therefore possible to state, in terms of the 66 participants of low cash float subsample, that these taxpayers were significantly more risk averse when assessing prospects of various levels of refund in *refund* Tax Case Y than when choosing among prospects involving terminal tax payments in *high pay* Tax Case X. They exhibited a clear increase in risk willingness when responding to the latter decision problem.¹⁰

When the focus was shifted to H_{2A} , a similar outcome was detected. When the participants of the low cash float subsample had been segregated in terms of their responses to the question soliciting their attitude towards the holding of private health insurance, the subjects who considered private health insurance necessary furnished evidence of a risk preference shift which allowed for the rejection of the hypothesis' null form. These participants were significantly less risk willing when undertaking financial planning with respect to their health than they were with respect to paying their taxes. This finding accords with both Dusenbury (1994) and with the results obtained from this study's full sample partitioned by attitude to private health insurance.

Conversely, those participants who did not value the holding of private medical insurance policies provided results which did not allow rejection of the hypothesis' null form; and in this respect they conformed to the behaviour of the full sample segregated in the same manner. This dichotomy was caused by the cultural difference described above in the discussion of H_{2A} results pertaining to the full sample.

¹⁰ Carroll (1989), op. cit. n. 8. This finding accorded with Carroll's recognition of *need* as a mobilising force in the framing of options by decision makers with respect to tax.

The low cash float participants also furnished results, on the basis of which the null form of H_{3A} could be rejected. These taxpayers were significantly more risk averse when deciding on the level of income to report in their tax returns than when choosing among an identical set of risks and outcomes presented in terms of a gamble. Again, this result corroborated Dusenbury's (1994) findings and concurred with the results obtained with respect to the full sample.

When the low cash float subsample was employed in the investigation of H_{5A} , the result was, once more, in keeping with the result produced from an examination of the full sample of 132 participants. The low cash float participants' treatment of information regarding prior withholdings made no relevant contribution to the risk level chosen.

Finally, when the low cash float subsample was employed in the testing of H_{4A} , a significant difference was detected between its *high pay* Tax Case X mean and the corresponding mean produced by the \$3,500 cash float subsample; but this was the only instance in which H_{4A} could be rejected. The indeterminacy of the outcomes of these comparisons provides one ground for concluding that a confounding influence associated with cash float levels exists.

A further ground for what may be thought of as the *cash float knowledge* factor (explained in the next paragraph) was provided by the results associated with the various permutations of the high cash float. In contrast to the clarity of outcomes associated with the low cash float subsample, the high cash float subsample produced results which did not allow for the rejection of the null forms of H_{1A} , H_{2A} , and H_{4A} (with the exception of the low cash float and \$3,500 cash float comparison mentioned in the last paragraph). This was found to be the case when the high cash float subsample was employed as a combined grouping of the 36 recipients of the \$3,500 float and the 30 recipients of the \$4,200 float, and also when these two subgroupings were employed as subsamples independently. The only exception to this trend was that the two cash float subsamples, combined high and \$3,500, provided evidence supporting rejection of the null form of H_{3A} . It would seem that the participants' willingness to adopt a risky stance in a gambling context was of a different order to their approach to sorting the relative risk levels appropriate for adoption in the tax and medical decision problems.

The failure to reject the null forms of both H_{1A} and H_{2A} , on the one hand, and H_{4A} on the other, lends support to the argument raised in Chapter Eight, Subsection 8.5.6, in summation of the evidence presented with respect to H_{1A} . This argument posits that the participants' awareness of the existence of two cash float levels, and their awareness of which particular level applied to them, acted as a confounding factor. The recipients of both levels of high cash float (\$3,500 and \$4,200) knew that an equal number of other participants received only \$2,100. This caused them to be more complacent in their selection of options, confident they would make money for the organisations for which they were fundraising as volunteers. Conversely, the low cash float recipients knew they had a significantly lower level of seed-money funding; and this led to the belief that they would have to be very careful indeed if they were to avoid *losing* money for the organisations for which they were fundraising as volunteers. This effect could be called the *cash float knowledge* factor.¹¹

The *cash float knowledge* factor may be eliminated as a possible confounding variable in future studies of taxpayers' decision frames by the simple expedient of keeping the participants unaware of the existence of any cash float other than their own. This would entail running separate sessions for each cash float grouping. It would also involve the breach of ethics inherent in running an experiment on subjects who have not been apprised of the true nature of the study. This would not have been a feasible proposition for the current study, since it required formal clearance by the University of Canterbury's Human Ethics Committee. An alternative and ethically more acceptable approach would be to inform participants that the study involves assessing the effect of a number of different levels of cash float on participants' financial decision making; but until the study has been completed, information about these cash float levels would not be disclosed. The salient information to be left undisclosed is what the range of cash floats are, and where a given participants' cash float is positioned within that range.

¹¹ This factor is similar to the *liquidity* variable discussed by Martinez-Vazquez, Harwood and Larkins (1992). However, Martinez-Vazquez et al compared situations in which participants were told either they had the ability to pay a tax bill, or they did not have the ability to pay because of a shortfall in their cash reserves. See Martinez-Vazquez, J., Harwood, G. B. and Larkins, E. R., (1992), "Withholding Position and Income Tax Compliance: Some Experimental Evidence", which was also cited in Chapter Three, Subsection 3.3.8.

Several further comments are pertinent to the risk preference patterns displayed by the New Zealand taxpayer-participants in this study. Like Dusenbury's (1994) subjects, the New Zealand participants were risk averse in absolute terms. Dusenbury noted that the mean risk level, in his study, for all three sets of tax case responses combined was only 17 percent.¹² When rounded to the nearest integer, the mean response over the three New Zealand study tax cases was also 17 percent. Furthermore, the relative shifts in risk preference were also small, as the largest mean risk level adopted was only nine percentage points greater than this mean tax level, at 26 percent in the Gamble Case (recorded by the full sample and the low cash subsample), which rose a further one point to 27 percent with respect to the \$3,500 cash float subsample. These results belong very strongly in the same ball park as those recorded by Dusenbury, as his equivalent was 27 percent, again recorded in the Gamble Case.¹³ Dusenbury noted that findings of this order were indeed risk averse, since many tax professionals would consider reasonable even the 40 percent risk associated with both studies' riskiest options.¹⁴

The results of the current study also concur in this respect with the findings of Chang, Nichols and Schulz (1987),¹⁵ who tested their laboratory subjects for risk willingness and risk aversion in terms of the subjects' choices between agreeing to play lotteries and accepting, instead, their *certainty* alternatives. Chang et al's subjects were found to be risk averse in all instances apart from when tax payments were viewed as a pure loss. The current study's result, in this respect, also concurs with Schepanski and Kelsey's (1990) finding of mildly risk averse behaviour when subjects, given a decision problem with loss frame characteristics, faced probabilities of detection greater than 20 percent.¹⁶

¹² Dusenbury (1994), op. cit. n. 1, p. 13. The mean of 15% (*refund* case), 20% (*high pay* case) and 15% (*low pay* case) is 16.67 in Dusenbury's study. The current study's equivalent was 17.33%, which is the mean of 16% (*refund* case), 21% (*high pay* case) and 15% (*low pay* case).

¹³ *Idem.*

¹⁴ *Idem.*

¹⁵ Chang, O. H., Nichols, D. R. and Schulz, J. J., (1987), "Taxpayer Attitudes towards Tax-Audit Risk", cited in Chapter Three, Subsection 3.3.2.

¹⁶ Schepanski, A. and Kelsey, D., (1990), "Testing for Framing Effects in Taxpayer Compliance Decisions". This was also cited in Chapter Three, Subsection 3.3.6.

A further valid point was made by Dusenbury regarding the recording of significantly different risk levels for paired scenarios containing identical sets of monetary risks and returns, but dissimilar in financial context. He noted that this implied that a single preference ordering (value function) would have to have two slopes at the same point. The absurdity of this implied that, instead, two separate value functions must exist, and that they were induced by the context-sensitivity of the participants.¹⁷ This, in turn, supports the assumption underlying both Dusenbury (1994) and the current study, that *Prospect Theory's* S-shaped value function best describes the participants' preference orderings. The shortfall of this assumption with respect to the Gamble Case has already been commented upon.

Responses to the nineteen question end-of-session questionnaire made possible the investigation of a number of further factors potentially impinging on the risk levels participants chose in the study. One of these factors was found to be influential in determining participants' risk choices in the Medical Insurance Case. This was the participants' attitudes to the holding of a private health insurance policy, solicited in Question 4 and mentioned earlier in this section.

A *gender* difference was also detected with respect to responses to H_{1A} . When the full sample of 132 participants and the low cash float subsample of 66 participants were both partitioned by *gender*, the women in both subsamples provided evidence which strongly supported rejection of the null hypothesis. The men, however, furnished results upon which the null hypothesis could not be rejected. An explanation, similar in nature to the explanation of cash float subsample differences, was advanced in Chapter Nine, Subsection 9.3.2.1. The women were possibly keener fund-raisers than the men, and may have regarded the challenge to end the experiment with a positive contribution for their organisation of recruitment origin more seriously. The dedication of the women as fund raisers was also reflected in the composition of

¹⁷ Dusenbury (1994), op. cit. n. 1, p. 13.

the sample, in that the four contributing organisations were able to recruit 83 women (63 percent) to only 49 men.¹⁸

An *age*-related difference was also detected in the experiment. Participants who were in the 31 - 40 bracket tended to be more susceptible to framing effects than were participants who were either older or younger. This result was confirmed in terms of both the full sample and the low cash float subsample.

Influences emanating from *nature of income earning experience (Q14)* and *size of household income (Q13)* were, however, found upon exhaustive investigation, to have an insignificant impact on the participants' risk preferences. In particular, it was established that the increased opportunity on the part of the self-employed, to declare less taxable income (or alternatively, declare more income-related expense deductions) did not carry over into an increased propensity for risk willingness.

In all other instances, the variables generated from responses to the questionnaire were found to have only a statistically insignificant impact on the participants' selection of risky (or riskless) choices over the five replication scenarios in the study. These variables were:

1. Attitude to fairness of tax with respect to changes in the public health system (Q5),
2. Attitude to non-compliance by high income earners (Q6),
3. Attitude to non-compliance by low income earners (Q7),
4. Feelings about actual tax paying (Q8),
5. Frequency of payment-due tax-filings (Q9),
6. Perception of likelihood of an audit by the tax authorities (Q10),
7. Feelings about gambling (Q11),
8. Number of contributors to household income (Q12)
9. Education (Q15),
10. Awareness of general compliance with tax laws (Q16)
11. Professional tax-filing experience (Q17),
12. Knowledge of *Expected Utility Theory* (Q18),
13. Knowledge of *Prospect Theory* (Q19).

It is recognised that investigation of all nineteen variables furnished by the end-of-session questionnaire was not exhaustive over all permutations of the sample in terms of partitioning

¹⁸ This differential recruitment phenomenon reached its extreme in the school subsample, which contained 19 women and one man.

by cash float level, or over all possible permutations of potential interaction effects among the nineteen variables themselves. However, to have pursued all of these investigations would have entailed moving away from the primary focus of the study, which was to evaluate the nature of the risk shift (if any) between *refund* Tax Case Y and *high pay* Tax Case X, which was codified in hypothesis H_{1A} .

11.3 LIMITATIONS

The limitations recognised by Dusenbury (1994) with respect to the North American study apply in the main in this New Zealand conceptual replication. In the first instance, the participants were not randomly selected. Therefore, in Dusenbury's words, "...the representativeness of the subject pool is unknown."¹⁹

This issue of the representational faithfulness of the sample was addressed in Chapter Nine, Section 9.2. It was noted that while the proportions of the sample by income-earning experience were similar (but not identical) with the corresponding proportions recorded in the 1991 Census of Population and Dwellings, the sample quite markedly differed from the 1991 Census benchmark in terms of household income size. The fact that the information contained in the Census was four years out of date at the time of the laboratory experiment, however, made the Census an unreliable benchmark. Nevertheless, the participants did not faithfully represent the New Zealand public in terms of religious denomination; and the fact that participation entailed an act of altruism was possibly highly unrepresentative of the nature of New Zealander taxpayers in general. But the validity of this conjecture is not easily ascertainable.

Nevertheless, the use of a sample of taxpayers recruited beyond the pool of undergraduate Commerce Faculty students or MBA students (which is the most readily accessible source of participants for studies of this nature), must be seen as a relative strength. Both undergraduate Commerce Faculty students and working adults enrolled in MBA programmes may be expected

¹⁹ Dusenbury (1994), op. cit. n. 1, p. 14.

to have been taught the concept of evaluation of risky alternatives by their expected values. Until non-university Commerce Faculty recruits are used in tax compliance laboratory experiments, the presence of a confounding variable based on training to think in terms of *Expected Utility Theory* precepts cannot be discounted fully. It was possible to discount the presence of this variable in the current study on both the ground that the participants were not recruited from commerce classes of any kind, and also that they were asked to disclose their prior knowledge of theories of decision making under risk. The incidence of such prior knowledge was minimal.²⁰

A second limitation is inherent in the selection of the levels of risk and return employed in the five options of each decision problem in the study. In all six of the scenarios which have been used and discussed, the parameters of uncertainty, risk levels, costs and benefits were copied from Dusenbury (1994); and Dusenbury himself noted that these settings might well be an inaccurate representation of all tax return-filings.²¹

An attempt was made, however, to provide a context recognisable to New Zealand taxpayers in each of the scenarios. The current New Zealand tax rates of 24 cents in the dollar for income under \$NZ30,875 and 33 cents for every dollar of income earned in excess of that benchmark, were employed. Furthermore, the story-lines were adapted to show the impact of prior withholdings on the riskless filing option.

What cannot be known is the actual level of risk taxpayers expose themselves to when they record unjustifiable net taxable income figures in their tax returns. The New Zealand Inland Revenue Department does not divulge information of this nature. In this study it was noted that the participants believed that the Inland Revenue Department's audit rate was 24 percent.²² The

²⁰ In contrast to the finding in the current study, Martinez-Vasquez, Harwood and Larkins (1992) found that their MBA subjects (recruited from the South East of the United States) tended to choose the option associated with the highest expected value in each decision problem, regardless of variations in items likely to give rise to framing effects. However, White, Harrison and Harrell (1993), who used 81 part-time students and 175 undergraduate business school students, found strong evidence of framing effects. See Martinez-Vasquez, Harwood and Larkins (1992) op. cit. n. 11; and White, R. A., Harrison, P. D. and Harrell, A., (1993), "The Impact of Income Tax Withholding on Taxpayer Compliance: Further Empirical Evidence". Both of these papers were also cited in Chapter Three, Subsection 3.3.8.

²¹ Dusenbury (1994), op. cit. n. 1, p. 14.

²² Question 10, *Perception of likelihood of an audit by the tax authorities*, contained 21 boxes corresponding with 5% increments in audit likelihood. The second box represented a 5% rate, and the 21st box represented a 100% rate. The first box

percentage of respondents to a mail survey conducted in the same New Zealand city in 1992, who said they had been subjected to a tax audit, was also 24 percent.²³ However, the belief (whichever way it is measured) that the audit rate is set at this level does not make it so in actuality.

A related limitation of the current study was the jettisoning of Dusenbury's uniformity of income figures given in the tax decision problems.²⁴ This potentially transformed the story-line income figure into a confounding variable. There was an increase in income between *refund* Tax Case Y and *high pay* Tax Case X. If this income information induced a decision frame in participants' minds, then the framing effect was one of a perceived gain. According to *Prospect Theory*, gain frames produce risk averse decision making; however, the observed shift in risk preference was towards greater risk willingness. This would suggest (as argued in Chapter Four, Subsection 4.4.2) that the perceptions of loss associated with choices amongst terminal tax payments induced a more robust framing effect. However, the risk shift observed in participants who claimed to have used the story-line information about prior withholdings in *high pay* Tax Case X was less pronounced and far less statistically significant than the shift recorded by the subsample of participants who claimed to have paid this information scant heed. This suggests there was an income effect, and that it did dampen the swing towards relatively greater risk willingness induced by the loss frame associated with the paying of terminal taxes. If this diagnosis is accurate, then it adds credence to the power of a multiplicity of motivations which may exist in participants' minds, as noted by Jackson and Jones (1985)²⁵ and Carroll (1992).²⁶

represented a 1% rate. Given that these boxes were transformed into a 21-point Likert scale, the mean response, 5.6484, corresponds with an audit rate which is more than 23% and less than 25%. The standard deviation was 4.3735, which suggests that there was some divergence of opinion; and the range of responses recorded actually incorporated both of the scale's extremes. However, if they were correct in their estimation of the actual (but unknown) audit rate, the participants of the current study showed themselves to be risk averse in their recording of mean responses with lower risk levels than 24 percent in all of the tax cases.

²³ Hasseldine, D. J. and Kaplan, S. E., (1992), "The Effect of Different Sanction Communications on Hypothetical Taxpayer Compliance: Policy Implications from New Zealand", Table 1, p. 51.

²⁴ This was discussed in Chapter Four, Subsection 4.4.2.

²⁵ Jackson, B. R. and Jones, S. M., (1985), "Salience of Tax Evasion Penalties versus Detection Risk". This was cited in Chapter Three, Subsection 3.3.2.

²⁶ Carroll (1992) op. cit. n. 6.

