

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Developing a Method of Collecting Purchase Probability Data in  
Telephone Interviews**

A thesis presented in partial  
fulfilment of the requirements  
for the degree  
of Masters Of Business Studies  
in Marketing at  
Massey University

Dean Lawrence Hini

1997

## Abstract

Purchase predictions is an important issue for both commercial and academic researchers. The Juster Scale is an eleven point purchase probability scale designed to collect purchase probability data in face-to-face interviews. The Verbal Purchase Probability scale is a variation of the Juster Scale designed specifically for use in telephone interviews.

The main focus of this study was to investigate ways in which the accuracy of predictions obtained using the Verbal Purchase Probability scale. This was achieved by testing two procedures designed to improve predictions: using respondent recall of previous purchase behaviour prior as a guide to making predictions; and a double question procedure where respondents were first asked to make purchase predictions for a longer time period (eight weeks), then for the time period of interest (four weeks). It was found that the technique using respondent recall as a guide was not effective at improving predictions, asking the prediction questions over the two time periods was.

Other findings included; that purchase level predictions could not be made with any less data than, the probability of purchasing any products, the number of product most likely to be bought and the probability of purchasing exactly that number without a significant reduction in the accuracy of the prediction. It was found that using respondent recall to test the accuracy of predictions resulted in significantly understated error. Accurate recall at the time of making a prediction lead to more accurate purchase level predictions being made, but not purchase rate predictions. "Non-users" had significantly larger errors in their purchase level predictions than "users", this lead to a new method of estimating purchase levels by assuming non-users have a zero purchase probability and thus the predicted purchase level is equal to the purchase level of users.

## **Acknowledgement**

I would like to thank Dr Mike Brennan, not only for his wisdom and guidance, but also for his eternal patience and support, without which this may never have come to fruition.

I'd also like to thank Carolyn the encouragement and support she provide throughout it all.

I'd also like to thank the Marketing Department the financial assistance provided towards this study, without which this thesis would not have been possible.

## Table of Contents

<b>1. SUMMARY</b> .....	<b>1</b>
<b>2. INTRODUCTION</b> .....	<b>6</b>
<b>3. LITERATURE REVIEW</b> .....	<b>12</b>
3.1 Introduction .....	12
3.2 The Importance of Being Able to Accurately Predict Behaviour .....	13
3.3 Development of Purchase Probability Measurement.....	14
3.4 Purchase Level Predictions .....	18
3.5 Purchase Probability Collection via Telephone Interviewing.....	21
3.6 Predicting Purchases of Branded Products.....	22
3.7 Using Purchase Probabilities to Calculate Demand Curves.....	23
3.8 Limitation of Tests of The Juster Scale and Verbal Purchase Probability Scale.....	25
3.9 The Accuracy of Recall Data.....	26
3.10 Averaging and Rounding .....	26
3.11 Omissions.....	27
3.12 Telescoping.....	28
3.13 Model of Subjective Probability .....	30
3.14 Effect of Inaccurate Recall on Purchase Probabilities .....	32
3.15 Reducing Recall Errors.....	34
3.16 Alternative Methods of Collecting Behavioural Data .....	38
<b>4. HYPOTHESES</b> .....	<b>42</b>
4.1 Introduction .....	42
4.2 Hypotheses 1, 2 and 3 .....	42
4.3 Hypothesis 4 .....	43
4.4 Hypothesis 5 .....	44
4.5 Hypothesis 6 .....	44
4.6 Hypothesis 7 .....	45
4.7 Hypothesis 8 .....	45
<b>5. METHOD</b> .....	<b>46</b>
5.1 Introduction .....	46
5.2 Sample .....	48

5.3 Response rate .....	49
5.4 Procedure .....	51
5.5 Treatments .....	53
5.6 Computer records .....	57
5.7 Interviewers .....	57
5.8 Instrument .....	59
5.8.1 Juster questionnaires .....	59
5.8.2 Follow-up questionnaires .....	62
5.8.3 Late follow-up interview .....	64
5.9 Limitations .....	67
5.10 Variable Operationalisation .....	68
5.10.1 Error calculations .....	68
5.10.2 Expected value .....	69
5.10.3 Expected value calculation for purchase rate .....	69
5.10.4 Expected value calculation for purchase level .....	70
5.10.5 Actual value .....	71
<b>6. RESULTS .....</b>	<b>75</b>
6.1 Introduction .....	75
6.2 Effect of Recall Assisted and Bounded Prediction Questions on Predictions .....	77
6.2.1 Introduction .....	77
6.2.2 Effect of recall assisted and bounded prediction questions on purchase rate predictions .....	78
6.2.3 Effect of recall assisted predictions and the effect of a bounded prediction question on purchase level predictions .....	80
6.2.4 Summary .....	83
6.3 Effect of Calculation Method of Purchase Level Predictions .....	84
6.3.1 Introduction .....	84
6.3.2 Effect of calculation method on purchase level errors .....	84
6.3.3 Summary .....	87
6.4 Effect of using recall as a surrogate for actual behaviour for testing the accuracy of predictions .....	89
6.4.1 Introduction .....	89
6.4.2 Effect of using recall as a surrogate for actual behaviour for testing the accuracy of purchase rate predictions .....	90
6.4.2 Effect of using recall as a surrogate for actual behaviour for testing the accuracy of purchase level predictions .....	91
6.4.3 Summary .....	93
6.5 Effect of accuracy of prior recall on predictions .....	94
6.5.1 Introduction .....	94
6.5.2 Effect of accuracy of prior recall on purchase rate predictions .....	95
6.5.3 Effect of accuracy of prior recall on purchase level predictions .....	95
6.5.4 Summary .....	97

6.6 Effect of usage .....	98
6.6.1 Introduction .....	98
6.6.2 Effect of usage on purchase rate predictions .....	100
6.6.3 Effect of usage on purchase level predictions .....	101
6.6.4 Summary .....	108
<b>7. DISCUSSION</b> .....	<b>110</b>
7.1 Introduction .....	110
7.2 Effect of Recall Assisted Predictions .....	110
7.3 Effect of the Bounded Prediction Procedure .....	112
7.4 Effect of Purchase Level Estimate Calculations .....	113
7.5 Effect of using Respondent Recall to Calculated Error in Predictions .....	113
7.6 Effect of Recall .....	116
7.7 Effect of usage .....	117
<b>8. CONCLUSIONS</b> .....	<b>119</b>

*REFERENCES*

*BIBLIOGRAPHY*

*APPENDICES*

*Questionnaire*

*Charts*

## List of Figures

<i>Figure 1. Beach, Barnes and Christensen-Szalanski's contingency model of subjective probability judgement.....</i>	<i>8</i>
<i>Figure 2. Probability scale used by Byrnes (1964).....</i>	<i>15</i>
<i>Figure 3. The Juster Scale .....</i>	<i>17</i>
<i>Figure 4. Beach, Barnes and Christensen-Szalanski's contingency model of subjective probability judgement.....</i>	<i>30</i>

## List of Tables

<i>Table 1. Summary of research investigating factors affecting reporting in diaries. ....</i>	40
<i>Table 2. Distribution of number of videos hired.....</i>	49
<i>Table 3. Response rate for the Juster interview.....</i>	49
<i>Table 4. Response rates for the follow up interview.....</i>	50
<i>Table 5. Phases of interviewing.....</i>	52
<i>Table 6. Effect of recall assisted and bounded prediction questions on the predictive accuracy of purchase rate estimates.....</i>	78
<i>Table 7. Effect of recall assisted and bounded prediction questions on the predictive accuracy of purchase level estimates.....</i>	81
<i>Table 8. Effect of purchase level calculation method on calculated error in purchase level.....</i>	86
<i>Table 9. Purchase rate error using computer records and recall as a surrogate for computer records in the error calculations.....</i>	90
<i>Table 10. Effect of using recall as a surrogate for actual behaviour for purchase level estimates.....</i>	92
<i>Table 11. Effect of accuracy of recall on predictive accuracy of purchase rate.....</i>	95
<i>Table 12. Effect of accuracy of recall on predictive accuracy on purchase level.....</i>	96
<i>Table 13. Effect of being a user or non-user on predictive accuracy of purchase rate.....</i>	100
<i>Table 14. Effect of being a user or non-user on predictive accuracy of purchase level.....</i>	102
<i>Table 15. Alternative method of calculating purchase level errors using computer data to determine users and non-users.....</i>	104
<i>Table 16. Alternative error calculation method using recall data to determine users and non-users.....</i>	107

## 1. SUMMARY

Predicting future behaviour is an important issue for both academic and commercial researchers. Traditionally, the prediction of future behaviour has been made with the use of intention scales, however more recently, the predictive validity of purchase probability scales have been tested. It has been found that probability scales can accurately predict behaviour. One purchase probability scale that has received much attention is the Juster Scale, an eleven point scale from 0 to 10 with labels at each of the scale points. Originally designed for predicting purchase rates (the proportion of people that will make at least one purchase in a given time period) in face-to-face interviews, it has been developed to the point where it has been tested in self completion questionnaires (Day, Gan, Gendall and Esslemont, 1991) and adapted for telephone interviews (the Verbal Purchase Probability scale) (Brennan, Esslemont and Hini, 1995; Brennan, Hini and Esslemont, 1994). Furthermore, the application of the Juster Scale has been extended. The Juster Scale has been used to predict purchase levels (the total number of products that will be sold). There have been two major ways in which purchase level estimates have been made. The first is by obtaining purchase probabilities for each quantity of products that may be bought, for example the chances of buying exactly one unit, exact two units and so on. The second method is a formula based on the variables of, the chances that any products will be bought, the number most likely to be bought and the chances that exactly that number will be bought. Using purchase level estimates at various price points has then enabled the construction of demand curves (Brennan, 1995; U 1991).

To test the accuracy of the predictions obtained using the Juster Scale or Verbal Purchase Probability scale the method frequently used has been to reinterview the respondent after the prediction period and ask them how many purchases they made over the prediction period. The number recalled as being purchased is then applied to the formula,  $(\text{Predicted} - \text{Number recall as being purchased}) / \text{Number recalled as being purchased} * 100$ . However, there are several biases that affect recall, namely omissions, averaging and telescoping, and thus may be affecting the error calculations.

The affect of inaccurate recall may have further reaching affects than just the error calculations. The subjective probability literature would suggest that when making a judgement, a respondent will need to recall information in order to make the judgement. Thus inaccurate recall may affect the accuracy of the prediction itself. Therefore, it would seem that improving recall could improve the predictions.

Of the three main recall biases, telescoping appears to have the largest affect. Telescoping describes the over-reporting frequently found in retrospective questioning. It has been found that the predictions obtained using the Juster Scale and Verbal Purchase Probability scale tended to be over-estimates. Conceptually, telescoping in recall and over-predictions are very similar in that in both situations, the respondent is over-reporting. It is claimed that bounded recall is effective at reducing telescoping in recall as it provides cues to the respondent to provide accurate answers (Loftus, Klinger, Smith and Fiedler, 1990). The third factor in Beach, Barnes and Christensen-Szalanski's model of subjective probability is motivation to provide accurate answers. Thus it may be that methods that are effect at reducing telescoping in retrospective questioning (such as bounded recall, that is, asking about a long time period and then the time period of interest) may also be effective at reducing over-estimates in predictions.

Several issues were addressed by this present study. The specific issues investigated were, the effect that asking recall questions prior to asking the prediction question has on predictive errors, the effect that asking a bounded prediction question has on the predictive error, the effect that reducing the questions used to calculate purchase level estimates has on predictive accuracy, the effect that using recall as a surrogate for actual purchase data has on the calculated error in predictions, the effect accuracy of recall has on predictive errors and the effect that usage levels has on predictive errors.

The method used in this study was a telephone interview with 1360 members of a video store. The respondents were assigned to one of six treatment groups. Respondents in the Single Recall group were asked to recall the number of videos they had hired in the previous four weeks and then asked to predict the number that they will hire in the next four weeks. Respondents in the Bounded Recall group were asked to recall the number of videos they

had hired in the previous eight weeks and then the number they had hired in the previous four weeks. They were then asked to predict the number that they would hire in the next four weeks. The Bounded Juster group respondents were asked to predict the number of videos they would hire in the next eight weeks, and then the next four weeks. There were three control groups. Respondents in the control groups were only asked to predict the number of videos they would hire in the next four weeks. With the co-operation of the video store, the computer records were kept which recorded the number of videos each of the respondents hired over the four week period. At the end of the prediction period, all respondents who completed the initial interview were reinterviewed to ask the respondent how many videos they had hired in the previous four weeks (since they were interviewed). After interviewing was completed, the sample was post-stratified on the number of videos hired, that is, after post-stratification in each of the treatment groups, there were exactly the same number of respondents who had hired no videos, one video, two videos and so on. This was performed to remove affect that the number of videos hired may have on predictive accuracy.

It was found that there was little evidence to support the hypothesis that obtaining recall data prior to the probability data would improve the accuracy of the predictions. However, it was concluded that although the results did not support the hypothesis, there was some evidence to suggest that the rationale was appropriate and that the unexpected results found here may be resulting from the product being predicted.

It appears that bounding the prediction question is an effective means of reducing predictive errors. The error in purchase rate and purchase level predictions from the Bounded Juster group were smaller than the errors for the control groups. The two main implications of this is that when making predictions, it appears that bounding the prediction question will improve the prediction and that there is evidence that telescoping is a name given to a general motivation affect that leads to over-statements. Thus a bounding technique may also be effective at reducing over-statement in other question types, such as the area of socially desirable topics where motivational affects may be leading to over reporting.

When the error in predictions using the formula (Predicted - Number from computer records) was compared to the error using the formula (Predicted - Number recalled as being hired), it

was found that when using recall as a surrogate for actual data, the errors calculated were significantly understated. However, although the errors were understated, the effect was consistent across the treatments, thus there was no evidence to suggest that conclusions regarding the relative superiority of experimental treatments would be misleading. The implications of these findings are that claims regarding the absolute accuracy of predictions are possibly over-stated and that more accuracy ways of collecting the recall data is needed. However, claims regarding the relative superiority of treatments are likely to be valid.

In testing the effect that accuracy of recall at the time of making predictions has on predictive accuracy, it was found that while accuracy of recall did not seem to affect purchase rate predictions, those with no error in their recall had significantly less purchase level predictive error than those with error in their recall. The theoretical implication of this result is that it provides evidence to support the hypothesis that recall is part of the subjective probability formulation process, and thus provides evidence to support both the use of Beach, Barnes and Christensen-Szalanski's contingency model of subjective probability judgement in this situation, and also supporting the structure of the model itself. This also provides a generalisable results in that other products which are known to be accurately recalled are likely to be accurately predicted.

The final area tested was the effect that usage has on predictive accuracy. It was found that for purchase rate estimates, level usage did not appear to have any affect. However, for purchase level estimates, those who had not used their video card made significantly larger errors in their predictions than those who had used their video card to hire at least one video. This finding lead to a new method of calculating purchase level estimates. By assuming that non-users have a purchase probability of zero, the purchase level of the market will be equal to the purchase level for users. Reanalysing the data found that if the purchase level was calculated from the purchase probabilities of those who had hired over the four week period, predictive error was greatly reduced. Having established that this alternative method of calculating purchase level reduced predictive error, the data was reanalysed again, using the respondent's recall to define users and non-users. A user was defined as person who at the time of making the predictions recalled hiring at least one video in the previous four weeks, whilst a non-user was a person who recalled they had hire no videos over the previous four

week period. It was again found that errors in purchase level predictions were much less than those calculated using the traditional method.

The implication of these results were that if there were a large proportion of non-users in a sample, then it is likely that using a traditional method of calculating purchase level estimates would be inaccurate, although an alternative calculation method has been developed that could potentially reduce the errors in predictions.

## 2. INTRODUCTION

Market researchers often need to predict what the future demand for a product will be. Practitioners often want to know what the demand for a product will be, while academics often need a surrogate for actual purchasing in-order to test their theories (for example, see Charlett, Garland and Marr 1995). This has traditionally been achieved through the use of purchase intention scales. However, more recently, it has been found that a purchase probability scale could be used to provide better estimates (Juster 1966). The Juster scale, an eleven point purchase probability scale, has been used to predict the purchase rate and purchase level of a wide variety of branded and unbranded products and product classes (for example see Day, Gan, Gendall and Esslemont 1991). Originally designed for face-to-face surveys (Juster 1966), the Juster Scale has since been used in self completion questionnaires (Gendall, Esslemont and Day 1991), and a Verbal Purchase Probability Scale, based on the Juster scale, has been developed for use in telephone interviews (Brennan, Hini and Esslemont 1994; and Brennan, Esslemont and Hini 1995).

To be able to determine the accuracy of predictions made with the Juster scale, a simple formula of the actual number of purchases minus the predicted number of purchases all divided by the actual number of purchases \* 100, has frequently been used (for example see Gendall et al 1991). However, usually, the respondents were reinterviewed at the end of the prediction period and asked how many purchases they had made during the prediction period. This recalled number was then used as a surrogate for the actual number of purchases made in the formula to calculate the accuracy of the prediction. As it is well known that respondent recall is often inaccurate (Sudman and Bradburn 1973), there is good reason to question the validity of using respondent recall to judge the accuracy of predictions. If the respondent's recall was inaccurate, the accuracy of the predictions must be misleading. Inaccurate recall could lead to the assessment of the accuracy of the predictions to appear either more or less accurate than the predictions actually are, for example, if the respondent was simply recalling what they had predicted they would buy, then using this recalled number to assess the predicted number would make the predictions appear to be rather accurate, irrespective of

how accurate the predictions actually were. Furthermore, the results of experimental work could be affected, resulting in false conclusion being drawn. Obtaining predictions using different treatments may give different stimuli to the respondents that may in-turn affect the respondent's recall in the second interview. If the accuracy of the respondents' recall varied with the prediction treatment, then the accuracy for each treatment could be misleading. For example, if two prediction treatments both over-estimated purchasing by 10%, but the first treatment produced stimuli that lead the respondents to over-report in the subsequent interview by 10%, while the respondents in the second treatment accurately recalled their purchases, it would be concluded that the first treatment method gave better predictions, even though it only appears to be more accurate in its predictive ability because the over-prediction was compared to the over-reported recall.

There are three main problems associated with recall; omissions, rounding and telescoping. Omissions refers to an event that is either forgotten or not reported. Omissions usually occur when an event has low level of reinforcement in memory, such as infrequent or low-involvement events, or if the respondent is unaware of the event occurring, such as the behaviour of a family member (Cook 1973; Sudman and Bradburn 1973).

Rounding is where a respondent recalls a quantity rounded to a prototypic value, that is, a commonly used grouping value. For example, four weeks may be rounded to one month, or numbers may be rounded to quantities of five or ten. Rounding responses to prototypic values has been found to lead to significant under and over reporting (Huttenlocher, Hedges and Bradburn 1990).

Related to rounding is averaging, where the respondent estimates a "usual" quantity associated with a convenient time period, and then multiplies this estimate by the number of time periods in question. For example, a respondent may estimate their usual weekly usage, and then multiply this by four to give an answer about monthly usage (Cook 1987).

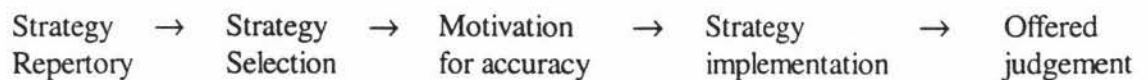
The third effect on recall, telescoping, is considered to be the major source of errors in recall studies (Sudman, Finn and Lannom 1984; Sudman and Bradburn 1973). Telescoping is where events that occurred outside a boundary, such as a point in time, are recalled as

occurring inside the boundary. Telescoping in recall can be both forwards and backwards. In forward telescoping, an event that occurred prior to a point in time is reported as occurring later than it actually did. Backward telescoping is where an event occurring after a point in time is recalled as occurring earlier than it actually did, to be included in the time period. For example, if a person was asked to recall how often they had performed a behaviour last week, backward telescoping would occur if the respondent reported any events that occurred in the current week as occurring last week. Backward telescoping usually is not a concern in recall studies, as the respondent is required to recall events that have occurred since some point in time. The end of the recall period is defined by the recall interview, hence, for backward telescoping to occur, the respondent would need to recall future events. Forward telescoping has however, been found to lead to significant over-reporting (Neter and Waksberg, 1964).

Not only is possible that the use of recall may be leading to false conclusions when it is used as a surrogate for actual behaviour in tests of the accuracy of predictions, but the respondent's inaccurate recall may also lead to the respondent making inaccurate predictions. For example, if predictions are based on the recall of previous behaviour, then if the recall was affected by telescoping, or some other effect that reduces the accuracy, then the predictions based on recall are likely to be inaccurate.

To date, only Beach, Barnes and Christensen-Szalanski (1986) have proposed a non-laboratory model of how subjective probabilities are formulated by respondents. The model can be seen in Figure 1.

Figure 1. Beach, Barnes and Christensen-Szalanski's contingency model of subjective probability judgement



The model begins with the strategies available for assessing a probability. These fall into two categories, knowledge and statistically based. Knowledge based strategies use the respondent's personal knowledge about the event situation, where statistically based strategies

draw on the respondent's knowledge of the frequency of an event. For example, if the best man of a wedding was asked the chances that a randomly selected couple who were recently married would end up getting divorced, it would be expected that the best man would use his knowledge of divorce rates to give the probability that the couple would get a divorce. However, if the same best man was asked the chances that the couple to whom he was the best man for would get a divorce, it would be expected that he would use his personal knowledge of the two people and their relationship to give a response.

Both of these strategies rely upon the respondent's recall to either assess a situation they are personally knowledgeable about, or to estimate the frequency with which an event occurs, whether it is recalling the married couple's relationship or national divorce rates. Accordingly, in the Beach, Barnes and Christensen-Szalanski's model, inaccurate recall would lead to inaccurate probability estimates. If the accuracy of the recall could be improved, then it would be expected that more accurate predictions would result.

To reduce the effects that inaccurate recall could be having on the predictions, and on the tests of the accuracy of the predictions, the recall needs to be improved. Two techniques have been shown to reduce the amount of telescoping (the main source of recall error) in recall. One is to use a landmark event that acts as a clear boundary of the recall period. Loftus and Marburger (1983) found that a landmark event can be a unique event such as the eruption of Mt St Helens, a personal event such as a birthday, or a public event, such as a public holiday, and that these landmark events can be effective at reducing the over-reporting found in recall studies.