The inability of the study to make definitive claims with respect to the absolute magnitudes of shifts in risk preferences constitutes a further limitation, which was also recognised by Dusenbury (1994).²⁷ Nevertheless, the five-option scale used by Dusenbury and in the current study has allowed for a more sophisticated mapping of these shifts than was possible in prior *Prospect Theory*-related tax compliance studies. The range of risks was kept relatively limited and low; but this possibly added to the realism, since the New Zealand participants believed mean tax audit rate to be only 24 percent.²⁸

The major strength of Dusenbury (1994), which carried over into the current study, was the use of *within-subject* contrasts.²⁹ This focussed attention on the risk preference shifts made by each participant without reference to a benchmark provided by any other participant.

11.4 EXTENSIONS

This study could be extended in the future in a number of directions. Dusenbury (1994) noted that his study could be reworked to see whether the setting of different tax rates, or greater or smaller risks, or greater or smaller monetary amounts made any difference to the risk preference shifts induced in participants.³⁰ Since the current study kept these variables constant, it could be reworked with any or all of those changes.³¹

²⁷ Dusenbury (1994) op. cit. n. 1, p. 14.

²⁸ See Schadewald, M. S., (1989), "Reference Point Effects in Taxpayer Decision Making", discussed in Chapter Three of this study. Schadewald's use of a 70% detection rate was subjected to critical analysis by Schepanski and Kelsey (1990), op. cit. n. 16. Schepanski and Kelsey found that as the detection rate rose to 45%, a framing effect, visible when the detection rate was 20%, disappeared. It is probable that taxpayers will be influenced by gain/loss framing effects within the range of detection probabilities they assume they encounter in reality; but where the stated audit rate is higher than experimental subjects' personal expectations of it, it may be expected to exert an undue dampening effect on their experimental responses.

²⁹ Dusenbury (1994), op. cit. n. 1, p. 14.

³⁰ *Ibid*, p. 15.

³¹ It was noted in Chapter Three, Subsection 3.3.3 that White, Harrison and Harrell (1993) argued that Hite, Jackson and Spicer's (1988) results did not show conclusive evidence of framing effects because Hite et al used probabilities of evasion detection which were too low for participants to consider significant. The discounting of low probabilities as negligible was posited by Kahneman and Tversky (1979) to occur as part of the editing phase operation they called *simplification*. Conversely, Schepanski and Kelsey (1990) noted that one of the flaws in Schadewald's (1989) experimental design was the selection of a detection rate which, at 70 percent, was too high and was therefore preternaturally dominant. Schepanski and Kelsey found that framing effects observed when the detection rate was 20 percent, disappeared when the probability of detection was raised to 45 percent. (See Hite, Jackson and Spicer (1988) op. cit. n. 5; and Kahneman, D. and Tversky, A., (1979), "Prospect Theory: An Analysis of Decision under Risk", p. 275. This was also mentioned in Chapter Two, Subsection 2.4.1.5. Also see

In particular, the varying of the tax rate would provide information which would be useful in maximising compliance via identification of an optimal tax rate — if such a rate exists. However, an extension of this nature would require significant rewording of the story-lines used in the decision problems in order to bring the tax rates out of the background obscurity (enjoyed by the two rates in the current study) and into the taxpayer-participants' conscious minds.

A second possible extension relates to the incorporation of varying levels of a cash float variable. In the current study, the size and relative position of each participant's cash float was disclosed information. It would be interesting to find out whether non-disclosure of this information, as suggested in Chapter Eight, Subsection 8.5.6 and in Chapter Ten, Section 10.2, would make any difference to the behaviour of participants with relatively higher levels of cash float.

The study could also be extended by reducing the low cash float to a level which caused cash flow difficulties to be experienced throughout the experiment, as distinct from just at the end of the session, which was the only time cash flow difficulties were experienced by any participants in the current study. Such an extension would involve a significant recasting of the experiment's format so that participants had an awareness throughout the experiment of the range of contingent financial events and *how* their long-term (within the experiment) cash flow could be expected to measure up against the income and outgoings occasioned by these events. This line of thinking is reminiscent of the two-period model developed by Robben, Webley, Elffers and Hessing (1990).³² Perhaps the extension could become a multiple-period simulation with tax problems in each period with a medical insurance problem and the opportunity to gamble presented in conjunction with the given tax problem in several of the periods. Furthermore, a lending arrangement could be incorporated in the experiment. However, with

Schadewald (1989) op. cit. n. 28; Schepanski and Kelsey, (1990), op. cit. n. 16; and White, Harrison and Harrell, op. cit. n. 20.

³² Robben, H. S. J., Webley, P., Elffers, H. and Hessing, D. J., (1990), "Decision Frames, Opportunity and Tax Evasion: An Experimental Approach". These researchers have produced a number of papers on this theme, which were surveyed in Chapter Three of this study.

an experiment of this complexity, it would be important to standardise the sequence of decision problems to a small number of orders to facilitate comparability.

A further extension was contemplated during the process of running the experiment used in the current study involving decision problems labelled *refund* Tax Case V and *high pay* Tax Case W. These contained tables in which the set of options had been reorganised so as to produce outcomes with standardised expected values. They were included in order that the effect of rejigging the expected values within scenarios might be gauged.³³

This issue was covered in part by Schepanski and Kelsey (1990)³⁴ to the extent that they provided a number of loss condition trials in which the expected value of the prospects were kept uniform. However each of Schepanski and Kelsey's trials contained a riskless and only one risky prospect as distinct from Dusenbury's array four risky prospects. The issue was also covered by White, Harrison and Harrell (1993),³⁵ who found evidence of a withholding framing effect which was stronger when the expected values of outcomes were bunched more closely together. In the light of these findings it would be useful to make a comparison between scenarios containing uniform expected values and scenarios with Dusenbury's (1994) non-uniform expected values. If it were to be undertaken, this extension would require the formulation of a number of different tax-related decision problems presented as linked pairs, identical in every respect with the exception of the configuration of the risks and returns. Such an experiment would throw light on the relative explanatory powers of the *Expected Utility Theory* derivative, *Deterrence Theory* and *Prospect Theory* with respect to a *within-subject* risk profile analysis in the area of tax compliance as affected by prior withholdings.

Finally, in keeping with Dusenbury's (1994) thoughts about a future extension of his work in a United States context,³⁶ it would be of value to investigate the impact of the institutional

³³ This extension was not pursued in the current study because it was not sufficiently well planned in advance. Nevertheless, the opportunity to make use of the participants already assembled for the current study was taken up. The addition of two extra scenarios did not make an impact on the participants' risk profiles with respect to the five replication scenarios because these extra two were always presented last, after the presentation of the three extra ratings questions.

³⁴ Schepanski, A. and Kelsey, D., (1990), op. cit. n. 16.

³⁵ White, R. A., Harrison, P. D. and Harrell, A., (1993), op cit. n. 20, p. 75.

³⁶ Dusenbury (1994), op. cit. n. 1, p. 16.

features of the income tax system in New Zealand with a view to determining how to maximise risk aversion in disclosure of annual taxable income.

11.5 IMPLICATIONS FOR TAX ADMINISTRATION

In accordance with the tax legislation, the Inland Revenue Department already makes use of withholding taxes with respect to income on interest and dividends, requires PAYE to be withheld from the paypackets of wage and salary earners, and requires prepayments (known as provisional tax) from income earners who are self-employed. The levying of these taxes, either at the time that income is earned, or even in advance (in the case of provisional tax), has been based on knowledge of developments in compliance research made in the United Kingdom, Australia, Canada, and, to a substantial degree, in the United States of America. The current study provides an insight into the validity of such prepayment requirements based on an analysis of a sample of New Zealand taxpayer-participants' propensities for taking risks in meeting their tax obligations.

The study presents strong evidence that New Zealanders, as represented by the taxpayer-participants, are susceptible to framing effects which may be manipulated by the judicious setting of tax policy. There is clearly a connection between perceptions of gain and relative risk aversion, and a connection between perceptions of loss and relative risk seeking behaviour. It is also clear that individual taxpayers alter their risk preferences according to whichever of these perceptions is generated by the financial figures with which the taxpayer commences computation of his or her end-of-year tax balance. This was manifest in the study by the use of *within-subject* GLM analysis of variance procedures. As a consequence, this finding also validates *Prospect Theory*, which in the wider tax compliance context may be viewed as a branch of the *fiscal psychology* paradigm, as a proper tool for tax research in New Zealand.

The swings in the level of risk adopted by subjects responding to a variety of framing effects (both taxation-based and independent of taxation) suggest that the *fiscal psychology* model's general approach to improvement in tax compliance has merit, since participants altered their preferred levels of risk in a manner which was, to a degree, independent of considerations of the unpleasantness of possible outcomes. However, the study has not necessarily proven that a

fiscal psychology approach to compliance reinforcement is sufficient to render obsolete the use of penalty level and audit rate manipulations advocated in the *economic deterrence* modeling of tax compliance behaviour.

12. REFERENCES

- Allais, M., [1953], "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'Ecole Americaine", *Econometrica*, Vol. 21, October, pp. 503 - 46.
- Allingham, M. G. and Sandmo, A., [1972], "Income Tax Evasion: A Theoretical Analysis", *Journal of Public Economics*, Vol 1, No. 3/4, November, pp. 323 - 338.
- Alm, J., Jackson, B. and McKee, M., [1992], "Deterrence and Beyond: Toward a Kinder, Gentler IRS", in J. Slemrod (ed), *Why People Pay Taxes: Tax Compliance and Enforcement*, University of Michigan Press, Ann Arbor.
- Alt, F., [1936], "Über die Messbarkeit des Nutzens", *Zeitschrift für Nationalökonomie*, Vol. 7, pp. 161 - 69. English translation: "On the Measurement of Utility", in J. S. Chipman, L. Hurwicz, M. K. Richter and H. F. Sonnenschein (eds), *Preferences, Utility, and Demand*, Harcourt Brace Jovanovich, New York, 1971, pp. 424 - 431.
- Arkes, H. R. and Blumer, C., [1985], "The Psychology of Sunk Cost", *Organizational Behavior and Human Decision Processes*, Vol. 35, No. 1, February, pp. 124 - 140.
- Arrow, K. J., [1971], *Essays in the the Theory of Risk-Bearing*, Markham, Chicago.
- Baldry, J. C., [1986], "Tax Evasion is Not a Gamble: A Report on Two Experiments", *Economics Letters*, Vol. 22, pp. 333 - 335.
- Bardsley, P., [1994], "Tax Compliance Research: An Economic Perspective on the Research Agenda", *Australian Tax Forum*, Vol. 11, No.3, pp. 271 - 290.
- Beck, P. J., Davis, J. S. and Jung, W., [1991], "Experimental Evidence on Taxpayer Reporting Under Uncertainty", *The Accounting Review*, Vol. 66, No. 3, July, pp. 535 - 558.
- Beck, P. J. and Jung, W. [1989], "Taxpayer Compliance under Uncertainty", *Journal of Accounting and Public Policy*, Vol. 8, No. 1, Spring, pp. 1 - 28.
- Bernoulli, D. , [1738], "Exposition of a New Theory on the Measurement of Risk", *Econometrica*, Vol. 22, No. 1, January 1954, (Translated from Latin by L. Sommer), pp. 23 - 36.
- Bhattacharjee, V., Cicchetti, C. J. and Rankin, W. F., [1993], "Energy Utilities, Conservation and Economic Efficiency", *Contemporary Policy Issues*, Vol. 11, No. 1, January, pp. 69 - 75.
- Borg, W. R. and Gall, M. D., [1983], *Educational Research: An Introduction*, Longman, Inc, White Plains, New York.
- Brockner, J., [1992], "The Escalation of Commitment to a Failing course of Action: Toward Theoretical Progress", *Academy of Management Review*, Vol. 17, No. 1, January, pp. 39 - 61.
- Budescu, D. V. and Weiss, W., [1987], "Reflection of Transitive and Intransitive Preferences: A Test of Prospect Theory", *Organizational Behavior and Human Decision Processes*, Vol 39, pp. 184 - 202.

- Carroll, J. S., [1989], "A Cognitive-Process Analysis of Taxpayer Compliance", in J. A. Roth, and J. T. Scholz (eds), *Taxpayer Compliance Vol. 2: Social Science Perspectives*, University of Pennsylvania Press, Philadelphia.
- Carroll, J. S., [1992], "How Taxpayers Think About Their Taxes: Frames and Values", in J. Slemrod (ed), *Why People Pay Taxes: Tax Compliance and Enforcement*, University of Michigan Press, Ann Arbor.
- Casey, J. T., [1994], "Buyers' Pricing Behavior for Risky Alternatives: Encoding Processes and Preference Reversals", *Management Science*, Vol. 40, No. 6, June, pp. 730 - 749.
- Casey, J. T. and Scholz, J. T., [1991], "Boundary Effects of Vague Risk Information on Taxpayer Decisions", *Organizational Behavior and Human Decision Processes*, Vol. 50, No. 2, December, pp. 360 - 394.
- Chang, O. H. and Schultz, J. J. Jr., [1990], "The Income Tax Withholding Phenomenon: Evidence from TCMP Data", *The Journal of the American Taxation Association*, Vol. 12, No. 1, Fall, pp. 89 - 93.
- Chang, O. H., Nichols, D. R. and Schultz, J. J., [1987], "Taxpayer Attitudes toward Tax-Audit Risk", *Journal of Economic Psychology*, Vol. 8, No. 3, September, pp. 299 - 309.
- Choquet, G., [1955], "Theory of Capacities", *Annales de l'Institut Fourier*, Vol. 5, pp. 131 - 295.
- Christian, C. W., Gupta, S., Weber, G. J. and Willis, E., [1994], "The Relation between the Use of Tax Preparers and Taxpayers' Prepayment Position", *The Journal of the American Taxation Association*, Vol 16, No. 1, Spring, pp. 17 - 40.
- Clotfelter, C. T., [1983], "Tax Evasion and Tax Rates: An Analysis of Individual Returns", *The Review of Economics and Statistics*, Vol. 65, No. 3, August, pp. 363 - 373.
- Cowell, F. A., [1985], "Tax Evasion with Labour Income", *Journal of Public Economics*, Vol. 26, No.1, February, pp. 19 - 34.
- Cox, D. and Plumley, A., [1988], Analyses of Voluntary Compliance Rates for Different Income Source Classes, Unpublished report, International Revenue Service, Research Division, Washington DC.
- Crane, S. E. and Nourzad, F., [1990], "Tax Rates and Tax Evasion: Evidence from California Amnesty Data", *National Tax Journal*, Vol. 43, No. 2, June, pp. 189 - 199.
- Darke, P. R. and Freedman, J. L., [1993], "Deciding on whether to Seek a Bargain: Effects of both Amount and Percentage Off", *Journal of Applied Psychology*, Vol. 78, No. 6, December, pp. 960 - 965.
- Dubin, J. A., Graetz, M. J. and Wilde, L. L., [1990], "The Effect of Audit Rates on the Federal Individual Income Tax, 1977 - 1986", *National Tax Journal*, Vol. 43, No. 4, December, pp. 395 - 409.
- Dusenbury, R., [1994], "The Effect of Prepayment Position on Individual Taxpayers' Preferences for Risky Tax-Filing Options", *The Journal of the American Taxation Association*, Vol. 16, No. 1, Spring, pp. 1 - 16.

Dusenbury, R., [1995], A Personal Communication via Email.

Einhorn, H. J. and Hogarth, R. M., [1986], "Decision Making Under Ambiguity", *Journal of Business*, Vol. 59, No. 4, Pt. 2, pp. s225 - s249.

Elffers, H., Weigel, R. H., and Hessing, D. J., [1987], "The Consequences of Different Strategies for Measuring Tax Evasion Behavior", *Journal of Economic Psychology*, Vol. 8, pp. 311 - 337.

Elliot, C. S. and Archibald, R. B., [1989], "Subjective Framing and Attitudes Towards Risk", *Journal of Economic Psychology*, Vol. 10, No. 3, November, pp. 321 - 328.

Emory, C. W. and Cooper, D. R., [1991], *Business Research Methods* (Fourth Edition), Richard D. Irwin Inc., Boston.

Fischer, C. M. and Russell, J. D., [1991], "Research Note: Replication of Empirical Research", *The Journal of the American Taxation Association*, Vol. 13, No. 1, Spring, pp. 73 - 91.

Fischhoff, B., [1983], "Predicting Frames", *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 9, No. 1, January, pp. 103 - 116.

Fishburn, P. C., [1988], *Nonlinear Preference and Utility Theory*, The Johns Hopkins University Press, Baltimore.

Fishburn, P. C. and Kochenberger, G. A., [1979], "Two-Piece Von Neumann - Morgenstern Utility Functions", *Decision Sciences*, Vol. 10, pp. 503 - 518.

Fountain, J., [1992], Quasi Rational Consumer Demand - Some Positive and Normative Surprises, Discussion Paper No. 9202, Department of Economics, University of Canterbury, New Zealand.

Fountain, J. and McCosker, M., [1993], Fans, Frames, and Risk Aversion: How Robust is the Common Consequence Effect?, Discussion Paper No. 9303, Department of Economics, University of Canterbury, New Zealand.

Fountain, J., McCosker, M. and Morris, D., [1994], Are People Really Risk Seeking When Facing Losses?, Discussion Paper No. 9407, Department of Economics, University of Canterbury, New Zealand.

Fountain, J., McCosker, M. and MacFarlane, B., [1995], Framing and Incentive Effects on risk Attitudes when Facing Uncertain Losses, Discussion Paper No. 9503, Department of Economics, University of Canterbury, New Zealand.

Frisch, R., [1926], "Sur un Probleme d'economie pure", *Norsk Matematisk Forenings Skrifter*, Vol. 16, pp. 1 - 40. English translation: "On a Problem in Pure Economics" in J. S. Chipman, L. Hurwicz, M. K. Richter and H. F. Sonnenschein (eds), *Preferences, Utility and Demand*, Harcourt Brace Jovanovich, New York, 1971, pp. 386 - 423.