The second procedure to reduce telescoping is called bounded recall. The original procedure involved the respondent being interviewed at the start of the recall period and again at the end of the recall period, with retrospective questions asked in each interview on the purchase of durable products and services. With access to the results from the first interview, the interviewer at the second interview could probe the respondents when they reported that they had bought a product or service that had been reported in the first interview to ensure that the purchase had been made since the time of the first interview (Neter and Waksberg 1964). This method of interviewing not only uncovered the telescoping, but also allowed the

researcher to subtract the amount of telescoping from the respondent's recall. The disadvantage was that two interviews were necessary, leading to increased costs and wasted data.

Sudman, Finn and Lannom (1984) extended on the work of Neter and Waksberg (1964) by asking respondents to recall events that occurred in a previous time period (the last month) followed by the current time period (this month), eliminating the need for two interviews. Loftus, Klinger, Smith and Fiedler (1990) used a single interview bounded recall technique similar to that of Sudman, Finn and Lannom (1984), but they used a long time period (the last six months) followed by the time period of interest (the last two months). It was found in both studies that asking about two time periods was effective at reducing the amount that respondents reported for the time period of interest compared to just asking the respondent to recall for just the time period of interest. Loftus et al (1990) also verified the recall of the respondents with actual records. It was found that asking about two time periods did improve the accuracy of the respondents' recall.

Conceptually, the problem of over-reporting in recall is the same as the problem of over-estimation in predictions, that is, the respondents tend to give answers that are too large. It may be that the bias that causes telescoping to occur in recall is the same bias that causes respondents to have a tendency to over-estimate their predictions made when using either the Juster scale (for example see Seymour, Brennan and Esslemont 1994). Therefore, techniques that reduce telescoping in recall may also reduce the over-estimation in predictions.

Loftus et al claim that their results suggest that the single interview bounded recall is effective at reducing recall errors because it "conveys to the respondent that the interviewer desires greater precision in responding" p344. Conveying the interviewer's desire for greater precision in responding could be interpreted as increasing the respondent's motivation for accuracy, which would be consistent with the third factor in Beach, Barnes and Christensen-Szalanski's model. This gives good reason to suggest that bounding the prediction question would improve the prediction.

There are several issues that arise from the literature that this study addresses. One of the

main purposes is to develop methods to improve the predictive accuracy of the Verbal Purchase Probability scale. The recall and subjective probability literature suggest two ways in which the predictions could be improved. These are requiring the respondent to accurately recall their previous behaviour prior to asking the prediction question, and secondly, using a bounded prediction question technique.

The second of the main purposes of this study is to investigate the affect of using recall data as a surrogate for actual behaviour has on the calculated error in predictions, compared with the error in predictions calculated using actual behavioural data.

Other issues investigated include, specifically testing the effect that the accuracy of recall has on predictive accuracy, the effect that usage level has on predictive accuracy and to examine the possibility of reducing the questions needed to calculate purchase level estimates without significantly affect predictive accuracy.

### **3. LITERATURE REVIEW**

#### **3.1 Introduction**

The purpose of this section is to present a review of the literature relating to purchase probabilities, and avenues that suggest methods of improving predictions.

This section begins by establishing the importance of being able to predict behaviour, both to academics, and to commercial researchers.

Having established the importance of being able to accurately predict behaviour, the development of the Juster Scale is traced, from the debate over whether intention or probability scales should be used, to the variations of the Juster Scale that have since been developed.

After examining the developments in the Juster Scale and the Verbal Purchase Probability scale, a general limitation of the studies which have tested and developed the two scales is identified. More specifically, the limitation is that the error in predictions was frequently calculated as  $(\text{Predicted} - \text{Actual})/\text{Actual}$ . However, rather than having actual purchase data to use in the error calculations, respondent recall had been used as a surrogate for actual behavioural data and there are several biases associated with recall which could affect the error calculation. This leads into a discussion on the biases associated with retrospective questioning, namely omissions, averaging and rounding, and telescoping.

Apart from the effect that inaccuracies in recall may have on the error calculations, it may also have an effect on the predictions themselves. A model of subjective probability is presented. This model identifies how recall affects predictions. A discussion on consumer purchasing behaviour follows establishing that future purchasing is related to previous purchasing, validating the need to have accurate recall of previous behaviour at the time predictions are made.

The third factor in the model of subjective probability, that is, motivation to provide accurate answers, is linked back to a method of reducing telescoping in recall (the single interview bounded recall technique), thus leading to a further technique for improving purchase prediction accuracy.

The final part of this section looks at alternatives to using retrospective questioning as a means of collecting behavioural data.

### **3.2 The Importance of Being Able to Accurately Predict Behaviour**

The problem of estimating purchase behaviour is an important issue. Academics often desire actual behaviour to test a theory, although in many cases collecting actual behaviour is neither practical nor possible. As an alternative to collecting actual behavioural data, a reliable method of predicting actual behaviour can be used as a surrogate for actual behaviour (for example, see Charlett, Garland and Marr 1995). Market researchers often want to predict future purchasing to test market demand for a product, or to test the effectiveness of marketing mixes (for example, see Baldwin, Gendall and Hoek 1995). Traditionally, purchase behaviour has been estimated using intention scales where people are asked to what extent they intend to perform the behaviour. However, what the researcher really wants to know is not their intention, but the probability of them purchasing a product.

Conceptually, intentions and probabilities are very different. Intentions are a state of mind, whereas probabilities are a prediction of future events (U 1991). For example, "Do you intend to get caught in a traffic jam tomorrow?" is clearly quite different from "What are the chances that you will be caught in a traffic jam tomorrow?".

However, there is a belief among survey statisticians that typical survey respondents do not understand the notion of probabilities (Byrnes, 1964). Thus intention scales are preferred over probability scales. Hamilton-Gibbs (1986) provides evidence to support the claim that respondents do not understand probabilities. When the respondents were asked their

probability of purchasing various numbers of units, (including the probability of purchasing none), the sum of the probabilities, which theoretically must sum to 1, often did not sum to 1, demonstrating that the respondents were not applying fundamental probability theory in their judgements. This study by Hamilton-Gibbs is discussed in more detail in a later section.

A further advantage of asking people questions such as "Do you intend to buy the product?", is that they can be easily divided into groups of intenders and non-intenders. Marketing programmes can then be tailored to target the appropriate segment. However as Juster (1966) points out, intentions are simply indications of purchase probabilities when it comes to predicting purchases, and furthermore, Wright and Esslemont (1994) question the general usefulness and necessity of segmentation at all.

Despite the general use of intention scales, it is probabilities that researchers are often interested in measuring, not intentions. There was merit in the argument that respondents do not understand probabilities, although this argument has since been addressed in research by Hamilton-Gibbs (1986), Seymour, Brennan and Esslemont (1994), Brennan, Hini and Esslemont (1995) (each of these studies are detailed in section 2.3). Therefore, there is good reason to investigate probability measurement further.

### **3.3 Development of Purchase Probability Measurement**

Ferber and Piskie (1965) were the first to attempt to measure purchase probabilities in a large scale survey. The researchers developed a scale which they called a plan-o-meter to measure respondents' purchase probabilities on a variety of durables and services. The plan-o-meter consisted of an eleven point (0 to 10) scale, with labels of "certain" at 10, "fifty-fifty" at 5, and "no plans at all" at 0.

Respondents were then asked, "Do you plan to purchase any of these goods (whether owned presently or not) between now and ....? Let's take the first one .... How likely are you to purchase it during this period?". The respondent then selected a number from the scale to represent their probability of purchase.

There was a trimodal distribution of responses obtained from the use of the plan-o-meter. This led Ferber and Piskie (1965) to suggest that consumers could only think in three groups of probabilities. Based on this, they believed that the use of a probability scale would provide no more information than a standard intention scale.

However, Juster (1966) claimed that the plan-o-meter was not actually collecting purchase probabilities. In fact, there is little in the plan-o-meter to suggest probabilities are being sorted, and asking about purchasing plans is more akin to asking about purchase intentions than purchase probabilities. Furthermore, as Juster (1966) points out, a respondent with no plans to make a purchase would presumably give an answer of 0 - no plans at all, no matter what their purchase probability was. For frequently purchased goods, consumers may not "plan" to purchase, even though they are aware of a probability of purchasing. Juster (1966) also argued that the trimodal distribution of responses was an artifact of the three labelled points on the plan-o-meter. The three points which are labelled would draw extra attention, which influenced the respondent's choice rather than the respondents possessing limited probability processing abilities.

In 1964 Byrnes set out to test whether measuring purchase probabilities for specific items within a specified time period would be feasible. Respondents were first asked if anybody in their household had purchased a variety of household durables, including an automobile. They were then asked to give the probability of anybody in the household buying the items they had just been asked about over periods of six months, one year and two years. The scale used was an eleven point scale with a descriptive label at each scale point (the scale used can be seen in Figure 2).

Figure 2. Probability scale used by Byrnes (1964)

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

From Juster (1966)

The results showed a close agreement between the expected and observed purchase rates, and therefore the usability of a purchase probability scale.

Although the work by Byrnes (1964) demonstrated that purchase probabilities can be successfully collected, the study did not assess the accuracy of the predictions. The only attempt to validate the predictions was a comparison between the expected purchase probability and recalled prior purchase behaviour. A proper test of the predictions require a comparison between the prediction for a certain time period and the actual purchases made during the same time period.

Analysis showed that there was still a distinct peak at the 5 point, although this was less than that found in the study by Ferber and Piskie (1965). Juster (1966) argued that the peak at the midpoint was due to the descriptive labelling on the scale referring to the chances of either greater or less than even (50-50).

Following on the work of Ferber and Piskie (1965) and Byrnes (1964), Juster (1966) set out to determine if predictions from probability scales are superior to predictions from intention scales.

In July of 1964, as part of the Quarterly Survey of Intentions (QSI), Juster (1966) interviewed 800 households using a conventional intention scale, then a few days later, the same respondents were reinterviewed with the purchase probability scale, now known as the Juster scale (see Figure 3).

Figure 3. The Juster Scale

10	-CERTAIN, PRACTICALLY CERTAIN (99 IN 100)
9	-ALMOST SURE (9 IN 10)
8	-VERY PROBABLE (8 IN 10)
7	-PROBABLE (7 IN 10)
6	-GOOD POSSIBILITY (6 IN 10)
5	-FAIRLY GOOD POSSIBILITY (5 IN 10)
4	-FAIR POSSIBILITY (4 IN 10)
3	-SOME POSSIBILITY (3 IN 10)
2	-SLIGHT POSSIBILITY (2 IN 10)
1	-VERY SLIGHT POSSIBILITY (1 IN 10)
0	-NO CHANCE, ALMOST NO CHANCE (1 IN 100)

from Juster 1966

The results from the use of the Juster scale were promising. It was found that the probability scale explained nearly twice as much of the variance in actual purchases than the intention scale. It was also found that when the purchase probabilities were cross-tabulated with the purchase intentions, there was a wide distribution of purchase probabilities associated with each intention scale point. By doing this, it was found that those who claimed to have "definite" purchase intentions, had a mean purchase probability of around .75, substantially less than 1, as well as those with "no" purchase intentions having a purchase probability significantly greater than 0.

Gruber (1970) undertook a study comparing the distribution of purchase probabilities and purchase intentions. This study also found a wide distribution of purchase probabilities, especially in the middle three purchase intention points. Although actual purchase behaviour was not collected, it was concluded (like Juster, 1966) that the purchase probability scale would be more accurate than intention scales.

### 3.4 Purchase Level Predictions

There had been much work demonstrating the superiority of the Juster scale in predicting the purchase rate, that is, the proportion of people who will make a purchase. For durable items where the respondent would only be making one purchase during the period of interest, the purchase rate is equivalent to the purchase level, that is the total number that will be sold. However, for fast moving goods, such as toothpaste, chickens, butter and the like, respondents could make several purchases of the product within the time period of interest. Therefore, purchase levels could be many times larger than the purchase rate, and purchase level (the number of products that would be sold) is of more interest than the purchase rate (the number of people that would make a purchase).

In order to address this issue, Hamilton-Gibbs (1986) developed two methods using the Juster scale to predict purchase levels. In the first, the multiple question method, the respondents were asked the probability of purchasing exactly no units, one unit, two units, and so on, of a particular product within the next four weeks, until the probability reached zero, that is, a quantity where the respondent claims there is no chance that they would purchase that many. Although the sum of the probabilities must theoretically sum to one, the sum of the probabilities often exceeded one. The responses were then scaled to adjust for this.

The second version which was called the "constant sum" method used ten counters. Each counter represented a one in ten chance of purchasing. The respondent then distributed the ten counters on a showcard with numbers on it, to represent their probability of purchasing that number of units. For example, three counters on the number four would represent a .3 probability of purchasing exactly four units. As the respondent is given ten counters, this forces the probabilities to sum to one, thereby eliminating the need to scale.

Purchase probabilities were found for a variety of fast moving goods for a four week period. Four weeks after the initial interview, respondents were reinterviewed to collect data regarding their purchasing of the products they were previously interviewed about. The actual (recall) data was then compared to the predictions to test the accuracy of the

predictions.

The results proved very encouraging, with most errors  $((\text{predicted} - \text{actual})/\text{actual} * 100)$  less than ten percent. Although the difference in mean predictive error (the mean of the absolute error of each of the products tested) between the two methods tested was not significant, the constant sum method was found to be more accurate overall. Interestingly, the constant sum method consistently overestimated purchases, while the multiple question method consistently underestimated purchases. The under-estimation of the multiple question method may have been due to the repetitive nature of the questioning. Respondents may have learnt that reporting a zero probability reduces the questioning, and so they reported a zero probability.

Following from the findings of Hamilton-Gibbs (1986), research has since looked at the purchase level problem to improve on the method initially devised. Seymour, Brennan and Esslemont (1994) extended on the work by Hamilton-Gibbs (1986) by experimenting with the presentation of the constant sum method.

Two versions of the constant sum method were used, one, the stacked sum method, involved the respondent stacking the counter on top of each other over the number of units. In the flat sum method, the counters were placed beside each other in a row beside the number of units, creating a histogram of probabilities for purchasing the various number of units. As well as the two constant sum methods, a variation on the wording of the multiple question method was also tested. Where Hamilton-Gibbs asked about the probability of purchasing firstly zero units, then one, two and so on until the probability equalled zero, Seymour et al (1993) asked what was the mostly likely number that they would buy in the time period (n), then the probability of purchasing (n-1), (n-2) and so on until a zero probability was reached. The respondent was then asked the probability of purchasing (n+1), (n+2) and so on until a probability of zero was reached. This alteration was designed to eliminate redundant questions, and reduce the laboriousness of the questioning.

Seymour et al's findings supported those of Hamilton-Gibbs, that is, the constant sum method was superior in predicting purchase levels for the limited range of fast moving goods tested, although there was no significant difference in the predictive accuracy of the two constant

sum methods.

In a further extension of the work by Hamilton-Gibbs (1986) and Seymour et al (1994), Brennan, Esslemont and Hini (1995) experimented with the constant sum method. Both methods began the same way. For each of the products, the respondents were asked the chances that they would buy at least one of each of the products. Both methods then used variations of Seymour et al's flat sum method.

In the first method, simply called the constant sum method 1 (CSM1), respondents were required to distribute the 10 counters on a showcard without a zero column. To do this, the respondents were asked to distribute the counters over the quantities in the way that would best represent the chances of them buying the product, if they were to buy any at all. The mean purchase probability was then weighted by the probability of buying any. For example, if a person had a probability of buying any of .7 and the counters were evenly distributed between 2 and 3, their expected number of purchases would be  $.7*((2*.5)+(3*.5))=1.75$  products.

In the second method (constant sum method 2, CSM2), after finding the respondent's probability of purchasing at least one product, the probability of buying zero products can be calculated as one minus the probability of buying at least one. The interviewer then placed the number of counters that represents the probability of buying zero units in the zero column. The respondent was then given the remaining counters to distribute over the number of units to represent their probability of purchasing the various number of units for each particular product. Overall, it was found that CSM2 provided the most accurate predictions, followed by the multiple question method and then CSM1.

An alternative method of calculating the purchase probability was also examined. Usually, the expected number is calculated as the sum of the probability of buying  $n$  units, multiplied by  $n$ , for all values of  $n$  with a non-zero probability of purchasing. In the alternative method, purchase levels were estimated based on the most likely number of units to be bought, the probability of purchasing exactly that number and the probability of buying any. This method of estimating purchase levels was found to produce more accurate results for five out of six

cases for both the constant sum methods, and for four out of six cases for the multiple question method. This method of calculating purchase levels using a reduced set of questions is a significant departure from the traditional method. The analysis by Brennan, Esslemont and Hini (1995) demonstrated that the formula used to calculate purchase levels does not have to be logically complete to be an accurate estimator. This then leads to the suggestion that it may be possible to further reduce the questions used to calculate purchase level estimates.

### **3.5 Purchase Probability Collection via Telephone Interviewing**

All of the studies mentioned thus far have been face-to-face studies, in which the respondent has a copy of the Juster scale in front of them to refer to as they answer. However, as telephone interviewing has become the main method of data collection for large samples (Aaker & Day, 1990), there is the need for a method of obtaining purchase probability data via telephone interviews. Two studies have looked at this issue of administering the Juster scale via telephone interviews.

A preliminary study by Brennan, Hini and Esslemont (1994) compared three methods of purchase probability data collection by telephone; an 11 point 0-10 verbal scale, a branched 11 point verbal scale and a pre-posted Juster scale. The verbal purchase probability scale involved reading a short description of the Juster Scale to the respondents. The description was the same as the description usually read to respondents in face-to-face interviews, plus an description of two additional scale points. In the Branched method, respondents were asked if the chances were over or under a 5 in 10 chance. They were then asked if the chances were between 5 and 7 or 8 and 10 chances in 10, or 0 and 2 or 3 and 5 chances in 10. Having reduced the probabilities available to the respondent to a range of .3 (for example from .8 to 1.0), the respondent was asked for a final probability from within that probability range. The third treatment, the Pre-post group, were sent a copy of the Juster Scale prior to the interview, the survey then proceeded in the same manner as the face-to-face interviews.

It was concluded that a branched method was too confusing and laborious for both the interviewer and respondent, but there was little difference in predictive accuracy between the verbal probability scale and the pre-posted Juster scale (although the verbal probability scale consistently produced more accurate predictions). As it was only a preliminary study using small sample sizes, this study was replicated using larger sample sizes, and only two data collection methods, the verbal probability scale and the pre-posted Juster scale (Brennan, Esslemont and Hini 1995). Again the two methods performed equally well, with the error of most predictions on the variety of fast moving goods of less than 20%. It was found that the pre-posted Juster scale performed marginally better than the verbal scale. Considering the similarity in predictive accuracy, it was concluded that the preferred method for a telephone survey would be the verbal probability scale as it is cheaper to use as there is no mailing costs, and can be used with random-digit dialling as it eliminates the need to generate a sample beforehand.

### **3.6 Predicting Purchases of Branded Products**

Despite the large amount of research that has gone into developing and testing the Juster scale to predict product classes, only two have looked at predicting branded products. Brennan and Esslemont (1994) tested the Juster scale for predicting the rate of purchase for two product classes, tinned soup and yoghurt.

In this study, the error was calculated as the predicted minus the actual. In this error calculation, the difference between the predicted and actual was not divided by the actual. The benefit is that errors can be calculated for each individual respondent. If the difference between the predicted and actual was divided by the actual, the error can not be calculated for each individual respondent because if an individual did not make a purchase, then the actual would be zero, and obviously dividing by zero is undefined. The benefit of calculating the error for each individual respondent is that the significance between errors of any treatment groups can be calculated (although significance tests were not used in the Brennan and Esslemont study). If the errors were calculated at the aggregate level, that is, the total predicted purchase quantity minus the total actually bought, the significance of the difference

in errors can not be calculated as there would be no distribution of errors (as there would be only one aggregate error) which is needed to calculate the standard error. The main disadvantage of calculating individual errors without dividing the difference between the predicted and actual by the actual is that the reported errors are not relative to the amount being purchased. For example, if the predicted was 100.5 products and the actual was 100 products, the error between the predicted and actual is 0.5 of a product, dividing this number by the actual gives an error of only 0.005%, or one half of one percent. However, if the predicted was 0.6 products and the actual was 0.1 products, the difference between the predicted and actual is still 0.5 of a product, but dividing this number by the actual gives an error of 500% error. When the actual purchasing is small, the error using only predicted minus actual results in errors that appear to be quite small, when they may be quite large relative to the actual level of purchasing.

In the Brennan and Esslemont (1994) study, although the largest error was less around fifteen percentage points, the results were described as disappointing. However, brand share was calculated as "the number of people buying the brand divided by the sum, over all brands, of the number brands, of the number of people buying each brand." (p50). This procedure resulted in a mean absolute error over all brands of only 3.1 percentage points. This result lead to another method of estimating the purchase rate being developed, based on brand share. By multiplying the estimated brand share for a brand, by the predicted number of purchases of that product class, an estimate of the number of purchases can be made. It was found that by using this method, the mean absolute error over all brands was just 1.7 percentage points.

One other study that has investigated predicting branded products (U 1991) is discussed in the following section (3.7).

### **3.7 Using Purchase Probabilities to Calculate Demand Curves**

While the previous study looked at predicting the purchase rate for branded goods, three studies have looked at using the Juster scale to construct demand curves, one for a

commodity, one for branded fast moving consumer goods, and for a new product.

Assendelft and Lewis (1991) used the Juster scale to construct demand curves for raw wool. The method involved asking the Hamilton-Gibbs purchase level style of question at various price levels. This results in an estimated number of wool bales that would be bought at various price points. A linear relationship between points was then fitted, forming the demand curve. Although at the time the study was published, the actual accuracy of the demand curves had not been established, it did demonstrate a method that could be used for commodity goods.

U (1991) looked at constructing a demand curve for branded fast moving consumer goods. The same method as used by Assendelft and Lewis (1991) was used by U (1991). Only two products were investigated, Coca Cola and Campbell's Red and White label soup. While Coca Cola was predicted very accurately, the prediction for Campbell's soup was highly inaccurate. It was suggested that the variation in predictive ability may have been due to how different the respondents perceive the products to be from other substitutes. For example, it may be that Coca Cola is perceived to be quite different from its substitutes, but Campbell's Red and White label soup may not be considered that different from other tinned soups. Respondents may have been considering Campbell's (or even any tinned soup) in general when giving their purchase probabilities, resulting in the massive overestimation. Belson (1986) has demonstrated that respondents do, sometimes, ignore boundaries placed in survey questions and so the respondent may have been considering tinned soup in general, rather than Campbell's Red and White label soup.