Garland, H. and Newport, S., [1991], "Effects of Absolute and Relative Sunk Costs on the Decision to Persist with a Course of Action", *Organizational Behavior and Human Decision Processes*, Vol. 48, No. 1, February, pp. 55 - 69.

- Gregory, R., [1986], "Interpreting Measures of Economic Loss: Evidence from Contingent Valuation and Experimental Studies", *Journal of Environmental Economics and Management*, Vol. 13, No. 4, December, pp. 325 - 337.
- Grether, D. M. and Plott, C. R., [1979], "Economic Theory of Choice and the Preference Reversal Phenomenon", *The American Economic Review*, Vol. 69, No. 4, September, pp. 623 - 638.
- Groenland, A. G. and van Veldhoven, G. M., [1983], "Tax Evasion Behavior: A Psychological Framework", *Journal of Economic Psychology*, Vol. 3, pp. 129 - 144.
- Hamm, S. and Shettleworth, S., [1987], "Risk Aversion among Pigeons", *Journal of Experimental Psychology: Animal Behavior Processes*, Vol.13, pp. 376 - 383.
- Hasen, R. L., [1990], "Efficiency under Informational Asymmetry: The Effect of Framing on Legal Rules", *UCLA Law Review*, Vol. 38, No. 2, December, pp. 391 - 438.
- Hasseldine, D. J. and Bebbington, K. J., [1991], "Blending Economic Deterrence and Fiscal Psychology Models in the Design of Responses to Tax Evasion: The New Zealand Experience", *Journal of Economic Psychology*, Vol. 12, pp. 299 - 324.
- Hasseldine, D. J. and Kaplan, S. E., [1992], "The Effect of Different Sanction Communications on Hypothetical Taxpayer Compliance: Policy Implications from New Zealand", *Public Finance / Finances Publiques*, Vol. 47, No. 1, pp. 45 - 60.
- Hasseldine, D. J., Kaplan, S. E. and Fuller, L. R., [1994], "Characteristics of New Zealand Tax Evaders: A Note", *Accounting and Finance*, Vol. 34, No. 2, November, pp. 79 - 93.
- Hershey, J. and Schoemaker, P. J. H., [1980], "Prospect Theory's Reflection Hypothesis: A Critical Examination", *Organizational Behavior and Human Performance*, Vol. 25, pp. 395 - 418.
- Hessing, D. J., Elffers, H. and Weigel, R. H., [1988], "Exploring the Limits of Self-Reports and Reasoned Action: An Investigation of the Psychology of Tax Evasion Behavior", *Journal of Personality and Social Psychology*, Vol. 54, No. 3, March, pp. 405 - 413.
- Hessing, D. J., Elffers, H., Robben, H. S. J. and Webley, P., [1992], "Does Deterrence Deter? Measuring the Effect of Deterrence on Tax Compliance in Field Studies and Experimental Studies", in J. Slemrod (ed), *Why People Pay Taxes: Tax Compliance and Enforcement*, University of Michigan Press, Ann Arbor.
- Hilton, R. W., [1988], "Risk Attitude under Two Alternative Theories of Choice under Risk", *Journal of Economic Behavior and Organization*, Vol. 9, No. 2, March, pp. 119 - 136.
- Hirst, D. E., Joyce, E. T. and Schadeveld, M. S., [1994], "Mental Accounting and Outcome Contiguity in Consumer-Borrowing Decisions", *Organizational Behavior and Human Decision Processes*, Vol. 58, No. 1, April, pp. 136 - 152.
- Hite, P. S., Jackson, B. R. and Spicer, M. W., [1988], The Effect of Framing Biases on Taxpayer Compliance, Unpublished paper received from P. S. Hite.
- Hogarth, R. M. and Einhorn, H. J., [1990], "Venture Theory: A Model of Decision Weights", *Management Science*, Vol. 7, No. 7, July, pp. 780 - 803.

- Holt, C. A., [1986], "Preference Reversals and the Independence Axiom", *American Economic Review*, Vol. 76, June, pp. 508 - 515.
- Jackson, B. R. and Jones, S. M., [1985], "Salience of Tax Evasion Penalties Versus Detection Risk", *The Journal of the American Taxation Association*, Vol. 6, Spring, pp. 7 - 17.
- Jackson, B. R. and Milleron, V. C., [1986], "Tax Compliance Research: Findings, Problems, and Prospects", *Journal of Accounting Literature*, Vol. 5, pp. 125 - 165.
- Jensen, N. E., [1967], "An Introduction to Bernoullian Utility Theory: 1. Utility Functions", *Swedish Journal of Economics*, Vol. 69, pp. 163 - 183.
- Kachelmeier, S. J. and Shehata, M., [1992], "Examining Risk Preferences Under High Monetary Incentives: Experimental Evidence from the People's Republic of China", *The American Economic Review*, Vol. 82, No. 5, December, pp. 1120 - 1141.
- Kahneman, D. and Tversky, A., [1979], "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica*, Vol. 47, No. 2, March, pp. 263 - 291.
- Kahneman, D. and Tversky, A., [1984], "Choices, Values, and Frames", *American Psychologist*, Vol. 39, No. 4, April, pp. 341 - 350.
- Kameda, T. and Davis, J. H., [1990], "The Function of the Reference Point in Individual and Group Risk Decision Making", *Organizational Behavior and Human Decision Processes*, Vol. 46, No. 1, June, pp. 55 - 76.
- Kanto, A. J., Rosenqvist, G. and Suvas, A., [1992], "On Utility Function Estimation of Racetrack Bettors", *Journal of Economic Psychology*, Vol. 13, No. 3, pp. 491 - 498.
- Kinsey, K. A., [1984], Survey Data on Tax Compliance: A Compendium and Review, Taxpayer Compliance Project working Paper 84 - 1, American Bar Association, Chicago, Illinois.
- Kotz, S., Johnson, N. L. and Read, C. B. (eds), [1988], *Encyclopedia of Statistical Sciences*, John Wiley and Sons, New York.
- Lakatos, I., [1974], "Falsification and the Methodology of Scientific Research Programmes", in I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*, Cambridge University Press, Cambridge.
- Lange, O., [1934], "The Determinateness of the Utility Function", *Review of Economic Studies*, Vol. 1, pp. 218 - 224.
- Lewis, A., [1982], *The Psychology of Taxation*, Martin Robinson, Oxford.
- Loewenstein, G. F., [1988], "Frames of Mind in Intertemporal Choice", *Management Science*, Vol. 34, No. 2, February, pp. 200 - 214.
- Loftus, E. F., [1985], "To File, Perchance to Cheat", *Psychology Today*, Vol. 19, No. 4, April, pp. 34 - 39.
- Loke, W. H., [1989], "The Effects of Framing and Incomplete Information on Judgments", *Journal of Economic Psychology*, Vol. 10, No. 3, November, pp. 329 - 341.

- Loomes, G. and Sugden, R., [1982], "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty", *Economic Journal*, Vol. 92, pp. 803 - 824.
- Luce, R. D. and Raiffa, H., [1957], *Games and Decisions*, John Wiley & Sons, Inc., New York.
- Machina, M. J., [1987], "Choice Under Uncertainty: Problems Solved and Unsolved", *Economic Perspectives*, Vol. 1, No. 1, Summer, pp. 121 - 154.
- Markowitz, H., [1952], "The Utility of Wealth", *Journal of Political Economy*, Vol. 60, No. 2, April, pp. 151 - 158.
- Martinez-Vazquez, J., Harwood, G. B. and Larkins, E. R., [1992], "Withholding Position and Income Tax Compliance: Some Experimental Evidence", *Public Finance Quarterly*, Vol. 20, No. 2, April, pp. 152 - 174.
- McCaffery, E. J., [1994], "Cognitive Theory and Tax", *UCLA Law Review*, Vol. 41, pp. 1861 - 1947.
- McGuire, T. W., Kiesler, S. and Siegel, J., [1987], "Group and Computer-Mediated Discussion Effects in Risk Decision Making", *Journal of Personality and Social Psychology*, Vol. 52, No. 5, May, pp. 917 - 930.
- Milleron, V. C. and Toy, D. R., [1988], "Tax Compliance: An Investigation of Key Features", *The Journal of the American Taxation Association*, Vol. 9, Spring, pp. 84 - 104.
- Minarik, J. J., [1988], "Individual Income Tax Issues as Revised by Tax Reform", *National Tax Journal*, Vol. 41, No. 3, September, pp. 291 - 301.
- Ott, L. and Mendenhall, W., [1985], *Understanding Statistics* (Fourth Edition), Duxbury Press, Boston.
- Payne, J. W., Laughhunn, D. J. and Crum, R., [1980], "Translation of Gambles and Aspiration Level Effects in Risky Choice Behavior", *Management Science*, Vol 26, pp. 1039- 11060.
- Payne, J. W., Laughhunn, D. J. and Crum, R., [1981], "Further Tests of Aspiration Level Effects in Risky Choice Behavior", *Management Science*, Vol. 27, pp. 953 - 958.
- Peckman, J. A., (1983), *Federal Tax Policy*, Paper produced for the Brookings Institute, Washington DC.
- Pommerehne, W. W., Schneider, F. and Zweifel, P., [1982], "Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination", *American Economic Review*, Vol. 72, No. 3, June, pp. 569 - 574.
- Porcano, T. M., [1988], "Correlates of Tax Evasion", *Journal of Economic Psychology*, Vol. 9, pp. 47 - 67.
- Pratt, J. W., [1964], "Risk Aversion in the Small and in the Large", *Econometrica*, Vol. 32, No. 1 - 2, January - April, pp. 132 - 136.
- Ravinder, H. V., Kleinmuntz, D. N. and Dyer, J. S., [1988], "The Reliability of Subjective Probabilities Obtained Through Decomposition", *Management Science*, Vol. 34, No. 2, February, pp. 186 - 199.

- Reilly, R. J., [1982], "Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design", *The American Economic Review*, Vol. 72, No. 3, June, pp. 576 - 584.
- Robben, H. S. J., Webley, P., Elffers, H. and Hessing, D. J., [1990], "Decision Frames, Opportunity and Tax Evasion: An Experimental Approach", *Journal of Economic Behavior and Organization*, Vol. 14, No. 3, December, pp. 353 - 361.
- Robben, H. S. J., Webley, P., Weigel, R. H., Warneryd, K., Kinsey, K. A., Hessing, D. J., Alvira Martin, F., Elffers, H., Wahlund, R. Van Langenhove, L., Long, S. B. and Scholz, J. T. , [1990], "Decision Frame and Opportunity as Determinants of Tax Cheating", *Journal of Economic Psychology*, Vol. 11, No. 3, September, pp. 341 - 364.
- Rothschild, M. and Stiglitz, J., [1970], "Increasing Risk 1: A Definition", *Journal of Economic Theory*, Vol. 2, No. 3, September, pp. 225 - 243.
- Salminen, P. and Wallenius, J., [1993], "Testing Prospect Theory in a Deterministic Multiple Criteria Decision-Making Environment", *Decision Sciences*, Vol. 24, No. 2, March/April, pp. 279 - 294.
- Sandford, C. and Wallschutzky, I., [1994], "Self-Assessment of Income Tax: Lessons from Australia", *British Tax Review*, pp. 213 - 220.
- Sanders, D. L. and Wyndelts, R. W., [1989], "An Examination of Tax Practitioners' Decisions under Uncertainty", *Advances in Taxation*, Vol. 2, pp. 41 - 71.
- SAS Institute Inc., [1985], *SAS Introductory Guide for Personal Computers, Version 6 Edition*, SAS Institute Inc., Cary, North Carolina.
- SAS Institute Inc., [1985], *SAS Procedures Guide for Personal Computers, Version 6 Edition*, SAS Institute Inc., Cary, North Carolina.
- SAS Institute Inc., [1989], *SAS/STAT User's Guide, Version 6, Fourth Edition*, SAS Institute Inc., Cary, North Carolina.
- Savage, L. J., [1954], *The Foundations of Statistics*, Wiley, New York.
- Schadewald, M. S., [1989], "Reference Point Effects in Taxpayer Decision Making", *The Journal of the American Taxation Association*, Vol. 10, Spring, pp. 68 - 84.
- Schaubroeck, J. and Davis, E., [1994], "Prospect Theory Predictions when Escalation is Not the Only Chance to Recover Sunk Costs", *Organizational Behavior and Human Decision Processes*, Vol. 57, No. 1, January, pp. 59 - 82.
- Schepanski, A. and Kelsey, D., [1990], "Testing for Framing Effects in Taxpayer Compliance Decisions", *The Journal of the American Taxation Association*, Vol. 12, Fall, pp. 60 - 77.
- Schisler, D. L., [1992], An Experimental Examination of Taxpayers' Aggressive Tax Preferences: A Prospect Theory Foundation, PhD Thesis, Memphis State University, Tennessee.
- Schisler, D. L., [1994], "An Experimental Examination of Factors Affecting Tax Preparers' Aggressiveness - A Prospect Theory Approach", *The Journal of the American Taxation Association*, Vol. 16, No. 2, Fall, pp. 124 - 142.

- Schneider, S. L. and Lopes, L. L., [1986], "Reflection in Preferences under Risk: Who and When May Suggest Why", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 12, No. 4, November, pp. 535 - 548.
- Schoemaker, P. J. H., [1991], "Choices Involving Uncertain Probabilities: Tests of Generalized Utility Models", *Journal of Economic Behavior and Organization*, Vol. 16, No. 3, December, pp. 295 - 317.
- Scholz, J. T., [1985], *Coping with Complexity: A Bounded Rationality Perspective on Taxpayer Compliance*, Proceedings of the Seventy-Eighth Annual Conference of the National Tax Association - Tax Institute of America, Columbus, Ohio.
- Shafir, E. B., Osherson, D. N. and Smith, E. E., [1993], "The Advantage Model: A Comparative Theory of Evaluation and Choice under Risk", *Organizational Behavior and Human Decision Processes*, Vol. 55, No. 3, August, pp. 325 - 378.
- Shafir, E., Simonson, I. and Tversky, A., [1993], "Reason-based Choice", *Cognition*, Vol. 49, pp. 11 - 36.
- Sheffrin, S. M., [1993], "What Does the Public Believe about Tax Fairness?", *National Tax Journal*, Vol. 46, No. 3, September, pp. 301 - 308.
- Simon, H. A., [1955], "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics*, Vol. 69, No. 1, February, pp. 99 - 118.
- Simon, H. A., [1956], "Rational Choice and the Structure of the Environment", *Psychological Review*, Vol. 63, No.2, pp. 129 - 138.
- Simon, H. A., [1976], "From Substantive to Procedural Rationality", in S. J. Latsis (ed), *Method and Appraisal in Economics*, Cambridge University Press, Cambridge.
- Simon, H. A., [1978], "On How to Decide What to Do", *Bell Journal of Economics*, Vol. 9, No. 2, pp. 494 - 507.
- Sinha, T., [1994], "Prospect Theory and the Risk Return Association: Another Look", *Journal of Economic Behavior and Organization*, Vol. 24, No. 2, July, pp. 225 - 231.
- Slemrod, J., [1992], *Why People Pay Tax: Tax Compliance and Enforcement*, The University of Michigan Press, Ann Arbor.
- Slovic, P. and Lichtenstein, S., [1983], "Preference Reversals: A Broader Perspective", *American Economic Review*, Vol. 73, No. 4, September, pp. 596 - 605.
- Smith, K. W. and Kinsey, K. A., [1987], "Understanding Taxpaying Behavior: A Conceptual Framework with Implications for Research", *Law and Society Review*, Vol. 21, No. 4, pp. 639 - 663.
- Smith, V., [1987], "Experimental Methods in Economics", in J. Eatwell, M. Milgate, and P. Newman (eds), *The New Palgrave: A Dictionary of Economics*, Vol. 2, Macmillan, London, pp 241 - 249.
- Smith, V. L. and Walker, J. M., [1992], *Monetary Rewards and Decision Cost in Experimental Economics*, Unpublished Manuscript, Economic Science Laboratory, University of Arizona, Arizona.

- Srinivasan, T. N., [1973], "Tax Evasion: A Model", *Journal of Public Economics*, Vol. 2, No. 4, November, pp. 339 - 346.
- Starmer, C. and Sugden, R., [1991], "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation", *American Economic Review*, Vol. 81, No. 4, September, pp. 971 - 978.
- Statistics New Zealand, [1992], *New Zealand Census of Population and Dwellings 1991*, Department of Statistics New Zealand, Te Tari Tatau O Aotearoa, Wellington.
- Statistics New Zealand, [1995], *New Zealand Official Yearbook 1995 (Te Pukapuka Honanga Whaimana o Aotearoa)*, 98th Ed., Statistics New Zealand (Te Tari Tatau), Wellington.
- Statistics New Zealand, [1995], *Key Statistics October 1995*, Statistics New Zealand (Te Tari Tatau), Wellington.
- Sternburg, T. J., [1993], An Empirical Investigation of Deterrence Theory and Influence Theory on Professional Tax Preparers' Recommendations, PhD Thesis, Arizona State University, Arizona.
- Swenson, C. W., [1988], "Taxpayer Behavior in Response to Taxation: An Experimental Analysis", *Journal of Accounting and Public Policy*, Vol. 7, No. 1, Spring, pp. 1 - 28.
- Tan, L. M. and Tooley, S., [1994], The Effectiveness of New Zealand Tax Simplification Initiatives: Preliminary Evidence from a Survey of Tax Practitioners, Working Paper, Massey University, New Zealand.
- Thaler, R. H., [1986], "The Psychology and Economics Conference Handbook: Comments on Simon, on Einhorn and Horgarth and on Tversky and Kahneman", *Journal of Business*, Vol. 59, No. 4, Pt. 2, October, pp. s279 - s284.
- Trimpop, R. M., [1994], *The Psychology of Risk Taking Behavior*, North-Holland Elsevier Science B. V., Amsterdam.
- Tversky, A., [1969], "Intransitivity of Preferences", *Psychological Review*, Vol. 76, No. 1, January, pp. 31 - 48.
- Tversky, A., [1972], "Elimination by Aspects: A Theory of Choice", *Psychological Review*, Vol. 79, No. 4, July, pp. 281 - 299.
- Tversky, A. and Kahneman, D., [1981], "The Framing of Decisions and the Psychology of Choice", *Science*, Vol. 211, No. 4481, 30 January, pp. 453 - 458.
- Tversky, A. and Kahneman, D., [1986], "Rational Choice and the Framing of Decisions", *Journal of Business*, Vol. 59, No. 4, Pt. 2, October, pp. s251 - s284.
- Tversky, A. and Kahneman, D., [1992], "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty*, Vol. 5, No. 4, pp. 297 - 323.
- Tversky, A., Slovic, P. and Kahneman, D., [1990], "The Causes of Preference Reversal", *The American Economic Review*, Vol. 80, No. 1, March, pp. 204 - 217.
- Tversky, A., Slovic, P. and Sattath, S., [1988], "Contingent Weighting in Judgement and Choice", *Psychological Review*, Vol. 95, No. 3, July, pp. 371 - 384.

- Von Neumann, J. and Morgenstern, O, [1947], *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.
- Wallschutzky, I. G., [1984], "Possible Causes of Tax Evasion", *Journal of Economic Psychology*, Vol. 5, No. 4, December, pp. 371 - 384.
- Wallschutzky, I. G., [1989], "The Effectiveness of Tax Office Enforcement Activities in Achieving Taxpayer Compliance", *Taxation in Australia*, Vol. 23, No. 7, February, pp. 495 - 503.
- Warneryd, K. and Walerud, B., [1982], "Taxes and Economic Behavior: Some Interview Data on Tax Evasion in Sweden", *Journal of Economic Psychology*, Vol. 2, No. 3, September, pp. 187 - 211.
- Wartick, M. L., [1994], "Legislative Justification and the Perceived Fairness of Tax Law Changes: A Referent Cognitions Theory Approach", *The Journal of the American Taxation Association*, Vol. 16, No. 2, Fall, pp. 106 - 123.
- Webley, P, Robben, H., Elffers, H. and Hessing, D., [1991], *Tax Evasion: An Experimental Approach*, Cambridge University Press, Cambridge, Great Britain.
- Weiss, L., [1976], "The Desirability of Cheating Incentives and Randomness in the Optimal Income Tax", *Journal of Political Economy*, Vol. 84, No. 6, December, pp. 1343 - 1352.
- Westat Inc., [1980], Individual Income Tax Factors Study: Qualitative Research Results, Paper prepared for the Internal Revenue Service, 4 February 1980.
- White, R. A., Harrison, P. D. and Harrell, A., [1993], "The Impact of Income Tax Withholding on Taxpayer Compliance: Further Empirical Evidence", *The Journal of the American Taxation Association*, Vol 15, No. 2, Fall, pp. 63 - 78.
- Whyte, G., [1993], "Escalating Commitment in Individual and Group Decision Making: A Prospect theory Approach", *Organizational Behavior and Human Decision Processes*, Vol. 54, No. 3, April, pp. 430 - 455.
- Yaniv, G., [1988], "Withholding and Non-Withheld Tax Evasion", *Journal of Public Economics*, Vol. 35, No. 2, March, pp. 183 - 204.
- Zahedi, F., [1986], "Group Consensus Function Estimation when Preferences are Uncertain", *Operations Research*, Vol. 34, No. 6, November/December, pp. 883 - 894.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all the people whose sacrifice of time and effort made possible the completion of this odyssey through the landscapes and seascapes of a social science laboratory experiment in taxpayers' risk preferences with respect to compliance. There were so many of you, both inside the University of Canterbury's Department of Accountancy, Finance and Information Systems (AFIS), and beyond.

Inside AFIS, I owe a particular debt of gratitude to Adrian Sawyer, my supervisor, for the hours and days he spent reading my drafts, and for the constructive criticisms and suggested directions of enquiry he afforded. I would also like to thank him for arranging for John Hasseldine of Indiana University, Bloomington, to take on the job of assessing the study as its external examiner.

I am grateful to John Hasseldine, in turn, not only for his efforts as an examiner, but also for the assistance he afforded earlier in proceedings by providing access to several papers produced in the United States which I might otherwise not have been given.

In this respect I also owe a debt of gratitude to Richard Dusenbury of Florida State University. His prompt replies to my several e-mail requests for information were of considerable value; and while my study developed along lines somewhat different in several respects from his own, I have the deepest respect and admiration for the elegance and appositeness of his research design. I would also like to acknowledge Peggy Hite of Indiana University's School of Business (Bloomington Campus) for forwarding a fax of her 1988 working paper in spite of the technical hitches which plagued our communications.

Closer to home I would like to thank Alan Robb of AFIS for the valuable insights he provided at the departmental launching of my research proposal; and for the positive reinforcement he provided then and later. I would also like to thank Alan for the assistance he gave in negotiating the recruitment of the first of the groups of participants who underwent the experiment. This was hard work; but he was unflinchingly optimistic that volunteers could and would be recruited; and he pulled this miracle off. As a consequence of his inspiration, I found

negotiating the recruitment of volunteers from the three other sources of participants a much easier task to undertake. I owe a debt of gratitude to Kerry Jacobs too, for his role in the initial negotiations with the second of the organisations which provided volunteers.

Several other people also were of assistance in the volunteer recruitment area. My thanks to Marilyn Staite, Christine Dann and to Sandra Jamison.

And then there were my AFIS classmates. No laboratory experiment is capable of running itself. I am grateful to Thai Mui-Leng, Julie Hercock, Dane Howarth, Mark Sowden, Mike Houston and James Doyle for acting as session supervisors. Some of these people helped out time after time when they also had theses to write. My debt of gratitude also extends beyond this group to Greg Hughes, who also gave up a lot of his time to work as a session supervisor.

Then there were the people who provided computing assistance. For help with SAS I would like to thank the AFIS technicians, Peter Hinchey and Brendan Queree. For insights into SAS which enabled me to get on top of this software programme, the two people of assistance were Beverly Lord of AFIS and Bill Rea of the University of Canterbury Computer Services Department.

I also owe a debt of gratitude to Frank Lad of the University of Canterbury's Mathematics and Statistics Department for his two tutoring sessions. Similarly, John Fountain and Mike McCosker of the University of Canterbury's Economics Department also provided useful suggestions with respect to extensions, possible and actual. Thanks.

I am also grateful to Neil Graham, who was roped in at the eleventh hour to give this document its final reading for errors, oversights and omissions.

I would also like to acknowledge the work of the University of Canterbury's Human Ethics Committee in vetting the procedures of the laboratory experiment at relatively short notice. It must have been a surprise that this one was proposed by a trainee accountant as distinct from a graduate psychologist.

The other 132 people to whom I owe a debt of gratitude are the taxpayer-participants. Thanks, strangers, for your patience and good humour. It has been a long journey.



Frank begins to seriously question which is worse-- death or taxes.

TABLE OF APPENDICES

APPENDIX A: DUSENBURY'S INSTRUMENT	289
A.1 Dusenbury's Training Script	289
A.2 Dusenbury's Five Scenarios	293
A.3 Additional Rating of Judgment Situations	298
APPENDIX B: INSTRUMENT USED IN THIS STUDY	299
B.1 Preliminary Document	299
B.2 Training Script	300
B.3 The Experimental Instrument	306
B.4 Questionnaire	313
B.5 Tally Sheet	317
B.6 Debriefing Document.	318
APPENDIX C: HOW THE GLM PROCEDURE WORKS	319
APPENDIX D: DATA GROOMING	321
D.1 The Basic Model GLM Procedure	321
D.2 The Repeated Measures GLM Procedure	321
APPENDIX E: VALIDATION STATISTICS	323
E.1 MANOVA Statistics	323
E.2 Basic Model GLM Procedure Results	324
E.3 Repeated Measures Results for Individual Cases	326
APPENDIX F: FURTHER HYPOTHESIS TEST RESULTS	327
F.1 Results For H_{1A}	327
F.2 Results for H_{2A}	328
F.3 Results for H_{3A}	331
APPENDIX G: SPECIMEN TESTS OF THE DEMOGRAPHIC VARIABLES	335

APPENDIX A: DUSENBURY'S INSTRUMENT

This appendix contains the training script as it was designed and used by Dusenbury (1994):

Section A.1 contains Dusenbury's training script (including introductory document).

Section A.2 contains Dusenbury's five scenarios.

Section A.3 contains Dusenbury's additional ratings question.

A.1 DUSENBURY'S TRAINING SCRIPT

[Participants Received a Copy of the Script.]

I. PURPOSE AND NATURE OF THE STUDY

The purpose of the project is to study personal judgments, so there are no right or wrong answers. Therefore, in each situation, choose the option you feel most comfortable with. The session will last no longer than 60 minutes. All your answers will be anonymous. Cash awards will be paid. The amount depends on your choices and on chance outcomes which will be resolved through the use of a game-type spinner. You may receive \$10 or more, but you may receive nothing.

You are not required to participate in this study. You may withdraw if you wish.

II. CASH PAYOFFS

You will get a fund of money to spend to accomplish several tasks. The tasks are to pay income taxes, to buy a Health Care Plan for your family, and to select a gamble. Payments you make reduce your fund. Refunds and receipts increase your fund. Any of the fund not spent will be converted to cash awards at a rate of \$.50 in cash for each \$100 left over at the end of the session.

III. OVERVIEW OF THE TASKS

In the tax situations, the prepayments through withholding have already been made, and your task is to make any payment due, or to claim a refund, at the time of the tax filing. If the amount you choose to report turns out to under-report your tax, you must pay the taxes owed plus \$300. The \$300 is to covers all monetary and nonmonetary costs such as lost work time, professional service fees, anxiety over the deficiency, etc.

In the tax cases, the correct amount of tax is uncertain at the time you file your return. For example, your right to claim a dependency exemption for an adult child may be uncertain because of the facts of the case. To give another example, an activity which you consider a business may later be deemed to be a hobby, and hobby-related losses are not deductible. Or the uncertain item might be the gain on a home where the costs of the home sold are not documented. Please picture the uncertain item to be something for which the correct amount cannot be determined exactly at the time you file your return. You will choose how much (all, most, half, some, or none) of the uncertain item to report, and none of these choices would be illegal choices.

In the health care situation, the task is to purchase a family health care plan. You choose from among five family health care plans which differ in the health services covered. A family member may require services not covered by the Plan you purchase. In that case you will have additional costs for the uncovered medical services needed by a member of your family.

In the gamble situation, you select one gamble from among five available choices. You may either win additional money, or you may have to make a payment from your fund.

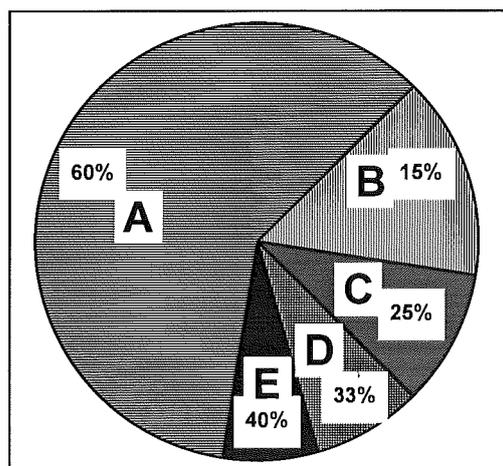
IV. DETERMINING FUTURE OUTCOMES

In reality, the consequences of choices like these would only be known in the future. Here a spinner will be used to determine the unknown outcomes. The spinner exactly reflects the chances stated in the options. For example, in the tax situations Option B has a 15% risk of under-reporting your tax. If you choose Option B, you are determined to have under-reported only if the spinner lands in the first 15% of the wheel. Likewise, if you choose Option C, you are determined to have under-reported only if the spinner lands in the first 25% of the wheel. Option D has a 33% risk of under-reporting your liability, and Option E always has a 40% risk of under-reporting your liability.

The spinner is pictured here.

[Spinner Face with B, C, D, & E regions here.]

[This graphics file was not received; but the following diagram is an approximation of Dusenbury's spinner face. Note that C includes all of the B sector; D includes the C and B sectors; and E includes the D, C and B sectors.]



V. THE PROCEDURES

1. You receive a Package of Cases, a PAYMENT ENVELOPE, and a pack of money.
2. After reading a case, you circle the option you prefer.
 - (a) If that option requires a payment, you enclose the amount of the payment in the PAYMENT ENVELOPE. Then you go on to the next case.
 - (b) If your preferred option is a filing for a refund, you raise your hand to receive the amount of the refund. Then you go on to the next case.
3. After you have made your choices in all the cases, you will then spin the spinner for each case in which you chose an uncertain outcome.
 - (a) If the spin is favorable, you make no further payment.
 - (b) If the spin is unfavorable, you hand over the payment stated in the option.
4. After completing all the spins, any of your fund remaining will be converted to cash awards at the rate of \$.50 for each \$100 left over.

VI. SAMPLE CASE

SAMPLE CASE: DAY CARE INCOME

You are married filing jointly with taxable income of \$25,780 not counting net income from providing day care in your home for several neighborhood pre-schoolers. The amount of this income you should include is uncertain because your costs for lunches, yard fencing, snacks, toys, use of your home, etc. are not documented.

You are now preparing your tax return. Your withholding totals \$5,600. Consider both the risk of under-reporting and the dollar outcomes in choosing the amount of this income to report.

In Option A you include the total earnings. There is no risk of under-reporting your tax (though you may be overpaying). In Options B, C, D, and E, you include less and less in net income, and both your refund and also your chance of under-reporting increases. The risk of under-reporting your tax is 15% in Option B, 25% in Option C, 33% in Option D, and 40% in Option E.

TAX FILING OPTIONS -- Circle One

AMOUNT INCLUDED	(A) ALL	(B) MOST	(C) HALF	(D) SOME	(E) NONE
REFUND CLAIMED	REFUND \$600	REFUND \$700	REFUND \$800	REFUND \$900	REFUND \$1,000
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$200	\$300	\$400	\$500

VII. TRIALS USING THE SAMPLE CASE

To include all of the Day Care Earnings, circle Option A, and raise your hand in order to be paid the \$600 refund. There is no risk of under-reporting in this Option, and you would make no spin.

To include most of the Day Care Earnings, circle Option B and raise your hand to be paid the \$700 refund. After making your choices in all the cases, you would then spin the spinner. If the spinner landed in the wedge with the letter B, you will be considered to have under-reported your true tax. You would have to pay \$200 out of your fund. If the spinner landed anywhere else in the wheel, you would be considered to be not under-reported.

To include half of the Day Care Earnings, circle Option C and raise your hand to be paid the \$800 refund. Later, you would spin the spinner. If the spinner landed in the wedge with the letter C, you will be considered to have under-reported your true tax. You would have to pay \$300 out of your fund.

And so on for Options D and E.

[End of Training Script.]

A.2 DUSENBURY'S FIVE SCENARIOS

[Beginning of Instrument. Each case on separate page.]

TAX CASE Y

The task is to file your income tax return. The correct amount of your income tax is uncertain because the correct amount of one deduction item is in doubt. You must choose the amount of this item to report. Your taxable income is \$27,780 not counting this item. Your withholding is \$5,400.

In Option A you deduct none of this item, and there is no risk of under-reporting your tax. In Options B, C, D, and E you deduct more and more of this item, and both your refund and also your chance of under-reporting increases. The cost if under-reported is unpaid taxes plus \$300. The risk of under-reporting your tax is 15% in Option B, 25% in Option C, 33% in Option D, and 40% in Option E.

TAX FILING OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
AMOUNT DEDUCTED	NONE	SOME	HALF	MOST	ALL
REFUND CLAIMED	REFUND \$400	REFUND \$500	REFUND \$600	REFUND \$700	REFUND \$800
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

Consider both the risk of under-reporting and the dollar amounts in choosing the amount of this item to deduct. Circle the Option you prefer, and raise your hand to receive your refund. Then go on to the next case.

TAX CASE Z

The task is to file your income tax return. The correct amount of your income tax is uncertain because the correct amount of one item of income is in doubt. You must choose the amount of this item to report. Your taxable income is \$25,780 not counting this item. Your withholding is \$4,800.

In Option A you include all this income, and there is no risk of under-reporting your tax. In Options B, C, D, and E, you include less and less of this income, and both your refund and your chance of under-reporting increases. The cost if under-reported is unpaid taxes plus \$300. The risk of under-reporting your tax is 15% in Option B, 25% in Option C, 33% in Option D and 40% in Option E.

TAX FILING OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
AMOUNT INCLUDED	ALL	MOST	HALF	SOME	NONE
REFUND OR PAYMENT DUE	PAYMENT \$200	PAYMENT \$100	NONE \$0	REFUND \$100	REFUND \$200
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

Consider both the risk of under-reporting and the dollar amounts in choosing the amount of this item to deduct. Circle the Option you prefer. If that Option requires a payment, enclose the payment in the PAYMENT ENVELOPE. If that Option has a refund, raise your hand to receive the refund. Then go on to the next case.