More recently, Brennan (1995) tested the validity of the demand curves for a new product, laser disc players. To construct the demand curves, Brennan (1995) used the same questioning method as used by Brennan, Esslemont and Hini (1995) at three price points (\$10, \$18 and \$25) for one group, while the order of the price points were reversed for the second group (\$25, \$18 and \$10). This questioning was asked in two consecutive years. In both years it was found that when the high price was asked about first, the probability of purchasing was greater at each price point than when the low price was asked first. It was suggested that the cause may have been due to a question order effect, as estimates for each

price point were obtained from each respondent, therefore, the estimates were not independent. After making a judgement about the chances of using at one price point, the respondents may feel that they must increase the probability when the price decreased, and decrease the probability as the price increased as this would be a rational response.

### **3.8 Limitation of Tests of The Juster Scale and Verbal Purchase Probability Scale**

Despite the many studies investigating the accuracy of the Juster scale to predict a range of products and services, all of the studies share two limitations. Firstly, the predictive ability for specific product classes and brands have been tested. However, the knowledge gained from these tests are rather limited in their generalisability to similar products. If the tests were directed to different purchase situations, for example, the accuracy of predictions made by users compared to non-users, then more generalisable results may result. Secondly, the studies testing the accuracy of predictions have used two interviews, one to collect purchase probability data, and the other to collect actual purchase behaviour data, to test the accuracy of predictions against. However, recall is known to be biased, and therefore, the validity of using recall data as a surrogate for actual behaviour must be questioned. It would be possible that using recall to verify predictions may have lead researchers to draw false conclusions concerning not only the accuracy of the predictions, but even the validity of the conclusions drawn from the experimental studies undertaken. For example, if the recall that is used as a surrogate for actual behaviour is closer to the predicted value than the actual behaviour is, then the calculated errors in predictions would appear to be smaller than if actual behavioural data had been collected and used to calculate the error in prediction. Furthermore, if in an experimental design different treatments were being tested, and the different treatments affected the subsequent recall differently, this could lead to false conclusions concerning the relative accuracy of the various treatment. To illustrate this point, consider the following example; two treatments both over-predicted actual purchasing by ten percent. However, the first treatment sent stimuli to the respondents that made them over-report their actual purchasing by ten percent, while the second treatment sent stimuli to the respondent which made them give completely accurate purchasing recall. By using the respondent's recall to

calculate the error in the prediction, even though there is no difference in the predictive ability of the two treatments tested, it would be concluded that using the first treatment method would give more accurate predictions than the second treatment simply because the over-prediction is being compared to the over-reported recall.

Given the potential problems with using recall data to calculate the error in predictions, it would be appropriate to investigate the biases affecting recall.

### **3.9 The Accuracy of Recall Data**

There are three common biases associated with recall, averaging and rounding, omissions and telescoping.

#### **3.10 Averaging and Rounding**

One of the common errors with recall data is that of averaging. This is when a respondent gives answers that are more like their usual behaviour. In a study by Hornik, Cherian and Zakay (1994), on the amount of time spent on various activities (such as television viewing and cigarette smoking) not only was it found that a significant proportion of respondents rounded their responses, but rounded responses resulted in significantly different results between the rounded and non-rounded responses.

It is claimed that averaging and rounding is used by respondents in-order to reduce the amount of mental processing required (Cook 1987). Huttenlocher, Hedges and Bradburn (1990) report the presence of rounding to prototypic values in a survey of recall of event dates. It was found that dates were rounded either by calendar intervals (7, 14, 21 days, and so on representing weeks, or 30, 60 days, and so on representing months), or by decimal rounding (such as 5, 10, 20 and so on).

Cook (1987) reports a survey in which it was noticed that when the respondents were asked about their evening meals, it was found that responses would be structured like "It was Monday, it must have been meat-loaf". This indicates that instead of actually recalling what was eaten that night, the respondent simply recalled their usual meal for a Monday night, and reported that. In order to overcome the problem of averaging, respondents were asked a series of questions on elements of the evening, such as preparation, nature of the occasion, and then the main dish and so on. It was also claimed (although there was no attempt to validate it) that the extra questioning would indicate the level of accuracy required of the respondent. This claim, that increased questioning serves as a cue to the respondent to provide more precise answers, is one which is also used by proponents of single interview bounded recall techniques. This technique is used to reduce recall error and is discussed in a later section.

### 3.11 Omissions

Omission is not only when an entire event is forgotten, but has also been defined as incomplete recall (Cook 1987), for example, when some part of an event is recalled, but other parts of the event are not.

It has been claimed that omissions are due to the exponential decay of short and intermediate memory (Sudman and Bradburn 1973). If memory did consistently decay exponentially, then it should be possible to determine a memory decay function, and estimates of the effect of decay should then be able to be made. However, exponential decay has only been shown in laboratory experiments (Baddeley 1979). Research on the decay rate of memories has shown that it is not necessarily exponentially shaped at all in real life situations (Warrington and Sanders, 1971). Considering the factors that influence recall, such as motivation (Hawkins and Coney, 1981), level of reinforcement (Cassells 1991), elapsed time since the last time the memory was used (Schwarz, Strack and Mai, 1991) to name just a few, it is not surprising that memory decay rates do not conform to a consistent pattern.

For short term memories, it has been shown that the age of the respondent appears to be the only factor which affects omissions. Research into the accuracy of the recall of famous faces (Warrington and Sanders 1971) has shown that the only factor that affected omissions of short term memory was the age of the respondent. Work by Sudman and Bradburn (1973) on factors affecting recall of a variety of events found that respondents over the age of 55 had increasingly inaccurate recall as the recall period increased from up to 13 weeks to 14 or more weeks, while the increasing recall period did not affect the accuracy of the recall of the under 55 year old respondents.

The length of the recall period, the degree of threat posed by the question, using face-to-face interviews opposed to self completion, open ended questions and questions at the start of the questionnaire can reduce the level of omission for long time periods (Sudman and Bradburn 1974). Omissions can be reduced with the use of aided recall, such as showcards (Sudman and Bradburn 1974).

### **3.12 Telescoping**

Of the three main recall biases, telescoping is considered to be the most important (Sudman and Bradburn 1974). Telescoping is generally considered to be when an event is recalled as occurring within a time period when in actual fact it occurred either before, or after the time period (although telescoping was originally found in the recall of crop yields (Mahalanobis, 1946)). The telescoping of time can be either forwards or backwards. In forward telescoping, events that occurred prior to start of the recall period are recalled as happening during the recall period, that is, the events are recalled as happening at a later point in time than it actually did, so that it can be reported that the event occurred within the required recall period. Backward telescoping is where an event that happened after a recall period is recalled as happening earlier than it actual did, so that it falls within the recall period. For example, backwards telescoping would occur if a respondent was asked to recall how often they had performed some behaviour in the previous calendar month, but the respondent included in their recall behaviours that were actually performed in the current calendar month. However, when a recall question asks how often a behaviour has been performed since some

point in time, the respondent cannot backwards telescope, as the end of the recall period is defined by the recall interview. To backwards telescope in such circumstances would involve the respondent "recalling" future events as occurring during the recall period.

Neter and Waksberg (1964) were the first to specifically study telescoping in a study of expenditure on household repairs and alterations. In their study, Neter and Waksberg (1964) used a procedure called bounded recall where respondents were interviewed to collect purchase data on household repairs and alterations at the beginning of the survey period. After the time period of interest, the respondents were interviewed about household repairs and alterations again. If the respondent recalled items that were reported in the first interview, the interviewer queried the response. Respondents could then decide if the purchase was actually made in the target time period, or was telescoped from before the target time period.

Using this technique, it was found that there was significant telescoping, and for larger jobs, there was generally a larger telescoping effect, although the effect was not significant due to the large error for the larger dollar amounts. It was also found that there was telescoping within the time period, that is, there were significantly more expenditure in the first month of a six month recall period, than there was in a one month recall period. It was concluded that the results showed a systematic tendency to misplace the date of expenditure in memory. However, Huttenlocher, Hedges and Prohaska (1988) claim that the results found by Neter and Waksberg (1964) were actually due to preventing forward telescoping, but not backwards telescoping (recalling the date of events earlier than they actually happened).

Although this technique would work well for infrequent purchases, it may not work so well if the respondent frequently purchased the product. If a purchase is infrequent, that is, no more than one purchase is made during the interview period, and the respondent had reported they had made a similar purchase prior to the last interview it is then unlikely that they would have made a second purchase. Telling the respondent that they had reported purchasing the item would send cues to the respondent to consider the behaviour more carefully as it is unlikely that they would have bought two items in the time period. However, if the item is purchased frequently, then it would be expected that several purchases would be made since the previous interview. Reminding the respondent that they reported that they had already

reported buying the frequently bought product would not send the respondent the cues to consider their response more carefully, as it is probable that they would have made additional purchases since the previous interview. Thus, there are three widely acknowledged recall biases, omission, averaging and rounding, and telescoping. The biases may affect the recall of respondents which have been used to test the accuracy of Juster Scale and Verbal Purchase Probability scale predictions, casting some doubt over the validity of the conclusions drawn from such studies. However, inaccurate recall may also affect the predictions themselves. This proposition is explored in the following section, 3.13.

### **3.13 Model of Subjective Probability**

The effect that inaccurate recall may have on the errors calculated have been discussed. However, it may be that if the respondent has inaccurate recall of their purchasing behaviour, this will have an adverse affect on the accuracy of their predictions. If respondents use their recall of previous purchasing to base their predictions about future purchasing on, then it would follow that if their recall was inaccurate, their predictions would also be inaccurate (assuming that their future purchasing is similar to their previous purchasing).

Considering there has been nearly 30 years of research into purchase probabilities and intentions, it would be reasonable to assume that there would be a well established model of subjective probability formulation upon which a rational for suggesting the accuracy of the respondent's recall would affect their predictions could be based on. However, despite the large amount of research that has gone into laboratory experiments on subjective probability formulation, there has been only one model suggested for non-laboratory probability tasks, and even this model has not been tested for its validity (Beach and Braun, 1994). The model proposed by Beach, Barnes and Christensen-Szalanski (1986) can be seen in Figure 4.

Figure 4. Beach, Barnes and Christensen-Szalanski's contingency model of subjective probability judgement.



The model begins with the strategies available to the respondent for making probability judgements. There are two main strategies available, knowledge based and statistically based. In knowledge based probabilities, the respondent uses their personal knowledge about the event to make a judgement. The other main strategy, statistically based, is where the respondent uses their knowledge of the frequency of the event to make a judgement of the probability of the event occurring. To contrast the use of the methods, consider the following questions.

- 1) How many cars are there in this country?
- 2) What are the chances that your best friend owns a car?

It is conceivable that either strategy could be used for both questions depending on the knowledge of the respondent. However, it would be probable that to answer the first question, respondents may need to consider the number of households in the country, and the number of cars per household to estimate the number of cars. For the second question, it would be expected that respondents would use their personal knowledge of their best friend to give the associated probability.