TAX CASE X

The task is to file your income tax return. The correct amount of your income tax is uncertain because the correct amount of one deduction item is in doubt. You must choose the amount of this item to report. Your taxable income is \$27,780 not counting this item. Your withholding is \$4,200.

In Option A you deduct none of this item, and there is no risk of under-reporting your tax. In Options B, C, D, and E you deduct more and more of this item, and your payment declines but your chance of under-reporting increases. The cost if under-reported is unpaid taxes plus \$300. The risk of under-reporting your tax is 15% in Option B, 25% in Option C, 33% in Option D, and 40% in Option E.

TAX FILING OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
AMOUNT DEDUCTED	NONE	SOME	HALF	MOST	ALL
PAYMENT DUE	PAYMENT \$800	PAYMENT \$700	PAYMENT \$600	PAYMENT \$500	PAYMENT \$400
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

Consider both the risk of under-reporting and the dollar amounts in choosing the amount of this item to deduct. Circle the Option you prefer, and enclose the payment due in the PAYMENT ENVELOPE. Then go on to the next case.

CHOOSE A GAMBLE

Your task is to choose one of the five gambles below. If you choose Option A, the outcome will be certain, and you will later pay \$200 for sure. In Options B, C, D, and E you will receive either the more favorable outcome OR the less favorable outcome depending on the spinner results. In Options B, C, D, and E, the more favorable outcome is increasingly more favorable, but your chance of getting the more favorable payoff decreases.

GAMBLES -- Circle One

OPTIONS	CHANCE	WIN/ PAY	AMOUNT
(A)	100%	PAY	\$200
(B)	85%	PAY	\$100
	15%	PAY	\$500
(C)	75%	PAY	\$0
	25%	PAY	\$500
(D)	67%	WIN	\$100
	33%	PAY	\$500
(E)	60%	WIN	\$200
	40%	PAY	\$500

Circle your preferred gamble. Then go on to the next case.

HEALTH CARE PLAN

The task is to purchase a Health Care Plan for your family. Your employer pays for part of the plan, and you pay the rest. You choose from among five plans which differ in the medical services covered. Plan A is the full coverage policy. In Options B, C, D and E, you buy Plans which have fewer and fewer covered services. If no one in your family needs any of the uncovered services, then the cheaper Plans provide coverage as good as that of Plan A.

If you choose Plan A, there is no risk of additional costs, but Plan A costs you more. In Options B, C, D, and E your net cost for the Health Plan decreases, but the chance that someone in your family will need uncovered medical services (costing you additional money) increases. The risk of added cost is 15% in Option B, 25% in Option C, 33% in Option D, and 40% in Option E.

HEALTH CARE OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
HEALTH PLANS	PLAN A	PLAN B	PLAN C	PLAN D	PLAN E
YOUR COST	\$800	\$700	\$600	\$500	\$400
RISK AND COST OF UNCOVERED SERVICES					
RISK	0%	15%	25%	33%	40%
YOUR COST	\$0	\$400	\$500	\$600	\$700

The chart shows your cost as well as the risk and cost of needing uncovered services for each Health Care Plan. Circle the Plan you prefer, and put the amount of your cost in the PAYMENT ENVELOPE. Then go on to the next case.

A.3 ADDITIONAL RATING OF JUDGMENT SITUATIONS

On this page, please prepare an additional rating for these five Judgment Situations. For this rating please use each Option (A, B, C, D, and E) once and only once. Assign the A (the least risky option) to the Judgment Situation for which you are least willing to take risk, and so on. This means you should Assign the E (the most risky option) to the Judgment Situation for which you are willing to take the most risk.

Please look back at all the Judgment Situations and compare them to make this additional rating.

USE EACH OPTION (A,B,C,D AND E) ONCE AND ONLY ONCE

SITUATION	OPTION
TAX CASE X	_____
CHOOSE A GAMBLE	_____
TAX CASE Y	_____
BUY HEALTH PLAN	_____
TAX CASE Z	_____

[WHEN YOU HAVE COMPLETED THIS ADDITIONAL RATING, PLEASE RAISE YOUR HAND.]

APPENDIX B: INSTRUMENT USED IN THIS STUDY

This appendix contains the complete set of documents used in the current study. The sections are laid out in the following order:

- Section B.1 contains the preliminary document titled “Purpose and Nature of the Study”.
- Section B.2 contains the training script and sample tax case.
- Section B.3 contains the five replication scenarios, the additional ratings questions used in the Summary Syndrome Extension, and the extension tax scenario, *high pay* Tax Case W, which was used in the Risk Profile Stability Extension.
- Section B.4 contains the end-of-session questionnaire.
- Section B.5 contains the tally sheet.
- Section B.6 contains the end-of session debriefing document titled “What This Was All About (In Brief).”.

B.1 PRELIMINARY DOCUMENT

[This document was required by the University of Canterbury’s Human Ethics Committee.]

PURPOSE AND NATURE OF THE STUDY

The purpose of the project is to study personal judgments, so there are no right or wrong answers. In particular the study is concerned about how people handle tax and medical decisions in situations where the law is not easily definable. Therefore in each situation, choose the option you feel most comfortable with.

The session will last about 60 minutes or up to 90 minutes at the very most. All your answers will be anonymous.

There is a questionnaire to fill out at the end of the session which will be used in the statistical analysis of the experimental results. Neither your personal results nor what you disclose in your answers to the questionnaire will become available to anybody other than the researcher (who only knows you as a number). A summary of results will be available at the end of the project (November) for anyone who wishes to obtain one; but neither individuals nor organisations will be identifiable in the end product. (There are a number of separate organisations contributing participants.)

This work is for a Master of Commerce thesis project involving decision making under uncertainty. While much of the experiment relates to tax payments, there is no connection between the researcher and the IRD other than that he is an ordinary taxpayer with a typical low student income. The IRD has no access to information from this experiment that could identify any individual in any way.

Real cash awards will be paid direct to your organisation. The amount depends on your choices and on chance outcomes which will be resolved through the use of a game-type spinner. The organisation may receive \$10 or more from your personal efforts, but it may receive nothing. You are not required to participate in this study. You may withdraw if you wish.

B.2 TRAINING SCRIPT

CONSENT

You are invited to participate in the research project with the working title, 'Personal Financial Judgements' by completing the following experiment and questionnaire. The aim of the project is to determine how people handle tax and medical decisions in situations where the law is not easily definable. Both parts of this undertaking are anonymous, and you will not be identified as an informant without your consent. You may at any time withdraw your participation, including withdrawal of any information you have provided. By completing the experiment and questionnaire, however, it will be understood that you have consented to participate in the project, and that you consent to publication of the results of the project with the understanding that anonymity will be preserved.

INTRODUCTION

There are a number of different tasks in this experiment:

Pay taxes,
Pay medical insurance,
Select a gamble.

All of these involve making decisions in the dark; and it is up to you how much risk you are prepared to take.

We would like you to imagine you are the person facing the situations in real life and act as you would act normally.

We would also like you to indicate what your second, third, fourth and fifth choices would have been (in order).

To make the payments you decide upon, you have a fund of play money amounting to \$3500 or \$2100. You can think of this as cash in your cheque account.

At the end of the session, the play money you still have is translated into real NZ money at NZ 30 cents per \$100 play money, and your total will be forwarded to [name of organisation supplying recruits] with everyone else's as a lump sum.

You may end up with a positive amount of play money or a negative amount (in debt play money-wise).

If you end up with a negative amount, the NZ 30 cents per \$100 play money is deducted from the [name of organisation]'s lump sum.

We have included the financial incentive in order to make the decisions as real as possible.

After the fifth situation we'd like you to answer a couple of general questions about what you have done so far.

Then there are two more situations and a questionnaire.

THE TASKS

TAX. In most of the situations we would like you to imagine that you are a tax-payer who has a certain income and who has had a certain amount of tax collected already by the tax authorities.

By the way, there is no connection between this experiment and any New Zealand tax body. The only connection between myself and the IRD is that I'm a student who usually gets his tax return in late.

Also the tax regulations in the situations we'd like you to look at are not necessarily real NZ tax regulations; but the tax rate was computed at the 24 cents in the dollar rate to \$30,875 and 33 cents in the dollar above that.

Anyway, you have had some tax collected already by the tax authorities:

This tax may have been the PAYE removed from your pay packet each pay day;
Or this tax may be provisional tax you have had to hand over earlier in the tax year.

Now it is time to file your annual tax return, and your task is to make any payment due, or to claim a refund. The correct amount of tax is uncertain at the time you file your return.

For example, your right to claim a rebate for a child may be uncertain because of the facts of the case.

Or an activity which you consider a business may later be deemed to be a hobby, and hobby-related losses are not deductible.

Please picture the uncertain item to be something for which the correct amount cannot be determined exactly at the time you file your return.

If the amount you choose to report turns out to under-report your tax:

You must pay the taxes owed plus \$300.

The \$300 is to cover all monetary and nonmonetary costs such as lost work time, professional service fees, anxiety over the deficiency, etc.

As that tax-payer, choose how much (all, most, half, some, or none) of the uncertain item to report. **None of these choices would be illegal choices.**

Medical Insurance. Here your task is to make a choice of medical insurance policies. The five policies on offer different levels of coverage.

A family member may require services not covered by the policy you choose.

In that case you will have additional costs for the uncovered medical services needed by a member of your family.

Gamble. In the gamble situation, you select one gamble from among five available choices. You may either win additional money, or you may have to make a payment from your fund.

DETERMINING FUTURE OUTCOMES

In reality, the consequences of choices like these would only be known in the future.

Here a bag containing 100 plastic chips will be used to determine the unknown outcomes. It exactly reflects the chances stated in the options:

Option B has a 15% risk of an unfavourable outcome.

Option C has an unfavourable outcome only if a chip with a number 25 or lower comes out of the bag of 100 chips.

Option D has a 33% risk.

Option E always has a 40% risk.

In the final two situations the risk levels are altered¹:

There are still five choices, but the final two (**D & E**) have higher risks.

Option D has a 50% risk.

Option E has a 57% risk.

¹ This information is relevant to *refund* Tax Case V and *high pay* Tax Case W. Since these cases pertained to a potential extension which was contemplated and then abandoned; this information is redundant in terms of the study both as it was initially conceived and in its final form. The two cases have not been included in this appendix. However, the version of *high pay* Tax Case W containing the unaltered set of risk levels, which was used in the Risk Profile Stability Extension, is provided at the end of Section B.3.

THE PROCEDURES

1. You receive a pack of play money and a large handout containing seven cases and the questionnaire.

2. **After reading a case**, circle the option you prefer.

(a) If that option requires a payment, put this payment into your PAYMENT ENVELOPE

(b) If you file for a refund, the supervisor give it to you.

(c) Then you go on to the next case.

3. **After you have made your choices** in all the cases:

A chip is drawn from the bag containing all 100 chips for each case in which you chose an uncertain outcome:

(a) If the chip has a number that is favourable, you make no further payment.

A favourable chip has a number higher than the level of risk.

(b) If the chip is unfavourable, you hand over the required payment.

An unfavourable chip has a lower or equal value to the level of risk.

4. **After determining all the outcomes by selecting chips**:

Any of your fund remaining will be recorded as a real cash contribution to be handed to the [name of organisation] at the rate of 30 cents for each \$100 of play money.

Or, if your fund ends up negative, the contribution to the [name of organisation] is reduced at the rate of 30 cents per \$100 play money.

SAMPLE CASE: DAY CARE INCOME

You are married filing jointly with **taxable income of \$20,833** — which includes every last cent of what you have earned from a second income source, which was providing day care in your home for several neighbourhood pre-schoolers.

But this of course does not take account of the expenses you incurred earning this day care income.

But the amount of this day care income you should include in your tax return is uncertain. This is because you have lost track of your receipts and other records detailing your costs for lunches, yard fencing, snacks, toys, use of your home, etc.

You are now preparing your tax return.

You have already made provisional tax payments totalling \$5,600, which is an **overpayment by \$600** on the \$20,833 income figure if the day care income is not reduced by some figure to account for the costs.

Consider both the risk of under-reporting and the dollar outcomes in **choosing the amount of this day care income** to report.

Option A: you include the total day care earnings. There is no risk of under-reporting your tax (though you may be overpaying).

In **Options B, C, D, and E**, you include less and less of the day care earnings in your reported total income, and both your refund and also your chance of under-reporting increases.

The risk of under-reporting your tax is:

- 15% in Option **B**,
- 25% in Option **C**,
- 33% in Option **D**,
- 40% in Option **E**.

TAX FILING OPTIONS -- Circle One

AMOUNT INCLUDED	(A) ALL	(B) MOST	(C) HALF	(D) SOME	(E) NONE
REFUND CLAIMED	REFUND \$600	REFUND \$700	REFUND \$800	REFUND \$900	REFUND \$1,000
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
You pay out	\$0	\$200	\$300	\$400	\$500

To include all of the Day Care Earnings, circle **Option A**, and raise your hand in order to be paid the \$600 refund.

There is no risk of under-reporting in this Option, and you would not need to select a chip

To include most of the Day Care Earnings, circle **Option B** and raise your hand to be paid the \$700 refund.

After making your choices in **all the cases**, you would then select chips from the bag.
If the chip has a number lower than or equal to 15 (the percentage level of **B**), you will be considered to have under-reported your true tax.
You would have to pay \$200 out of your fund.
If the chip has a higher number than 15 you would be considered to be not under-reported.

And so on for **Options C, D and E**.

When you have made your choice, then write the letter of your second, third fourth and fifth choice in the place provided:

My second choice would have been: _____
My third choice would have been: _____
My fourth choice would have been: _____
My fifth choice would have been: _____

END OF TRAINING SCRIPT.

B.3 THE EXPERIMENTAL INSTRUMENT

TAX CASE Y

The task is to file your income tax return. The correct amount of your income tax is uncertain because the correct amount of one item you could claim a rebate on is in doubt.

You must choose the amount of this item to report.

Your taxable income is \$24,166 not counting this item. **You have already made PAYE payments totalling \$5,400** which is **\$400 too much** on your income (ignoring the possible rebate).

In **Option A** you do not claim the rebate, and there is no risk of under-reporting your tax.

In **Options B, C, D,** and **E** you claim more and more of the rebate item's cost, and both your refund and also your chance of under-reporting increases.

The tax authorities will consider a disallowed rebate in the same light as under-reported tax. The cost if tax is under-reported is **unpaid taxes plus \$300**.

The risk of under-reporting your tax is shown in the lower part of the table, and the cost to you is shown in the bottom line.

Consider both the risk of under-reporting and the dollar amounts in choosing the amount of this rebate item to claim. Circle the Option you prefer, and raise your hand to receive your refund. Then fill in the list below before moving on to the next case.

TAX FILING OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
AMOUNT DEDUCTED	NONE	SOME	HALF	MOST	ALL
REFUND CLAIMED	REFUND \$400	REFUND \$500	REFUND \$600	REFUND \$700	REFUND \$800
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

My second choice would have been: _____

My third choice would have been: _____

My fourth choice would have been: _____

My fifth choice would have been: _____

TAX CASE Z

The task is to file your income tax return. The correct amount of your income tax is uncertain because you are not sure how much exactly you earned on one item of income. You must choose the amount of this item to report.

Your taxable income is \$25,780 not counting this item. You have already made provisional tax payments totalling \$4,800.

In **Option A** you include all possible income from the uncertain item, and there is no risk of under-reporting your tax.

In **Options B, C, D, and E**, you include less and less of this uncertain item income, and both your refund and your chance of under-reporting increases.

The cost if under-reported is unpaid taxes plus \$300.

The risk of under-reporting your tax is shown along the bottom of the table along with the possible costs.

Consider both the risk of under-reporting and the dollar amounts in choosing the amount of this item to deduct. Circle the Option you prefer. If that Option requires a payment, make the payment into the PAYMENT ENVELOPE.

If your Option has a refund, your supervisor will give you the refund.

After filling in your order of preference in the place for it below the box, go on to the next case.

TAX FILING OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
AMOUNT INCLUDED	ALL	MOST	HALF	SOME	NONE
REFUND OR PAYMENT DUE	PAYMENT \$200	PAYMENT \$100	NONE \$0	REFUND \$100	REFUND \$200
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

My second choice would have been: _____
 My third choice would have been: _____
 My fourth choice would have been: _____
 My fifth choice would have been: _____

TAX CASE X

The task is to file your income tax return. The correct amount of your income tax is uncertain because the correct amount of one business expense is in doubt.

You must choose the amount of this expense to report. **Your taxable income is \$40,000 — NOT counting this expense.**

You have already made provisional tax payments totalling \$9,621, which is \$800 short of what you would be expected to pay on \$40,000 if the expense is ignored.

In **Option A** you deduct none of this expense, and there is no risk of under-reporting your tax.

In **Options B, C, D, and E** you deduct more and more of this expense, and your payment declines but your chance of under-reporting increases.

The cost if under-reported is unpaid taxes plus \$300.

The risks and costs of under-reporting your tax are laid out at the bottom of the table. Consider both the risk of under-reporting and the dollar amounts in choosing the amount of this item to deduct. Circle the Option you prefer, and put your tax payment in your PAYMENT ENVELOPE.

After filling in your preference order in the place below the box, go on to the next case.

TAX FILING OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
AMOUNT DEDUCTED	NONE	SOME	HALF	MOST	ALL
PAYMENT DUE	PAYMENT \$800	PAYMENT \$700	PAYMENT \$600	PAYMENT \$500	PAYMENT \$400
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

- My second choice would have been: _____
- My third choice would have been: _____
- My fourth choice would have been: _____
- My fifth choice would have been: _____

CHOOSE A GAMBLE

Your task is to choose one of the five gambles below.

If you choose **Option A**, the outcome will be certain, and you will later pay \$200 for sure.

In **Options B, C, D, and E** you will receive either the more favourable outcome OR the less favourable outcome depending on the spin of the wheel at the end of the session.

In **Options B, C, D, and E**, the more favourable outcome is increasingly more favourable, but your chance of getting the more favourable payoff decreases.

Circle your preferred gamble. (No money is paid or given till the end when the chips are drawn.) Then go on to the next case.

GAMBLES -- Circle One

OPTIONS	CHANCE	WIN/ PAY	AMOUNT
(A)	100%	PAY	\$200
(B)	85%	PAY	\$100
	15%	PAY	\$500
(C)	75%	PAY	\$0
	25%	PAY	\$500
(D)	67%	WIN	\$100
	33%	PAY	\$500
(E)	60%	WIN	\$200
	40%	PAY	\$500

My second choice would have been: _____

My third choice would have been: _____

My fourth choice would have been: _____

My fifth choice would have been: _____

PAYING MEDICAL INSURANCE

The task is to pay medical insurance to cover your family. Your employer pays for part of the plan, and you pay the rest.

You choose from among five insurance policies which differ in the medical services covered.

Policy A is the full coverage policy.

In **Options B, C, D** and **E**, you opt for policies which have fewer and fewer covered services. If no one in your family needs any of the uncovered services, then the cheaper policies provide coverage as good as that of **Policy A**.

If you choose **Policy A**, there is no risk of additional costs, but **Policy A** costs you more.

In **Options B, C, D**, and **E** the cost of medical insurance goes down for you. But (as luck would have it!) the chance that someone in your family will need uncovered medical services (costing you additional money) increases.

The risk of added cost is **15%** in Option **B**, **25%** in Option **C**, **33%** in Option **D**, and **40%** in Option **E**.

The chart shows your cost as well as the risk and cost of needing uncovered services for each Insurance Policy Plan. Circle the Plan you prefer, and put your insurance premium in your payment envelope. After filling in your preference order in the place below the box, go on to the next case.

HEALTH CARE OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
POLICIES	PLAN A	PLAN B	PLAN C	PLAN D	PLAN E
YOUR COST	\$800	\$700	\$600	\$500	\$400
RISK AND COST OF UNCOVERED SERVICES					
RISK	0%	15%	25%	33%	40%
YOUR COST	\$0	\$400	\$500	\$600	\$700

- My second choice would have been: _____
- My third choice would have been: _____
- My fourth choice would have been: _____
- My fifth choice would have been: _____

ADDITIONAL RATINGS OF JUDGMENT SITUATIONS

On this page, please prepare three additional ratings for five of these Judgment Situations.

For this rating please **use each rating (1,2,3,4 and 5) once and only once**. Assign the 1 (the least risky option) to the Judgment Situation for which you are least willing to take risk, and so on. This means you should assign the 5 (the most risky option) to the Judgment Situation for which you are willing to take the most risk.

Please look back at all five of these Judgment Situations and compare them to make this additional rating.

Use each rating (1,2,3,4,5) once and only once:

TAX CASE X _____
 CHOOSE A GAMBLE _____
 TAX CASE Y _____
 PAY MEDICAL INSURANCE _____
 TAX CASE Z _____

Please indicate how significant you considered the facts of these two scenarios to be to you by ticking the most appropriate box:

My choice of option was influenced by the level of PAYE said to be withheld in **Tax Case Y**:

- Not at all — I ignored it.
- Only in a slight fashion
- It was significant in my thinking
- It was quite important
- Of major importance — it totally shaped my thinking.

My choice of option was influenced by the level of provisional tax said to be withheld in **Tax Case X**:

- Not at all — I ignored it.
- Only in a slight fashion
- It was significant in my thinking
- It was quite important
- Of major importance — it totally shaped my thinking.

Now please move on to the final two cases.²

[_____ Page Break _____]

² Only the version of *high pay* Tax Case W used in the risk Profile Stability Extension is reprinted here.

TAX CASE W

The task is to file your income tax return. The correct amount of your income tax is uncertain because the correct amount of depreciation claimable on a major piece of equipment is in doubt.

You must choose the amount of this depreciation to report. **Your taxable income is \$41,148** — if this depreciation expense is **NOT** brought in to reduce it.

You have already paid provisional tax totalling \$10,000, which is **\$800 short** of what you would be expected to pay on \$41,148 if the expense is ignored.

In **Option A** you deduct none of this expense, and there is no risk of under-reporting your tax.

In **Options B, C, D, and E** you deduct more and more of this expense, and your payment declines but your chance of under-reporting increases.

The cost if under-reported is **unpaid taxes plus \$300**.

The risks and costs of under-reporting your tax are laid out at the bottom of the table. Consider both the risk of under-reporting and the dollar amounts in choosing the amount of this item to deduct. Circle the Option you prefer, and put your tax payment in your PAYMENT ENVELOPE.

TAX FILING OPTIONS--Circle One

	(A)	(B)	(C)	(D)	(E)
AMOUNT DEDUCTED	NONE	SOME	HALF	MOST	ALL
PAYMENT DUE	PAYMENT \$800	PAYMENT \$700	PAYMENT \$600	PAYMENT \$500	PAYMENT \$400
RISK & COST OF UNDERPAYMENT					
RISK	0%	15%	25%	33%	40%
TOTAL COST	\$0	\$400	\$500	\$600	\$700

Now please go on to the final scenario; or if the one on this page is the last, please go on to the questionnaire.

B.4 QUESTIONNAIRE

The purpose of this questionnaire is to provide a background of solicited opinions for statistical comparison with the responses the entire group has made to the five payment-making scenarios.

The answers you make are completely confidential; and you personally will not be able to be identified in any way.

Please tick only one box for each question:

1. Gender:

- Male Female

2. Age:

Please tick the box which covers your age group:

- 18 - 20 21 - 25
 26 - 30 31 - 40
 41 - 60 >60

3. Tax filing experience in years:

Please tick the box which covers the approximate number of years you have filed tax returns in New Zealand or elsewhere:

- 1 - 5 6 -10
 11 - 15 16 - 20
 >20

4. Attitude to health insurance:

Given the current state of the public health system in New Zealand, how worthwhile do you consider a personal health insurance scheme to be for you or other members of your family?:

Please tick the box which most closely approximates your judgement of it:

- It is ridiculous
 It is not very necessary
 I am neither for nor against
 I think it is a reasonable thing to have
 I would not be without it

5. Attitude to fairness of tax with respect to changes in the public health system:

Given the level of tax levied at present in New Zealand, do you consider that changes in the public health system are:

- unacceptable
 have some, but overall questionable merit
 undecided
 are basically OK, but have some flaws
 acceptable

6. Attitude to non-compliance #1:

Given the tax laws that are in place in New Zealand at the present time, would you consider it acceptable for a taxpayer earning over \$35,000 to under-report income that he/she would have the power to conceal?

- never
 seldom
 on average
 mostly
 always

7. Attitude to non-compliance #2:

Given the tax laws that are in place in New Zealand at the present time, would you consider it acceptable for a taxpayer earning less than (say) \$18,000 to under-report income that he/she would have the power to conceal?

- never
 seldom
 on average
 mostly
 always

8. Feeling about actual taxpaying:

When you have found yourself in the position of owing more tax to the IRD, which single box of the following would come closest to most accurately describing what you have experienced:

- I am very unhappy: This is a loss to me and an imposition
 This is something I feel moderately unhappy about; but it is necessary
 No strong feeling
 I'm moderately happy to do the right thing
 I'm happy to fulfil this requirement: it is my contribution to society

9. Frequency of refund-due and payment-due tax filings:

When you have filed your annual income tax return a rough estimate of the times you have had to make a tax payment to the IRD or similar tax authority elsewhere would be:

- never 20% of years filed
 40% of years filed 60% of years filed
 80% of years filed 100% of years filed
 Don't know

10. Perception of likelihood of an audit by the tax authorities:

At present you believe that the IRD fully audits (as distinct from just checking the arithmetic) what proportion of tax returns that cross its desks? Tick the box which most closely approximates your own estimate:

- 1 percent 5 percent 10 percent
 15 percent 20 percent 25 percent
 30 percent 35 percent 40 percent
 45 percent 50 percent 55 percent
 60 percent 65 percent 70 percent
 75 percent 80 percent 85 percent
 90 percent 95 percent 100 percent

11. Feelings about Gambling:

This question is about how you view the rights and wrongs of gambling. Your attitude to spending say \$5 a week on a Lotto ticket is most closely approximated by which single box:

- I would never buy a ticket because gambling is always wrong
- I am happy to support a charity that funds good causes; but I consider regular ticket-buying to be not right for me
- I do not have an opinion either way
- I am quite happy to have a regular flutter of this sort; but if Lotto was banned, I would not miss it.
- I feel good about playing Lotto and would resent it if it were banned

12. Household income #1:

How would you describe the household in which you live (including benefits/ pensions/ official income support):

- The household has one income-earner
- The household has two income earners
- Other

13. Household income #2:

If total household income is defined as your income and the income of your partner/spouse added together then a ball-park figure of your household income would be in the range of:

- Less than \$20,000
- \$20,000 to \$30,000
- \$30,001 to \$40,000
- \$40,001 to \$50,000
- > \$50,000.

14. Nature of Income Earning Experience:

Which of the following best describes your major income-earning experience:

- State service or SOE employee
- Private sector employee
- Self employed (No employees - with exception that spouse may work with you)
- Self employed (You also employ non-family members)
- I am not directly an earner since my main role is managing the household or caring for the family.
- Other: Please state: _____

15. Education:

For this question, tick every box which applies to you. Which of the following describes your educational history:

- School education
- Completed a polytechnic qualification
- Completed a university qualification
- Completed some other form of tertiary education qualification
- Self-taught in your area of expertise

16. Awareness of general compliance with tax laws:

You believe (or have at least the gut feeling) that taxpayers in general, where they have freedom of action to do so, will tend to:

- Fully report all taxable income to the very last cent
- Report all taxable income give or take the last few dollars
- Sometimes under-report taxable income
- Usually under-report taxable income
- Under report as much and as often as they possibly can.

17. Professional tax-filing assistance:

This question is about whether you seek professional advice from an accountant or other tax return preparer. (If you own/manage a company and have responsibility for the company tax return, treatment of the company return is covered in this question.):

- I always prepare my tax return (or with my spouse/partner)
- I sometimes pay a professional to prepare my tax return
- I usually pay a professional to prepare my tax return
- I always pay a professional to prepare my tax return

18. Knowledge of theories of decision making under risk and under uncertainty #1:

How would you describe your knowledge of Von Neumann and Morgenstern's *Expected Utility Theory*:

- I was totally unaware of it
- I have heard the name but don't know the details
- I am moderately aware of the basic concepts
- I have studied this theory in the past.

19. Knowledge of theories of decision making under risk and under uncertainty #2:

How would you describe your knowledge of Kahneman and Tversky's *Prospect Theory*:

- I was totally unaware of it
- I have heard the name but don't know the details
- I am moderately aware of the basic concepts
- I have studied this theory in the past.

B.5 TALLY SHEET

The name of the organization recruiting the participant was printed on these forms. It has been replaced by XXX in this instance.

[Portion for the study's records]

Case	Option chosen	Risk level	Chip drawn	
Tax Case Y				
Medical Insurance				
Tax Case X				
Gamble				
Tax Case Z				
Tax Case V				
Tax Case W				
AMOUNT IN FUND AT END:				
Amount owed to Organisation XXX				

[Tear-off portion for the individual participant.]

AMOUNT IN FUND AT END:	
Amount owed to Organisation XXX	

B.6 DEBRIEFING DOCUMENT.

[This was required by the University of Canterbury's Human Ethics Committee.]

WHAT THIS WAS ALL ABOUT (IN BRIEF)

As you will have gathered from the experience you have just had, the main issue in this experiment was how people evaluate the choices they have to make in various financial contexts. You have just worked through five tax scenarios, a medical insurance scenario, and a gamble. The research question was, do people treat these quite different financial undertakings in an identical fashion, or do they approach each in a measurably different fashion? In particular the scenarios were set up in such a way that the predictions of *Expected Utility Theory* (from the discipline of Economics) and *Prospect Theory* (from the discipline of Psychology) could be tested in a tax-paying context.

As stated at the start of the session, there is **no connection** between this experiment and any New Zealand tax authority; and your anonymity is fully guaranteed.

Nevertheless you have the right to withdraw from the experiment even at this point if you feel uncomfortable about having participated. If you withdraw, your data, of course, cannot be used.

To complete the data I need to complete the experiment, please fill in the questionnaire.

The full experiment will not be processed and written up until November. When it is in its final form I will be happy to explain the overall findings in depth and detail with any of you who want a proper follow-up.

[Name and university address of researcher stated here.]

APPENDIX C: HOW THE GLM PROCEDURE WORKS

According to its creators, SAS's GLM procedure works as follows. Suppose an experiment involves one dependent variable (responses from participants) and one categorical variable (participants by category). Let there be four categories of participant: this means each category is represented by a level of the variable. The GLM procedure uses the fact that analysis of variance models can be written as linear models of the type¹:

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \dots + \beta_4 X_{4i} + \varepsilon_i$$

This states that observation i of the dependent variable Y is related to one, some or all observations i of the independent variables, X on the expression's right hand side. However, since there is only one categorical variable on the right hand side, and it contains four levels, the GLM procedure creates a dummy or *indicator* variable for each level of the categorical variable, denoted as A_1 , A_2 , A_3 and A_4 . The equation for a given observation may therefore be written²:

$$Y_i = \beta_0 + \beta_1 A_{1i} + \beta_2 A_{2i} + \beta_3 A_{3i} + \beta_4 A_{4i} + \varepsilon_i$$

In this expression, β_0 is the intercept, and only one of the four $\beta_k A_{ki}$ take a non-zero value. For the sake of simplification, let there be two responses per level of the categorical variable. GLM creates the set of indicator variables³:

INTERCEPT	A1	A2	A3	A4
1	1	0	0	0
1	1	0	0	0
1	0	1	0	0
1	0	1	0	0
1	0	0	1	0
1	0	0	1	0
1	0	0	0	1
1	0	0	0	1

The sums of squares derived from the processing of all observations i are computed by the GLM procedure and divided by the available degrees of freedom to produce an F statistic which measures, for the purpose of hypothesis testing, changes in the mean of the dependent variable associated with changes in levels of the categorical variable.

¹ The SAS Procedures Guide for Personal Computers (Version 6 Edition), Volume 1, p. 24.

² Idem.

³ Idem. While the Guide does not specifically state this, the procedure probably drops the fourth dummy variable on the ground that the fourth level of the categorical variable is implied by zero values given to the other three dummy variables.

APPENDIX D: DATA GROOMING

D.1 THE BASIC MODEL GLM PROCEDURE

In order to run the *basic model* GLM procedure, the raw data was sorted into a set of variables as follows. The participants' choice of response to each of the decision problems was labelled CASEDAT, which was the procedure's dependent variable. Since (in terms of the three replication hypotheses and the cash flow extension hypothesis) there were five scenarios, CASEDAT contained 660 unique observations — five for each of the 132 participants. This bundling of the responses into one variable entailed creating a categorical variable, CASENO to distinguish between the five scenarios. CASENO was the important independent variable in the procedure.

If the GLM procedure produced a significant F statistic with respect to CASENO, then a significant difference was said to exist between the means of the CASEDAT observations as sorted by CASENO.

In order to make the data-processing design workable, the 132 observations for group origin were sequentially repeated to yield five identical sets of observations totalling 660 (one for each category of CASENO) for the categorical variable, GROUP; and the same was done for session attended (SESSION), and order of presentation of cases (CASEORD). Because session-related main or interaction effects were not expected if group-related ones were not found, testing for session-related effects was restricted to instances in which a significant effect was either found or predicted with respect to GROUP.¹

D.2 THE REPEATED MEASURES GLM PROCEDURE

The *repeated measures* model was set up differently. Because this particular GLM procedure repeats an analysis of not one, but a set of dependent variables, in terms of a set of categorical variables, it was not necessary to collate subject responses as one variable. This meant that CASEDAT was dropped in favour of a series of dependent variables, each with 132 observations.

Since there was no longer any need for the classification variable, CASENO, to distinguish between cases, it was also dropped; and the length of the entire data set was standardised at 132 observations. As for the *basic model* GLM procedure, the *repeated measures* GLM involved monitoring of main and interaction effects in terms of the classification variables, GROUP, SESSION and CASEORD. The *repeated measures* procedure furnished F -statistics for both *between-subjects* and *within-subjects* effects; but when the *basic model* procedure was used, a separate run was required for each set in order that the appropriate error term would be reported.

¹ This becomes apparent only when attention is turned to tests of the first three hypotheses, since validation tests of the high cash float subsample are not reported in Section 8.4 of Chapter Eight.

The variables were labelled C1 to C5 as shown here²:

Variable	Case
C1	Refund Tax Case Y
C2	High Pay Tax Case X
C3	Medical Insurance Case
C4	Low Pay Tax Case Z
C5	Gamble

Note that when responses to questions from the end-of-session questionnaire were coded as variables, the variable was named after the number of the question. For instance, when the responses to Question 13 were used in a repeated measures GLM procedure, the categorical variable was called Q13.

² Initially only variables relating to the conceptual replication of Dusenbury (1994) were considered. Later, as and when they were required, the sets of observations of first choices made in the extension tax cases, *Refund Case V* and *High Pay Case W*, were incorporated as c6 and c7 respectively. Note that *Refund Case V* was ultimately discarded.