It appears that the strategy that would produce the most accurate predictions depends upon the knowledge of the respondent. If the respondent has knowledge about the event, then use

of a knowledge based strategy would produce more accurate estimates. If the respondent does not have any specific knowledge about the event, then use of a statistical method is likely to produce more accurate estimates. Scott Armstrong, Denniston Jr. and Gordon (1975) provide some evidence to support part of this claim that a statistical strategy would result in better judgements if the respondent does not have specific knowledge when they used a decomposition method to improve the accuracy of estimates. Respondents were asked a series of questions which were constituent parts to an overall question, for example, to estimate the number of families living in the USA in 1970, respondents were asked the population of the USA in 1970, and then the average size of a family. These two estimates were then used by the researchers to calculate the number of households. It was found that this decomposition method resulted in more accurate estimates than simply asking the overall question of how many families were there living in the U.S.A. in 1970. Parfitt (1967) compared the results of asking questions such as "In an average week, how much instant coffee do you buy?", with the combination of "About how often do you buy instant coffee?" and "About how much instant coffee do you buy at a time?". Although the results were not conclusive, there was a marginal improvement in accuracy using the second, decomposition question method. In effect, a decomposition method actually appears to force the respondent to make a well thought out "average" response.

### **3.14 Effect of Inaccurate Recall on Purchase Probabilities**

Both the statistical and knowledge based strategies require the respondents to recall some information, whether it be the number of households, or if a friend owns a car. If the respondent inaccurately recalls the information, it would follow that their final answer would be inaccurate. By aiding the respondent's recall of the necessary information to ensure the information is as accurate as possible, the resulting answer should be more accurate too. However, if the prediction of future purchases are to be improved through the use of accurate recall of previous purchasing, it must be established that future purchases are similar to previous purchases. If future purchases were consistently different from previous purchases, then accurately recalling previous purchases would not aid in the prediction of future purchases.

In a study by Bird and Ehrenberg (1966), a relationship between intentions to buy and claimed brand usage was found. In more than 20 different product fields, intentions to buy could be modelled from current usage with a mean deviation of around 3 percentage points. Although the formula models intentions from current usage, it does not establish the relationship between intentions to buy and actual purchasing. Thus, the model does not predict purchasing from usage, but it does demonstrate a relationship between intentions and current usage.

In a study of frozen orange juice, Kuehn (1962) has shown that the probability that a particular brand is bought on the next purchase is highly dependant on the previous brand bought. The more often a particular brand is purchased, the more the chances are that, that particular brand will be purchased again. Unfortunately, Kuehn's (1962) work only predicts the next purchase, not the number of items that would be purchased over a time period. This research does demonstrate the relationship between previous purchasing and future purchasing.

Given the link between past and future usage establish by Bird and Ehrenberg (1966) and Kuehn (1962), there is strong reason to use the respondent's recall of previous purchases in their prediction, not just because future purchases are related to past purchases, but Schuman and Presser (1981) have shown that at times, respondents will use the answer to a question as a guide to a later question. Therefore, the recall of previous usage may be used as a guide to questions concerning the probabilities of future usage.

Although Beach, Barnes and Christensen-Szalanski's (1986) model has strategy selection as the second stage, followed by motivation, it seems that the motivation to give accurate answers would influence the strategy selection as well, for example, if a respondent has no motivation to give accurate answers, and wishes to do the minimal amount of work to provide an estimate, then guessing would be an appropriate strategy. In fact, it could be argued that the whole process, except the strategy repertory could be influenced by the motivation to give accurate answers.

The motivation to give accurate answers is the only part of this model which has been tested so far (Beach and Braun, 1994). In a study on auditor performance by Johnson and Kaplan (1991), it was found that by increasing the motivation to give accurate answers by telling auditors that their work was going to be reviewed, the resulting work was more consistent than the control group who were given no indications about a review of their work. Other researchers have also found that motivation to provide answers, as opposed to motivation to provide correct answers, can affect responses (for example, Hawkins and Coney, 1981; Clark and Tiff, 1966 and Fisher 1993).

The final two stages involve the respondent going through the mental process of implementing the strategy they will use and then offering their judgement.

Beach, Barnes and Christensen-Szalanski's model of subjective probability formulation provides some evidence to suggest that accurate recall should lead to accurate predictions if the future behaviour is related to past behaviour. The work of Bird and Ehrenberg (1966) and Kuehn (1962) has shown that future purchasing behaviour is related to previous purchasing. Therefore, if the error in recall of previous purchasing could be reduced prior to asking the prediction question, the subsequent predictions should be more accurate the smaller the error in recall is.

### **3.15 Reducing Recall Errors**

As more accurate recall of previous purchasing should lead to more accurate predictions, ways of reducing recall error need to be investigated. Two techniques have been shown to be effective at reducing telescoping (the main source of recall error) in recall.

The first technique is the use of landmark events, that is, an event which clearly marks the beginning of the recall period.

In a survey of being a victim of a robbery, Loftus and Marburger (1983) asked half of their sample questions in the form of, "In the last 6 months, did anyone try to rob you?", while the other half of the sample was asked questions in the form of "Since the first major eruption of

Mt. St. Helens, has anyone tried to rob you?". The results showed that a significantly higher level of crime was reported when the question asked "In the last 6 months...", compared to asking "Since the first major eruption of Mt. St. Helens...". The higher reporting of crime from the "6 month" question was expected due to the effect of telescoping being lessened for the "Mt. St. Helens" question, which demonstrates that a landmark event can be effective at reducing over-reporting in recall questions. However, it should be noted that the actual victimisation rate was not verified in any way. In further experiments, Loftus and Marburger (1983) also found that personal events such as a birthday, and public events such as a public holiday can both be effective at reducing the amount of over-reporting in recall.

Research by Huttenlocher, Hedges and Bradburn (1990) demonstrated that there are at least two ways in which the timing of an event is recalled from memory, these are by calendar recall or sequence recall. Under calendar recall, the date of events are recalled by a calendar date, for example Christmas Day is the 25th of December. The other method, sequence, is where events are remembered in the order in which they occurred. As a new event occurs, it is placed at the end of the sequence. To recall an event from sequence memory requires the position of the event in the sequence being identified, it can then be reported as happening before, or after some other event. It was found by Huttenlocher, Hedges and Bradburn (1990) that, for recalling the date of an interview at least, respondents spontaneously used a sequence method for recalling the date. Although it has not been specifically tested, landmarks may be effective because respondents use a sequence method of recalling when events occur. By placing a clear boundary at the beginning of the recall period, the respondent only needs to recall events in the sequence that have occurred since the beginning of the recall period, rather than having to recall the specific date that an event occurred on.

The second method of reducing the amount of telescoping in recall is called bounded recall. Bounded recall was developed by Neter and Waksberg (1964), where they used two interviews to identify purchases that were telescoped. However, there are some limitations to using two interviews. One that had already been discussed is the possible inappropriateness of the two interview method for frequently purchased items. The others include the extra expense of having to interview twice, and the data collected in the first interview is not useful in terms of being accurate recall data (other than to identify purchases already made).

In an adaptation of the work by Neter and Waksberg (1964), Sudman, Finn and Lannom (1984) developed a method of asking respondents to recall their behaviour relating to aspects of their health for the previous month, and then for the current month, in effect, a single interview bounded recall. As the respondents were asked about the current month, which means they are recalling for a period of less than one complete month, the responses were adjusted to make them comparable to the recall of the previous month. It was found that for every event the respondents were asked to recall, the bounded recall (the recall for the current month) was less than for the unbounded recall (the recall for the previous month), and that the bounded recall was significantly less for three of the four events. In another study reported by Sudman, Finn and Lannom (1984), the time period was reversed, asking about the current week, followed by the previous week, and as a control, asking about the previous week, followed by the current week. It was again found that the lower levels of purchasing was reported for the second (bounded) time period asked about, with five of the six being significantly lower than the level reported from the unbounded recall.

Loftus, Klinger, Smith and Fiedler (1990) also experimented with the single interview bounded recall technique. Where Sudman et al asked about the previous month and then the current month, Loftus, Klinger, Smith and Fiedler (1990) asked about the last six months, followed by the last two months. An important distinction between the method used by Sudman et al and Loftus et al is that the last two months is a subset of the last six months, whereas Sudman et al used the previous month and the current month. In Loftus et al's methodology, it would be more probable that after giving a response to the six month question, the respondent may simply divide the answer to the six month question by three to give an answer to the two month question (in effect, averaging their response). In Sudman's method, as the time period of interest was the current month, and therefore not a complete month, and since the time period of interest was not a subset of the longer time frame, it would be expected that it would be less likely that the respondent would simply give an answer to the second time that was simply a function of the response to the first question, as it would require much more mental processing than the Loftus et al method would require.

The subject of Loftus et al's study was whether a variety of medical procedures had been

performed. The respondents were asked questions in the form of, "During the past  $N$  months, since (date), have you had any of the following procedures done under your GHC coverage?". Asking about a long time period, followed by the time period of interest; and a short time period followed by the period of interest were tested and compared to asking just about the period of interest. The recall was compared to the hospital records to assess the level of accuracy in the recall (although it was acknowledged that there may be some error in the hospital records). It was found that asking about a long period, followed by the period of interest was effective at reducing the level of error in the respondent's recall. Asking about a short period followed by the period of interest was less consistent with only weak evidence that asking about a short time period followed by the period of interest would reduce the level recall error.

The problem of over-reporting in recall is well documented, as too is the over-estimation in predictions obtained using the Juster scale (for example, see Hamilton-Gibbs 1986). It may be that the bias that leads respondents to over-report in recall questions is similar to the bias that leads respondents to over-estimate when making predictions. It has been suggested earlier that the over-estimation in predictions may be due to respondents basing their predictions on inaccurate, exaggerated recall. However, there may be more than just one factor that leads to over estimations in predictions. There may be a response bias that leads respondents to give answers that are generally too large, explaining why recall tends to be over-stated and why predictions tend to be over-estimated.

There have been several theories as to why telescoping may occur. Bradburn, Rips and Shevell (1987) claim that as memories age, the clarity of the memory serves as a cue to the respondent as to the age of the event. Implicit in this theory then are the degree of learning involved and the importance of the issue to the respondent as these would affect the clarity of memory, and therefore telescoping. However, work by Thompson, Skowronski and John Lee (1988) has shown that the clarity with which an event is recalled does not affect telescoping.

Sudman and Bradburn (1974) suggested that telescoping is a function of the exponential memory decay and the linear increase of error in memory as it ages. It was claimed that a

function of the two variables could then be found which would estimate the telescoping effect present. However, there seems to be no logical reason why the error in the aged memory should be on average in a positive direction, in order to give the frequently observed phenomena of telescoping. Furthermore, as has already been discussed, memory decay is not necessarily predictably exponential, and therefore, a function predicting, or modelling the effect of telescoping could not developed based on the exponential decay of memory.