APPENDIX E: VALIDATION STATISTICS

This appendix contains further validation statistics not reported in the text. These include *repeated measures* GLM procedure MANOVA results. They invariably accorded with the *repeated measures* GLM univariate statistics. The format of this appendix is as follows:

- Section E.1 contains MANOVA results.
- Section E.2 contains basic model GLM results for the full sample and the low cash float subsample.
- Section E.3 contains univariate results provided by the SAS *repeated measures* GLM procedure with respect to the individual decision problems.

E.1 MANOVA STATISTICS

TABLE E.1
MANOVA Output for All Five Cases.
Full Sample (N = 132)

Manova Test Criteria and Exact F Statistics for the Hypothesis of no CASE Effect:

H = Type III SS&CP Matrix for CASE E = Error SS&CP Matrix S=1 M=1 N=38

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.6601	10.0396	4	78	0.0001
Pillai's Trace	0.3399	10.0396	4	78	0.0001
Hotelling-Lawley Trace	0.5149	10.0396	4	78	0.0001
Roy's Greatest Root	0.5149	10.0396	4	78	0.0001

Manova Test Criteria and F Approximations for the Hypothesis of no CASE*GROUP Effect:

H = Type III SS&CP Matrix for CASE*GROUP E = Error SS&CP Matrix S=3 M=0 N=38

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.8568	1.0356	12	206.6601	0.4174
Pillai's Trace	0.1464	1.0262	12	240	0.4252
Hotelling-Lawley Trace	0.1633	1.0432	12	230	0.4100
Roy's Greatest Root	0.1370	2.7403	4	80	0.0342

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Manova Test Criteria and F Approximations for the Hypothesis of no CASE*CASEORD Effect:

H = Type III SS&CP Matrix for CASE*CASEORD E = Error SS&CP Matrix S=4 M=6.5 N=38

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.4469	0.9758	72	309.0729	0.5373
Pillai's Trace	0.7248	0.9959	72	324	0.4933
Hotelling-Lawley Trace	0.8993	0.9555	72	306	0.5815
Roy's Greatest Root	0.3122	1.4048	18	81	0.1521

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Manova Test Criteria and F Approximations for the Hypothesis of no CASE*GROUP*CASEORD Effect:

H = Type III SS&CP Matrix for CASE*GROUP*CASEORD E = Error SS&CP Matrix S=4 M=12 N=38

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.2896	0.9862	116	312.5359	0.5267
Pillai's Trace	1.0379	0.9787	116	324	0.5463
Hotelling-Lawley Trace	1.5052	0.9926	116	306	0.5099
Roy's Greatest Root	0.6557	1.8314	29	81	0.0179

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

It is clear from this MANOVA that there is a significant CASE effect, which tells us that the means of the responses to the five decision problems of the replication exercise are significantly different at better than the one percent level of probable error ($Pr > F = 0.0001$), while the *within-subjects* interaction effects remain statistically insignificant.

E.2 BASIC MODEL GLM PROCEDURE RESULTS

The *basic model* GLM procedure was applied to both the full sample and to the various subsamples as determined by levels of the cash float variable. It provided stronger evidence of an intrinsic difference among the five replication scenarios than did the *repeated measures* GLM procedure. The following table contains results for the full sample:

TABLE E.2
Tests for Main and Interaction Effects on the Five Replication Scenarios.
Full Sample (N = 132)¹

PANEL A

Between-Subjects Main and Interaction Effects Computed with the Basic Model GLM Procedure²

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	10.2395	3.4132	2.02	0.1100
CASEORD	18	29.6157	1.6453	0.97	0.4888
GROUP*CASEORD	29	52.6364	1.8150	1.07	0.3634
Error	609	1029.2093	1.6900		

PANEL B

Within-Subjects Main and Interaction Effects Computed with the Basic Model GLM Procedure

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	4	49.6754	12.4188	7.76	0.0001
CASENO*GROUP	15	26.1678	1.7445	1.09	0.3633
CASENO*CASEORD	90	126.8580	1.4095	0.88	0.7654
CASENO*GROUP*CASEORD	145	208.1003	1.4352	0.90	0.7783
Error	405	648.1107	1.6003		

No between-subject effects are statistically significant; but CASENO is strongly significant with only a one in ten thousand probability of error. There is no significant impact from group origin, case order or any multiplicative effect in which these are combined with each other or

¹ Note that the configuration required for this procedure involved treating all observations recorded for all five decision problems as observations on the single variable *CASEDAT* ($132 * 5 = 660$); and the categorical variable, *CASENO* was introduced to differentiate between decision problems. Hence, the degrees of freedom add up to 659 for the two procedures reported in the table.

² Note that the between-subjects procedure and the within-subjects procedure were processed separately. While *F*-statistics associated with Type I sums of squares were calculated, along with those associated with Type III sums of squares, in this procedure, only the Type III SS statistics are reported here.

with CASENO. This indicates that the means of the different sets of decision problem responses are significantly different from each other.

A similar pattern emerges from the output obtained from *basic model* procedures run on the low cash float subsample:

TABLE E.3
Tests for Main and Interaction Effects on the Five Replication Scenarios.
Low Cash Float Subsample (N = 66)

PANEL A

Between-Subjects Main and Interaction Effects Computed with the Basic Model GLM Procedure

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	4.5919	1.5306	0.90	0.4396
CASEORD	15	33.1567	2.2104	1.31	0.1975
GROUP*CASEORD	19	39.2538	2.0660	1.22	0.2397
Error	292	494.5000	1.6935		

PANEL B

Within-Subjects Main and Interaction Effects Computed with the Basic Model GLM Procedure

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	4	35.7636	8.9409	6.47	0.0001
CASENO*GROUP	15	24.5616	1.6374	1.18	0.2903
CASENO*CASEORD	75	117.5276	1.5670	1.13	0.2601
CASENO*GROUP*CASEORD	95	146.1912	1.5389	1.11	0.2796
Error	140	193.5000	1.3821		

It is clear that none of the *between-subjects* main or interaction effects were statistically significant; and that the only within-subject effect to attain significant status, as expected, was CASENO.

E.3 REPEATED MEASURES RESULTS FOR INDIVIDUAL CASES

This is a sample only. This particular table was constructed from results obtained with respect to the low cash float subsample only.

TABLE E.4

**Univariate Output of the Repeated Measures General Linear Model Procedure:
Responses to Individual Cases.**

Low Cash Float Subsample (N = 66)

Refund Tax Case Y (C1)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	4.3888	1.4629	1.40	0.2622
CASEORD	15	14.5148	0.9677	0.93	0.5455
GROUP*CASEORD	19	7.6889	0.4047	0.39	0.9821

High pay Tax Case X (C2)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	1.2264	0.4088	0.30	0.8267
CASEORD	15	21.4981	1.4332	1.04	0.4442
GROUP*CASEORD	19	26.7000	1.4053	1.02	0.4672

Medical Insurance Case (C3)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	1.6119	0.5373	0.36	0.7814
CASEORD	15	30.4279	2.0285	1.36	0.2316
GROUP*CASEORD	19	54.4815	2.8674	1.93	0.0558

Low pay Tax Case Z (C4)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	5.2804	1.7601	1.69	0.1910
CASEORD	15	19.6342	1.3089	1.26	0.2895
GROUP*CASEORD	19	17.8421	0.9391	0.90	0.5835

Choose a Gamble Case (C5)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	12.0541	4.0180	2.04	0.1311
CASEORD	15	31.4527	2.0968	1.06	0.4277
GROUP*CASEORD	19	39.4786	2.0778	1.05	0.4395

APPENDIX F: FURTHER HYPOTHESIS TEST RESULTS

This appendix contains further results concerning the three hypotheses tested in this study with respect to replicating Dusenbury (1994).

Both the *basic model* GLM and the *repeated measures* GLM analysis of variance procedures were used in the testing of all three of these hypotheses, producing very similar results.

Section F.1 contains H_{1A} results.

Section F.2 contains H_{2A} results.

Section F.3 contains H_{3A} results.

F.1 RESULTS FOR H_{1A}

Results obtained from the *basic model* GLM procedure for the sample and each cash float subsample are produced in Table F.1. The information reported is restricted to *within-subject* interactions, and to recording the Shapiro-Wilks' W statistic, which was generated from an analysis of the residuals obtained in the *basic model* procedure.¹ This analysis was achieved via SAS's independent UNIVARIATE procedure.

The F statistic reported in Panel A for the categorical variable CASENO is significant at the two percent level for the full sample while main and interaction effects relating to group and case order remain insignificant at the five percent level.

In Panel B, CASENO is even more strongly significant at close to the 0.3 percent level of error, with respect to the low cash float subsample. Again there are no significant group or case order interaction effects.

In Panel C the response difference traceable to framing effects caused by scenarios is the least significant variable. This variable, CASENO, has the lowest F -statistic and the second-to-highest level of error.

In Panel D, there is no meaningful difference between the two tax cases (as reported by CASENO) at all. This is also true in Panel E. These results concur with those obtained from the *repeated measures* GLM analysis of variance procedures reported in Chapter Eight.

¹ Note that in Table F.1 and in all ensuing tables using the same format, that the number of degrees of freedom for the model sum to one digit less than double the number of participants providing observations (here 263 for 132 people). This is because each person contributed two observations used as CASEDAT data-points. Hence the model, with respect to Table F.1 has 264 observations and 263 degrees of freedom. Information concerning the Shapiro-Wilk W statistic was computed not in the *GLM* procedure, but in a related *UNIVARIATE* procedure executed in the same SAS programme. The statistic was inserted here for convenience.

TABLE F.1
Refund Tax Case Y versus High Pay Tax Case X:
Basic Model GLM Analysis of Variance Results.

PANEL A: Full Sample (N = 132)		Shapiro-Wilks		$W^2 = 0.9541$;	Pr <
W = 0.0001					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	7.1045	7.1045	5.28	0.0228
CASENO*GROUP	6	12.2491	2.0415	1.52	0.1754
CASENO*CASEORD	36	38.6907	1.0747	0.80	0.7831
CASENO*GROUP*CASEORD	58	71.9836	1.2411	0.92	0.6310
Error	162	217.9119	1.3451		
PANEL B: Low Cash Float (N = 66)		Shapiro-Wilks		W = 0.9219; Pr < W = 0.0001	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	11.4107	11.4107	9.45	0.0033
CASENO*GROUP	6	5.6152	0.9359	0.78	0.5927
CASENO*CASEORD	30	36.0129	1.2004	0.99	0.4940
CASENO*GROUP*CASEORD	38	34.3889	0.9050	0.75	0.8250
Error	56	67.6000	1.2071		
PANEL C: Combined High Cash Float (N = 66)		Shapiro-Wilks		W = 0.9073; Pr < W = 0.0001	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	0.2532	0.2532	0.15	0.7013
CASENO*GROUP	6	16.4848	2.7475	1.61	0.1608
CASENO*CASEORD	34	41.5276	1.2214	0.72	0.8496
CASENO*GROUP*CASEORD	34	50.1678	1.4755	0.87	0.6690
Error	56	95.4167	1.7039		
PANEL D: \$3,500 Cash Float (N = 36)		Shapiro-Wilks		W = 0.8911; Pr < W = 0.0001	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	0.0057	0.0057	0.00	0.9558
CASENO*GROUP	4	16.4537	4.1134	2.24	0.0918
CASENO*CASEORD	22	32.3796	1.4718	0.80	0.6974
CASENO*GROUP*CASEORD	18	12.2769	0.6821	0.37	0.9832
Error	26	47.6667	1.8333		
PANEL E: \$4,200 Cash Float (N = 30)		Shapiro Wilks		W = 0.8050; Pr < W = 0.0001	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	2.1275	2.1275	1.28	0.2721
CASENO*SESSION	6	5.6004	0.9334	0.56	0.7540
CASENO*CASEORD	28	40.4303	1.4439	0.87	0.6375
CASENO*SESSION*CASEORD	6	13.7366	2.2894	1.38	0.2750
Error	18	29.8333	1.6574		

F.2 RESULTS FOR H_{2A}

Table F.2 contains the *repeated measures* GLM analysis of variance procedure results for the \$3,500 cash float subsample and the \$4,200 cash float subsample, which were not printed in

² The SAS Procedures Guide for Personal Computers, Version 6 Edition, p. 350, describes the Shapiro Wilks W statistic in the following terms: "The W statistic is the ratio of the best estimator of the variance (based on the square of a linear combination of the order statistics) to the usual corrected sum of squares estimator of variance. W must be greater than zero and less than or equal to one, with small values of W leading to rejection of the null hypothesis [that the input data values are a random sample from a normal distribution.]"

Chapter Eight. Note that these results were obtained in the absence of partitioning by the variable, *attitude to private medical insurance*, generated from the participants' responses to Question 4 of the end-of-session questionnaire. The null form of H_{2A} cannot be rejected.

TABLE F.2
High Pay Tax Case X versus Medical Insurance Case:
General Linear Models Procedure Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects.

PANEL A: \$3,500 Cash Float Subsample

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.4281	0.4281	0.26	0.6212
CASE*GROUP	2	6.4852	3.2426	1.94	0.1831
CASE*CASEORD	11	24.8224	2.2566	1.35	0.2997
CASE*GROUP*CASEORD	9	6.3750	0.7083	0.42	0.8995
Error(CASE)	13	21.7250	1.6712		

PANEL B: \$4,200 Cash Float Subsample

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	0.0026	0.0026	0.00	0.9735
CASE*SESSION	3	4.2922	1.4307	0.64	0.6090
CASE*CASEORD	14	16.7097	1.1936	0.53	0.8597
CASE*SESSION*CASEORD	3	0.0806	0.0269	0.01	0.9981
Error (CASE)	9	20.1667	2.2407		

PANEL C: Means Information Both Samples

Response Variable		Sample	N	Mean	Std Dev
High Pay Tax Case X	(C2)	\$3,500	36	2.4444	1.1819
Medical Case	(C3)	\$3,500	36	2.2500	1.1557
High Pay Tax Case X	(C2)	\$4,200	30	2.8333	1.3153
Medical Case	(C3)	\$4,200	30	2.4667	1.2243

To test H_{2A} with the *basic model* GLM analysis of variance procedure, the variable CASEDAT was restricted to observations relating to choices made in *high pay* Tax Case X and in the medical insurance case, consequently the categorical variable, CASENO possessed only two levels. The results obtained prior to partitioning by responses made to Question 4, *Attitude to private medical insurance*, are contained in Table F.3. Neither CASENO nor any of the interaction effects are significant at the five percent probability of error benchmark, although the three-way interaction effect, CASENO*GROUP*CASEORD is weakly significant with respect to the low cash float subsample at the 8.4 percent error level.

TABLE F.3
High Pay Tax Case X versus Medical Insurance Case:
Basic Model GLM Analysis of Variance Results.

PANEL A: Full Sample (N = 132)

Shapiro-Wilks W = 0.9558; Pr < W = 0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	1.0115	1.0115	0.61	0.4355
CASENO*GROUP	6	7.9314	1.3219	0.80	0.5721
CASENO*CASEORD	36	48.8701	1.3575	0.82	0.7537
CASENO*GROUP*CASEORD	58	91.9262	1.5849	0.96	0.5651
Error	162	268.0690	1.6547		

PANEL B: Low Cash Float (N = 66)

Shapiro-Wilks W = 0.9159; Pr < W = 0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	0.2711	0.2711	0.19	0.6649
CASENO*GROUP	6	2.8383	0.4731	0.33	0.9179
CASENO*CASEORD	30	51.9260	1.7309	1.21	0.2639
CASENO*GROUP*CASEORD	38	81.1815	2.1364	1.49	0.0844
Error	56	80.0667	1.4298		

PANEL C: Combined High Cash Float (N = 66)

Shapiro-Wilks W = 0.9297; Pr < W = 0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	1.4286	1.4286	0.77	0.3830
CASENO*GROUP	6	14.9282	2.4880	1.35	0.2522
CASENO*CASEORD	34	46.3246	1.3625	0.74	0.8278
CASENO*GROUP*CASEORD	34	40.2710	1.1844	0.64	0.9164
Error	56	103.4500	1.8473		

PANEL D: \$3,500 Cash Float (N = 36)

Shapiro-Wilks W = 0.9274; Pr < W = 0.0004

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	0.4281	0.4281	0.25	0.6223
CASENO*GROUP	4	9.8979	2.4745	1.44	0.2501
CASENO*CASEORD	22	37.8100	1.7186	1.00	0.4976
CASENO*GROUP*CASEORD	18	7.6825	0.4268	0.25	0.9983
Error	26	44.7833	1.7224		

PANEL E: \$4,200 Cash Float (N = 30)

Shapiro-Wilks W = 0.93202; Pr <

W = 0.0030

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	2.4500	2.4500	1.25	0.2719
CASENO*CASEORD	28	34.9667	1.2488	0.64	0.8818
Error	30	58.6667	1.9556		

Table F.4 contains the between-subjects effects recorded when the full sample was partitioned with respect to responses made to Question 4, *Attitude to private medical insurance*. No main or interaction effects achieved statistical significance. The *within-subject* results are reported in Table 8.15 in Chapter Eight.