The most empirically supported reason for the occurrence of telescoping has been put forward by Loftus, Klinger, Smith and Fiedler (1990) who found that by asking respondents about two time periods, the amount of over-reporting for the second time period is reduced. From this result, it was claimed that the reduction in the over-reporting was because asking about the second time period sends cues to the respondent that there is a need for greater accuracy, that is, the second question "conveys to the respondent that the interviewer desires greater precision in responding" (p344). The conveyance of the need for greater precision would be consist with Beach, Barnes and Christensen-Szalanski's model of subjective probability formulation, where the third factor is the motivation to provide accurate answers. This suggests that if the prediction question was bounded, that is, the respondent was first asked to predict for a long time period, followed by a prediction for the period of interest, there would be a reduction in the over-estimation, and therefore, more accurate predictions would result.

One of the main limitations of the tests of the accuracy of the predictions remains that recall data which is known to be subject to several biases, is used to calculate the error in the predictions. What is needed, is a method of collecting actual behavioural data that is not biased in any way.

### **3.16 Alternative Methods of Collecting Behavioural Data**

The recall method is often used to collect behavioural data, despite its well known limitations. However, there are several alternative methods for collecting purchase data, each of which have their advantages and disadvantages.

Possibly the "Rolls Royce" experimental and purchase data collection method is single source data. With single source data, a person is given an identification card which they use when they make a purchase. The identification card instructs the computer to send the list of items purchased to the research company. The household is also equipped so that promotional campaigns can be manipulated for experimental purposes.

The obvious advantage is that purchases are recorded with little effort by the respondent. However, the purchases must be made at a store equipped with a scanner, otherwise the purchase is not recorded. Furthermore, single source panels are restricted to isolated communities (for experimental purposes), and require considerable resources to set up.

Sudman (1964) suggested secretly following members of the household and recording the purchases made. The limitations of such an approach are obvious. The cost and logistical difficulties make this method quite inhibitive.

Pantry checks have been used, but these only give a snap shot in time. Repeated pantry checks would face the problem of not knowing if a product seen twice is in fact the same product, or a replaced product. A simple mark on the package could reduce this problem, but would not be suitable for products bought from "bulk bins", such as flour, which may be added to existing stock. Furthermore, repeated pantry checks would not record products bought and consumed between checks.

Other methods such as using seller or manufacturer records have been suggested, however, these only give results at an aggregate level.

The diary method has been called the "single best alternative to recall procedures" (Sudman and Ferber, 1974, p128). In a study by Parfitt (1967), it was found that for 12 consumer products, recall consistently reported higher levels of purchasing than that reported in the diary, a result that would be expected due to telescoping. The level of over-reporting ranged from 16% for tea to 296% for floor and furniture polish.

Although the level of reported purchasing was exaggerated, the relative frequency of purchasing across products was reasonably consistent between data collection methods. A similar result was found by Wind and Lerner (1979) where the ranking of purchasing for brands of margarine was consistent irrespective of whether a diary or recall method was used. It was also found by both Parfitt (1967) and Wind and Lerner (1979) that the errors increased as the frequency of purchase reduced. In other words, for infrequently purchased items, recall was less accurate

There has been much research into the factors affecting the level of reporting using a diary method. A summary of some research investigating factors that affect reporting in diaries can be seen in Table 1.

Table 1. Summary of research investigating factors affecting reporting in diaries.

Author	Finding
Kemsley and Nicholson (1960)	Personally distributed diaries resulted in higher reporting than distribution by post. A blank page did not have a significant affect compared to a structured page. Payments increased response rate. Reporting in the 1st week was significantly higher than the 2nd, 3rd or 4th weeks.
Sudman (1964)	Front pages of the diary had higher reporting than the middle or rear pages. Non-food, non-grocery items were less accurately reported than food or non-food grocery items. Entries using a prominent entry were more accurately reported then entries using a lighter type.
Sudman and Ferber (1971)	Individuals did not keep individual diaries, but could report their purchases to a record keeper. More structured diaries were easier to use. Less than half of those asked kept diaries for four weeks. Gifts increased participation. Large (durable) items may be better picked up using bounded recall.
Sudman and Ferber (1974)	Compensation improves co-operation and gave higher reports of expenditure. Auspices of the organisation did not make much difference in the total reported. A daily telephone method of diary keeping appears to be not as complete or accurate.

As can be seen in the table above, there are many factors that have been shown to influence the level of reporting in diaries. This clearly demonstrates that although diaries may be the "best" alternative to recall, the diary method is still subject to many biases itself.

Ideally what is needed to test recall, diaries and any method designed to record behaviours, is a behaviour that can only be performed where a record must be kept. However, even when records must be kept, and kept accurately, some errors in records can occur (Martin 1987). Ideally, what is needed is a behaviour that can not be performed unless an accurate record is kept. For example, the use of a credit card requires all transactions to be electronically recorded, therefore, the behaviour must be recorded, and it should be recorded without any error, so long as there is no error in the hardware or software recording the transaction. Unfortunately, there are only a few products or services that require accurate records to be kept.

## **4. HYPOTHESES**

### **4.1 Introduction**

The Literature Review presented the relevant literature on purchase probability measurement, retrospective question bias and subjective probability. In doing so, the Literature Review identified several issues that require further investigation. This section formally states eight hypotheses that resulted from the examination of the literature.

### **4.2 Hypotheses 1, 2 and 3**

Research into consumer behaviour and questionnaire design has produced evidence which, when applied to subjective probability formulation provides a framework for suggesting a method which will affect a respondent's predictions, leading the respondent to make predictions that more closely reflect their actual future behaviour.

Starting with the premises that for frequently purchased items the chances that a consumer will choose a particular brand is related to the previous brand bought and that respondents have been shown to use the answer to a previous question as a guide to subsequent questions in a questionnaire, it would follow that requiring a respondent to accurately answer a recall question prior to a prediction question would have two affects. Firstly, accurate recall would explicitly force the respondent to think about their previous behaviour, and secondly, thinking about their previous behaviour would act as a guide with respect to their future behaviour. The results of these two affects should improve the accuracy of predictions.

As bounded recall (the technique of asking a respondent to recall behaviour for two time periods, one longer than of interest, and another for the time period of interest) increases the accuracy of recall, asking a bounded recall question would be more effective at reducing predictions errors than a single recall question. Hence:

- H1 Asking a single recall question prior to the prediction questions will result in smaller errors in predictions than asking the prediction questions without any previous recall questions.
- H2 Asking a bounded recall question prior to the prediction questions will result in smaller errors in predictions than asking the prediction questions without any previous recall questions.
- H3 Asking a bounded recall question prior to the prediction questions will result in smaller errors in predictions than asking a single recall question prior to the prediction questions.

#### 4.3 Hypothesis 4

When making predictions, people go through a series of stages and influences that affect the prediction they make. These stages and influences have been modelled by Beach et al's model of subjective probability. One of the influences in the model is the motivation to provide accurate answers. It would follow that if a respondent's motivation to provide accurate answers could be increased, then their resulting predictions would be more accurate.

One method for increasing respondent motivation is bounded questioning. It has been claimed that a bounding technique is effective at reducing telescoping in recall questions by increasing the respondent's motivation to provide accurate answers. Therefore, a bounding technique of asking two prediction questions, one for a time period longer than of interest, and the other for the time period of interest may be effective at increasing the accuracy of predictions through increased motivation to provided accurate answers. Hence:

- H4 Using a bounded prediction question technique will result in smaller errors in predictions than predictions from using only a single prediction question.

#### 4.4 Hypothesis 5

Various methods have been used to calculate the purchase level. These range from the logically complete, but practically cumbersome multiple question method, to the reduced question method used by Brennan, Esslemont and Hini (1995), without adversely affecting prediction accuracy. Having established that purchase level calculation do not need to be logically complete to be accurate, it may be that further reductions in the questioning required is possible, without a significant decrease in prediction accuracy. Hence:

H5 Reduced prediction questioning procedures will not affect the error in purchase level predictions.

#### 4.5 Hypothesis 6

Recall data has often been used as a surrogate for actual behaviour when testing the accuracy of predictions from the Juster Scale and the Verbal Purchase Probability Scale. Recall is known to be affected by several biases, predominately telescoping which causes over-reporting. However, the use of a landmark event (an event which clearly defines a time boundary) has been shown to be effective at reducing over-reporting in recall questions. When recall data has been used to test the accuracy of predictions, the interview in which the predictions were obtained may have acted as a landmark event. As there is a landmark event aiding the accuracy of the respondent's recall, the affect of the recall biases would be reduced, and the respondent's recall would be more similar to that of unbiased, accurate behavioural data. Hence:

H6 The error in predictions calculated using recall as a surrogate for actual behaviour will not be different from the error in predictions calculated using actual behavioural data.

#### 4.6 Hypothesis 7

In the model of subjective probability formulation, the respondent is required to recall previous events in order to make a prediction about the future. Assuming that future behaviour is similar to past behaviour, and as discussed earlier, brands purchased usually are, it follows then, that as the accuracy of recall increases, so too would the accuracy of predictions. Hence:

- H7 Respondents with accurate recall at the time of making a prediction will have less error in their predictions than those with inaccurate recall at the time of making the prediction.

#### 4.7 Hypothesis 8

It has been found in a previous study (U 1991) that when predictions are made for products which have low levels of usage, the errors in predictions are high. Parfitt (1967) found that recall was less accurate for infrequently purchased items. Hypothesis 7 states that respondents with accurate recall at the time of making a prediction will have less error in their predictions than those with inaccurate recall at the time of making predictions. Therefore, if recall is less accurate for infrequently purchased products, by assuming that infrequently purchased items indicates infrequent usage, it would follow that predictions would be less accurate for infrequently used products. Hence:

- H8 Errors in predictions for low users will be greater than the errors in predictions for high users.

## 5. METHOD

### 5.1 Introduction

In the previous section, based on the theoretical framework developed in the Literature Review section, eight hypotheses were developed. This Method section details the sample, procedures, treatments (and instruments), the limitations of the study and operationalises the variables test the hypotheses.

The section begins by detailing the sample used in the survey, 1360 members of a local video store who were selected based on a variety of criteria to ensure even representation in each of the treatment groups. The response rate achieved is then presented.

The procedures for the telephone interviewing are then presented, followed by a detailed explanation of the treatments involved in the study, that is, one Single Recall Group, one Bounded Recall Group, one Bounded Prediction Group and three Control Groups. Specifics of the questionnaires associated with each of the treatment groups are then given, including both the questionnaires where the purchase probability data is collected, and the recall questionnaires where respondent recall data is collected.

The limitations of the study are then discussed. The limitations include being unable to control which video stores the respondent may hire from, and other people who may use their video card to hire videos. Having acknowledged these limitation, it is explained how these limitation have been minimised in the design of the questionnaires.

This section concludes with the definition of the variables. This includes an explanation on the calculation to estimate purchase rate and purchase levels. The purchase rate calculations are much more complex than that which have used in previous studies investigating purchase rates as the purchase rates have previously been calculated for the respondent personally making a purchase. However, in this study, the purchase rate must be calculated for any

person hiring a video. Having operationalised the predictions, there is a discussion on the operationalisation of the actual number of videos hired. This assumes particular importance as it was necessary to be able to match computer records with respondent recall. However, this was complicated because the computer records could only show the total number of videos hired on any one day, but a respondent could be interviewed at any time during the day. Thus, it is possible that at the time a respondent is interviewed, they may not have hired any videos, but proceeded to do so, later in the day. The computer records would show that videos were hired on that day, but the respondent data would show that they indicated they had not. How these two pieces of information were matched is discussed.