TABLE F.4
High Pay Tax case X versus the Medical Insurance Case
Partitioned by Q4 (Attitude to Health Insurance).
Full Sample (N = 132)
General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

PANEL A

Participants who Support Private Medical Insurance (N = 104)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	0.5266	0.1755	0.12	0.9457
CASEORD	17	16.6061	0.9768	0.69	0.8012
GROUP*CASEORD	21	42.3603	2.0172	1.42	0.1428
Error	62	87.8679	1.4172		

PANEL B

Participants indifferent to, or against Private Medical Insurance (N = 28)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	7.1845	2.3948	0.59	0.6470
CASEORD	12	20.2339	1.6862	0.42	0.9008
GROUP*CASEORD	7	8.8460	1.2637	0.31	0.9196
Error	5	20.2500	4.0500		

Repeated measures GLM procedures were performed on the \$3,500 and \$4,200 cash float subsamples partitioned by the Question 4 variable, *Attitude to private medical insurance*; but no statistically significant results were obtained. Furthermore, three of the four subsamples, obtained in the partitioning, were too small for there to be sufficient degrees of freedom for the computation of meaningful results.

F.3 RESULTS FOR H_{3A}

The results obtained from basic model GLM analysis of variance procedure are contained in Table F.5. The CASENO *F*-statistics recorded in Panels A to D are supported at levels of significance which allow for the rejection of the null form of H_{3A} in the case of the full sample, the low cash float subsample, the combined high cash float subsample, and the \$3,500 cash float subsample.

In Panel E, CASENO also furnishes a weakly significant *F*-statistic with respect to the \$4,200 cash float subsample ($F = 4.09$, $Pr > F = 0.0522$). However, the associated Shapiro-Wilks statistic indicates that the probability that the distribution of the residuals could be described as normal has a much higher level of error ($Pr < W = 0.0113$).

TABLE F.5
Low Pay Tax Case Z and the Gamble:
Basic Model GLM Analysis of Variance Results.

PANEL A: Full Sample (N = 132)		Shapiro-Wilks		W = 0.96566;	Pr <
W = 0.0004					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	36.7463	36.7463	20.44	0.0001
CASENO*GROUP	6	11.7730	1.9622	1.09	0.3697
CASENO*CASEORD	36	62.6092	1.7391	0.97	0.5282
CASENO*GROUP*CASEORD	58	82.1955	1.4172	0.79	0.8510
Error	162	291.2143	1.7976		
PANEL B: Low Cash Float Subsample (N = 66)		Shapiro-Wilks		W = 0.9090;	Pr < W = 0.0001
W = 0.0001					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	22.4902	22.4902	14.95	0.0003
CASENO*GROUP	6	17.3345	2.8891	1.92	0.0935
CASENO*CASEORD	30	51.0869	1.7029	1.13	0.3376
CASENO*GROUP*CASEORD	38	57.3207	1.5084	1.00	0.4891
Error	56	84.2667	1.5048		
PANEL C: Combined High Cash Float (N = 66)		Shapiro-Wilks		W = 0.9219;	Pr < W = 0.0001
W = 0.0001					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	27.8263	27.8263	13.91	0.0004
CASENO*GROUP	6	14.7790	2.4632	1.23	0.3042
CASENO*CASEORD	34	63.4240	1.8654	0.93	0.5791
CASENO*GROUP*CASEORD	34	39.3344	1.1569	0.58	0.9552
Error	56	112.0000	2.0000		
PANEL D: \$3,500 Cash Float (N = 36)		Shapiro-Wilks		W = 0.8721;	Pr < W = 0.0001
W = 0.0001					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	15.9298	15.9298	8.70	0.0066
CASENO*GROUP	4	4.5128	1.1282	0.62	0.6547
CASENO*CASEORD	22	21.2178	0.9644	0.53	0.9345
CASENO*GROUP*CASEORD	18	8.4757	0.4709	0.26	0.9979
Error	26	47.5833	1.8301		
PANEL E: \$4,200 Cash Float (N = 30)		Shapiro-Wilks		W = 0.9421;	Pr < W = 0.0114
W = 0.0114					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	8.7781	8.7781	4.09	0.0522
CASENO*CASEORD	28	76.0500	2.7161	1.26	0.2639
Error	30	64.4167	2.1472		
PANEL F: \$4,200 Cash Float (with SESSION)		Shapiro Wilks		W = 0.8533;	Pr < W = 0.0001
W = 0.0001					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASENO	1	6.0261	6.0261	2.47	0.1338
CASENO*SESSION	6	6.0803	1.0134	0.41	0.8597
CASENO*CASEORD	28	70.1624	2.5058	1.03	0.4896
CASENO*SESSION*CASEORD	6	14.2796	2.3799	0.97	0.4707
Error	18	44.0000	2.4444		

However, when the classification variable, SESSION, was taken into account in Panel F, the difference in means between the Low Pay Tax Case Z and the Gamble Case disappeared. That the inclusion of SESSION produced a better information set was reflected in the GLM procedure's higher R^2 statistic: 0.7100 versus 0.5755 in the absence of SESSION. It is to be noted that SESSION contributes no interaction effect.

The between-subjects effects recorded in the repeated measures GLM procedure performed on the \$4,200 cash float subsample are reported in Table F.6.

TABLE F.6
Low Pay Tax case Z versus the Gamble Case:
General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects.
\$4,200 Cash Float Subsample (N = 30)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
SESSION	3	1.8221	0.6074	0.18	0.9051
CASEORD	14	51.2662	3.6619	1.10	0.4537
SESSION*CASEORD	3	10.4140	3.4713	1.05	0.4179
Error	9	29.8333	3.3148		

The failure of the \$4,200 cash float subsample to reject the null form of H_{3A} was examined with respect to Question 11, *Attitude towards gambling*, in terms of the *basic model* GLM procedure. According to the results shown in Table F.7, Q11 produced a weakly significant main effect on the \$4,200 cash float subsample (Panel B), but no effect of any credibility on the full sample (Panel A).

TABLE F.7
Responses to the Gamble Case versus Responses to Question 11:
Basic General Linear Model Analysis of Variance.

PANEL A					
Full Sample (N = 132)		Shapiro Wilks' W = 0.87173; Pr < W = 0.0001			
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q11	4	10.7973	2.6993	1.33	0.2613
Error	127	257.1724	2.0250		
PANEL B					
\$4,200 Cash Float Subsample (N = 30)		Shapiro Wilks' W = 0.911808; Pr < W = 0.0181			
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q11	3	19.6667	6.5556	2.56	0.0765
Error	26	66.5000	2.5577		

A basic model GLM procedure also found that Question 6, *Attitude to non-compliance by higher income earners*, Question 7, *Attitude to non-compliance by lower income earners*, and Question 8, *Feeling about actual taxpaying*, produced no significant main effects. This is shown in Table F.8.

TABLE F.8

Responses to Low Pay Tax Case Z versus Responses to Questions 6, 7 and 8:
Basic General Linear Model Analysis of Variance.

PANEL A

Full Sample (N = 132)

Shapiro-Wilks' W = 0.939776; Pr < W = 0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q6	4	5.4182	1.3545	1.11	0.3547
Q7	4	4.5109	1.1277	0.92	0.4520
Q8	4	4.6056	1.1514	0.94	0.4410
Error	119	145.1017	1.2193		

PANEL B

\$4,200 Cash Float Subsample (N = 30)

Shapiro Wilks' W = 0.899357; Pr < W = 0.0083

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q6	2	0.6130	0.3065	0.17	0.8478
Q7	3	2.3417	0.7806	0.42	0.7378
Q8	4	0.8620	0.2155	0.12	0.9748
Error	19	34.9603	1.8400		

APPENDIX G: SPECIMEN TESTS OF THE DEMOGRAPHIC VARIABLES

A large number of repeated measures GLM procedures incorporating categorical variables furnished by the end-of-session questionnaire were necessary for the compilation of results written up in Chapter Ten. The following tables represent one small branch of these investigations by way of an example. The variables chosen were *gender* (Q1), *age* (Q2), and *education level* (Q15). Of these, *age* (Q2) is pursued to a deeper level. To conserve space, the Greenhouse-Geisser Epsilons and the Huynh-Feldt Epsilons have been discarded.

TABLE G.1
Tax Cases X, Y and Z, and Gender, Age and Educational Level.
Full Sample (N = 132)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q1	1	7.9928	7.9928	4.58	0.0350
Q2	5	14.9741	2.9948	1.71	0.1388
Q1*Q2	5	2.3083	0.4617	0.26	0.9315
Q15	4	6.3113	1.5778	0.90	0.4653
Q1*Q15	4	2.4920	0.6230	0.36	0.8388
Q2*Q15	12	14.9858	1.2488	0.72	0.7334
Q1*Q2*Q15	6	4.0371	0.6728	0.39	0.8868
Error	94	164.1757	1.7465		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	7.6494	3.8247	3.64	0.0282
CASE*Q1	2	0.4922	0.2461	0.23	0.7916
CASE*Q2	10	13.9150	1.39150	1.32	0.2208
CASE*Q1*Q2	10	6.3169	0.6317	0.60	0.8122
CASE*Q15	8	7.2451	0.9056	0.86	0.5506
CASE*Q1*Q15	8	9.3420	1.1677	1.11	0.3582
CASE*Q2*Q15	24	41.4883	1.7287	1.64	0.0360
CASE*Q1*Q2*Q15	12	13.5644	1.1304	1.07	0.3838
Error(CASE)	188	197.7632	1.0519		

TABLE G.2
Tax Cases X, Y and Z, and Gender, Age and Educational Level.
Low Cash Float Subsample (N = 66)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q1	1	7.7807	7.7807	4.89	0.0334
Q2	5	9.9702	1.9940	1.25	0.3051
Q1*Q2	3	3.7665	1.2555	0.79	0.5078
Q15	4	8.5475	2.1369	1.34	0.2729
Q1*Q15	4	4.8092	1.2023	0.76	0.5608
Q2*Q15	10	18.9910	1.8991	1.19	0.3272
Q1*Q2*Q15	1	0.0542	0.0542	0.03	0.8546
Error	36	57.2659	1.5907		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	12.8747	6.4373	7.48	0.0011
CASE*Q1	2	0.3214	0.1607	0.19	0.8301
CASE*Q2	10	22.9187	2.2919	2.66	0.0080
CASE*Q1*Q2	6	1.9081	0.3180	0.37	0.8960
CASE*Q15	8	6.7137	0.8392	0.98	0.4624
CASE*Q1*Q15	8	7.7831	0.9730	1.13	0.3537
CASE*Q2*Q15	20	18.5355	0.9268	1.08	0.3918
CASE*Q1*Q2*Q15	2	0.2697	0.1349	0.16	0.8553
Error(CASE)	72	61.9698	0.8607		

The identical procedure was run on the three high cash float subsamples; but the results are not reported.

The issue of *gender* alone was discussed in Section 9.3.2.1. The following four tables record the results obtained from *repeated measures* GLM procedures testing H_{1A} on the full sample and the low cash float subsample, both partitioned by *gender* (Q1).

TABLE G.3
Refund Tax Case Y and High Pay Tax Case X.
All Males Subsample (N = 49)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	8.7924	2.9308	1.89	0.1601
CASEORD	15	35.9864	2.3991	1.54	0.1692
GROUP*CASEORD	7	19.6312	2.8045	1.81	0.1346
Error	23	35.7321	1.5536		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	2.6304	2.6304	1.45	0.2412
CASE*GROUP	3	0.4275	0.1425	0.08	0.9711
CASE*CASEORD	15	11.6052	0.7737	0.43	0.9543
CASE*GROUP*CASEORD	7	5.3725	0.7675	0.42	0.8783
Error(CASE)	23	41.8036	1.8175		

PANEL C

Variable	N	Mean	Std Dev
C1	49	2.5102	1.1924
C2	49	3.0204	1.3615

TABLE G.4
Refund Tax Case Y and High Pay Tax Case X.
All Females Subsample (N = 83)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	2.5967	0.8656	0.54	0.6582
CASEORD	16	18.6576	1.1661	0.73	0.7518
GROUP*CASEORD	22	30.7663	1.3984	0.87	0.6279
Error	41	65.8333	1.6057		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	4.7098	4.7098	6.55	0.0143
CASE*GROUP	3	2.3099	0.7700	1.07	0.3720
CASE*CASEORD	16	19.6141	1.2259	1.70	0.0849
CASE*GROUP*CASEORD	22	9.9927	0.4542	0.63	0.8748
Error(CASE)	41	29.4810	0.7190		

PANEL C

Variable	N	Mean	Std Dev
C1	83	2.1687	0.9855
C2	83	2.6506	1.1416

TABLE G.5
Refund Tax Case Y and High Pay Tax Case X.
Male Low Cash Float Subsample (N = 19)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	1	5.5104	5.5104	6.01	0.0703
CASEORD	10	18.9924	1.8992	2.07	0.2516
GROUP*CASEORD	2	1.8452	0.9226	1.01	0.4425
Error	4	3.6667	0.9167		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	8.7971	8.7971	3.64	0.1291
CASE*GROUP	1	0.5104	0.5104	0.21	0.6697
CASE*CASEORD	10	15.4773	1.5477	0.64	0.7418
CASE*GROUP*CASEORD	2	0.2738	0.1369	0.06	0.9457
Error(CASE)	4	9.6667	2.4167		

PANEL C

Variable	N	Mean	Std Dev
C1	19	2.1579	1.3023
C2	19	3.1053	1.4101

TABLE G.6
Refund Tax Case Y and High Pay Tax Case X.
Female Low Cash Float Subsample (N = 47)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
GROUP	3	1.5584	0.5195	0.30	0.8247
CASEORD	13	15.8777	1.2214	0.71	0.7362
GROUP*CASEORD	11	8.8734	0.8067	0.47	0.9019
Error	19	32.8583	1.7294		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	1	6.6064	6.6064	11.93	0.0027
CASE*GROUP	3	1.8778	0.6259	1.13	0.3620
CASE*CASEORD	13	10.8779	0.8368	1.51	0.2012
CASE*GROUP*CASEORD	11	5.0402	0.4582	0.83	0.6167
Error(CASE)	19	10.5250	0.5539		

PANEL C

Variable	N	Mean	Std Dev
C1	47	2.1915	0.8505
C2	47	2.8936	1.1274

APPENDIX G: SPECIMEN TESTS OF THE DEMOGRAPHIC VARIABLES

A large number of repeated measures GLM procedures incorporating categorical variables furnished by the end-of-session questionnaire were necessary for the compilation of results written up in Chapter Ten. The following tables represent one small branch of these investigations by way of an example. The variables chosen were *gender* (Q1), *age* (Q2), and *education level* (Q15). Of these, *age* (Q2) is pursued to a deeper level. To conserve space, the Greenhouse-Geisser Epsilons and the Huynh-Feldt Epsilons have been discarded.

TABLE G.1
Tax Cases X, Y and Z, and Gender, Age and Educational Level.
Full Sample (N = 132)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q1	1	7.9928	7.9928	4.58	0.0350
Q2	5	14.9741	2.9948	1.71	0.1388
Q1*Q2	5	2.3083	0.4617	0.26	0.9315
Q15	4	6.3113	1.5778	0.90	0.4653
Q1*Q15	4	2.4920	0.6230	0.36	0.8388
Q2*Q15	12	14.9858	1.2488	0.72	0.7334
Q1*Q2*Q15	6	4.0371	0.6728	0.39	0.8868
Error	94	164.1757	1.7465		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	7.6494	3.8247	3.64	0.0282
CASE*Q1	2	0.4922	0.2461	0.23	0.7916
CASE*Q2	10	13.9150	1.39150	1.32	0.2208
CASE*Q1*Q2	10	6.3169	0.6317	0.60	0.8122
CASE*Q15	8	7.2451	0.9056	0.86	0.5506
CASE*Q1*Q15	8	9.3420	1.1677	1.11	0.3582
CASE*Q2*Q15	24	41.4883	1.7287	1.64	0.0360
CASE*Q1*Q2*Q15	12	13.5644	1.1304	1.07	0.3838
Error(CASE)	188	197.7632	1.0519		

TABLE G.2
Tax Cases X, Y and Z, and Gender, Age and Educational Level.
Low Cash Float Subsample (N = 66)

PANEL A

General Linear Models Procedure Repeated Measures Analysis of Variance:
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Q1	1	7.7807	7.7807	4.89	0.0334
Q2	5	9.9702	1.9940	1.25	0.3051
Q1*Q2	3	3.7665	1.2555	0.79	0.5078
Q15	4	8.5475	2.1369	1.34	0.2729
Q1*Q15	4	4.8092	1.2023	0.76	0.5608
Q2*Q15	10	18.9910	1.8991	1.19	0.3272
Q1*Q2*Q15	1	0.0542	0.0542	0.03	0.8546
Error	36	57.2659	1.5907		

PANEL B

General Linear Models Procedure Repeated Measures Analysis of Variance:
Univariate Tests of Hypotheses for Within Subject Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CASE	2	12.8747	6.4373	7.48	0.0011
CASE*Q1	2	0.3214	0.1607	0.19	0.8301
CASE*Q2	10	22.9187	2.2919	2.66	0.0080
CASE*Q1*Q2	6	1.9081	0.3180	0.37	0.8960
CASE*Q15	8	6.7137	0.8392	0.98	0.4624
CASE*Q1*Q15	8	7.7831	0.9730	1.13	0.3537
CASE*Q2*Q15	20	18.5355	0.9268	1.08	0.3918
CASE*Q1*Q2*Q15	2	0.2697	0.1349	0.16	0.8553
Error(CASE)	72	61.9698	0.8607		

The identical procedure was run on the three high cash float subsamples; but the results are not reported.

The issue of *gender* alone was discussed in Section 9.3.2.1. The following four tables record the results obtained from *repeated measures* GLM procedures testing H_{1A} on the full sample and the low cash float subsample, both partitioned by *gender* (Q1).