## 5.2 Sample

The sample was made up of 1360 members of a local video store who had hired at least one video in the previous two months. Pre-recorded video tape hire was selected as the behaviour being predicted as computer records must be kept each time a video is hired. Thus, through the co-operation of the video store, a record of the actual behaviour can be kept.

Only those who had hired in the previous two months were selected to ensure that the people in the sample were still current users of their video cards and that they are likely to use their video cards again during the interview period. Members of the sample were systematically assigned to one of six treatment groups (described below). To control for any effect resulting from the level of usage, the treatments were balanced by the number of videos that had been hired in the previous two months. This data was obtained from the video store's computer records of videos hired.

Some of the treatments required the respondent to recall previous behaviour. Thus people with a birthday falling within the months that the interviewing was conducted were omitted from selection in the sample to avoid any landmark effects their birthday may have. In other words, people with birthdays during the interview period were not eligible for selection to eliminate any possible effects on recall due to their birthday acting as a landmark event. Furthermore, the month of birth of the respondents were balanced across treatments to reduce any landmark effects that may result from a birthday near the time of the interview.

Although the sample was stratified by the number of videos hired in the last four weeks, the sample was post-stratified to ensure that the treatments were matched exactly with respect to this criteria. This post-stratification was achieved by eliminating respondents from each strata within each treatment group until there were equal numbers of respondents in each strata of each treatment group. For example, if one group had 11 respondents who had hired two videos and another group had 16 respondents, five respondents were randomly deleted from the second group so that there were 11 respondents in each treatment group who had hired exactly two videos. This procedure was undertaken instead of weighting so that significance

testing could still be performed. The final sample size for each treatment group can be seen in Table 2.

Table 2. Distribution of number of videos hired

Number of videos hired	Total in sample	Total in each treatment
0	414	69
1	60	10
2	66	11
3	36	6
4	12	2
5	6	1
6	6	1
Total	600	100

### 5.3 Response rate

In the initial interview, of the 1360 in the entire sample, 273 either no longer lived at the address that was recorded on the video store computer, and the respondent's new phone number was not known, or there was no such phone number. The 788 successful interviews completed represents a response rate of 72% (details can be seen in Table 3).

Table 3. Response rate for the Juster interview

Outcome	Treatment <sup>1</sup>						Total
	T1	T2	T3	T4	T5	T6	
Total	227	227	227	226	227	226	1360
No such number	18	25	37	28	23	23	154
Respondent not living there	20	19	15	22	22	21	119
Respondent refusals	31	24	22	33	29	14	153
Household refusal	2	2	3	2	4	5	18
Could not call back	2	1	0	1	0	1	5
Uncontactable	11	17	10	17	23	11	89
Other	5	8	5	2	3	11	34
Successful	138	131	135	121	123	140	788
Response rate (%) <sup>2</sup>	73	72	77	69	68	77	72

Note: 1 The treatments are detailed in the following section.

2 Response rate = successful / (Total - No such number - Respondent not living there) \* 100.

In the recall interviews, there were a total of 788 respondents available to be reinterviewed. Of those 788, 30 had moved without leaving a forwarding address, or had their phone disconnected. A total of 667 follow-up interviews were successfully completed, representing a response rate of 88% (details can be seen in Table 4).

Table 4. Response rates for the follow up interview

Outcome	Treatment <sup>1</sup>						Total
	T1	T2	T3	T4	T5	T6	
Total	138	131	135	121	123	140	788
No such number	2	4	3	5	2	3	14
Respondent not living there	5	4	3	3	3	4	16
Respondent refusals	2	2	1	2	2	0	9
Household refusal	0	0	0	0	0	1	1
Uncontactable <sup>2</sup>	5	16	18	5	7	15	77
Other	0	0	0	3	0	1	4
Successful	124	105	110	103	109	116	667
Response rate (%) <sup>3</sup>	95	82	85	91	92	83	88

Note: 1 The treatments are detailed in the following section.

2 There was a large difference between the number of respondents who were uncontactable in treatments T2, T3 and T5 compared with T1, T4 and T6. T2, T3 and T5 were interviewed four weeks later than T1, T4 and T6. It is believed that the increase in respondents that were uncontactable was because the respondents were students (although the occupation of the respondent is not known). The timing of the recall interview for treatments T2, T3 and T5 fell shortly after the end of the university year, and so many of the students that would have been successfully interviewed in the initial survey may have left the city for the summer vacation by the time of the second interview.

3 Response rate = successful / (Total - No such number - Respondent not living there) \* 100.

## 5.4 Procedure

Each respondent was interviewed twice by telephone. In the first interview the chances of hiring pre-recorded video cassettes over the next four (or eight) weeks were collected (according to the treatment procedure described below). In the second interview, the respondents were asked to recall how many videos they had hired over the interview period (four weeks or eight weeks, depending on the treatment). During the interviewing period and prior to the start of the interviewing, the computer records of the number of videos hired each day were collected for each respondent. The actual data was then used to test the accuracy of the predictions.

As well as for the present study, the data from the interviewing was also being used in another study on retrospective question biases. Some of the treatments for this Verbal Purchase Probability Scale study required recall questions, and some treatments in the recall study required an interview conducted four weeks before the recall question was asked. Generally, the two studies complemented each other well. However, one concession was necessary. To minimise the effect of external events that may affect video hiring, such as changes in the weather, it was deemed necessary that the recall period be the same for all treatments. Therefore, it was necessary to interview in three phases. The three phases started at four week intervals so that the recall period for the retrospective question bias study were covering the same time period. Table 5 shows the type of interview procedure conducted at each phase of the interviewing.

Table 5. Phases of interviewing

Hypothesis tested	Treatment	Phase		
		1	2	3
H1,2,5,6,8	Control A (T1)		J4: 4 week Juster	R4: 4 week Recall
H1,3,5,6,7,8	Single Recall (T2)		R4/J4: 4 week Recall : 4 week Juster	R4: 4 week Recall
H2,3,5,6,7,8	Bounded Recall (T3)		R8/R4/J4: 8 week/4 week recall : 4 week Juster	R4: 4 week Recall
H4,5,6,8	Control B (T4)	J4: 4 week Juster	R4: 4 week Recall	
H4,5,6,8	Control C (T5)	J4: 4 week Juster	R8/R4: 8 week/4 week Recall	
H4,5,6,8	Bounded Juster (T6)	J8/J4: 8 week/4 week Juster		R4/R8: 4 week/8 week Recall

Note: J4: 4 week Juster = Juster predictions for the next 4 weeks  
 J8/J4: 8 week/4 week Juster = Juster predictions for the next 8 weeks then predictions for the next 4 weeks  
 R4: 4 week recall = Recall of hiring over the last 4 weeks  
 R8/R4: 8 week/4 week recall = Recall of hiring over the last 8 weeks then recall of hiring over the last 4 weeks  
 R4/R8: 4 week/8 week recall = Recall of hiring over the last 4 weeks then recall of hiring over the last 8 weeks

## 5.5 Treatments

The respondents were assigned to one of the following treatments:

### Control A (T1)

As the probabilities were collected in Phase 2, this group is the control group for the Single Recall (T2) and Bounded Recall (T3) groups.

The respondents were asked the chances that they personally would hire at least one video, the number of videos they were most likely to hire, and the chances of them personally hiring exactly that number of videos over the next four weeks. They were then asked the chances of other people hiring at least one video, how many videos other people were most likely to hire and the chances of other people hiring exactly that number over the next four weeks. Four weeks later, these respondents were reinterviewed and asked how many videos they personally had hired, and how many videos other people had hired since they were last interviewed.

### Single Recall (T2)

The respondents were asked to recall how many videos they personally had hired, and how many other people had hired over the previous four weeks. The respondents were then asked the chances that they personally would hire at least one video, the number of videos they were most likely to hire and the chances of them personally hiring exactly that number of videos over the next four weeks. They were then asked the chances of other people hiring at least one video, how many videos other people were most likely to hire and the chances of hiring exactly that number over the next four weeks. Four weeks later, these respondents were reinterviewed and asked how many videos they personally had hired and how many videos other people had hired since they were last interviewed.

This treatment was used to specifically test the hypotheses that asking a single recall question prior to the prediction questions will result in smaller errors in prediction than asking the prediction question without any previous recall questions (H1) and that asking a bounded recall question prior to the prediction questions will result in smaller errors in predictions than asking a single recall question prior to the prediction questions (H3).

The data collected from the Single Recall (T2) group was also used to test the hypotheses that reduced prediction questioning procedures will not affect the errors in purchase level predictions (H5), that using recall as a surrogate for actual behaviour will not have an effect on the accuracy of predictions when compared to the accuracy of predictions calculated using actual behavioural data (H6), respondents with accurate recall will have less error in their predictions than those with inaccurate recall (H7), and that errors in predictions for low users will be greater than the errors in predictions for high users (H8).

### **Bounded Recall (T3)**

The respondents were asked to recall how many videos they personally had hired, and how many videos other people had hired over the previous eight weeks. They were then required to recall how many videos they personally had hired, and how many videos other people had hired over the previous four weeks. The respondents were then asked the chances that they personally would hire at least one video, the number of videos they personally were most likely to hire and the chances of them personally hiring exactly that number of videos over the next four weeks. They were then asked the chances of other people hiring at least one video, how many videos other people were most likely to hire and the chances of other people hiring exactly that number of videos over the next four weeks. Four weeks after that interview, these respondents were reinterviewed and asked how many videos they personally had hired and how many videos other people had hired since they were last interviewed.

This treatment was used to specifically test the hypotheses that asking a bounded recall question prior to the predictions will result in smaller errors in predictions than asking the prediction questions without any previous recall questions (H2) and that asking a bounded

recall question prior to the prediction questions will result in smaller errors in prediction than asking a single recall question prior to the prediction questions (H3).

The data collected from the Bounded Recall (T3) group was also used to test the hypotheses that reduced prediction questioning procedures will not affect the errors in purchase level predictions (H5), that using recall as a surrogate for actual behaviour will not have an effect on the accuracy of predictions when compared to the accuracy of predictions calculated using actual behavioural data (H6), respondents with accurate recall will have less error in their predictions than those with inaccurate recall (H7), and that errors in predictions for low users will be greater than the errors in predictions for high users (H8).

#### **Control B (T4)**

This group is the control group for the Bounded Juster group (T6) only.

The respondents were asked the chances that they personally would hire at least one video, the number of videos they personally were most likely to hire and the chances of them personally hiring exactly that number of videos in the next four weeks. They were then asked the chances of other people hiring at least one video, how many videos other people were most likely to hire and the chances of other people hiring exactly that number over the next four weeks. Four weeks later, these respondents were reinterviewed and asked how many videos they personally had hired and how many videos other people had hired since they were last interviewed. Procedurally, this treatment is identical to Control A (T1), the Phase 2 control group. The only difference is that interviewing for Control B (T4) began in Phase 1, four weeks earlier than Control A (T1).

#### **Control C (T5)**

This group is also a control group for the Bounded Juster treatment (T6). There was no requirement for T6, the Bounded Juster treatment to have two control groups, however, as

this treatment (T5) was necessary for the study on retrospective question bias which made use of some of the data collected in the series of interviews, this treatment provided more usable data for this study.

The purpose of this treatment in the retrospective question bias study was to test the affect of having a landmark event at the beginning of a recall period (the Juster interview) combined with bounded recall. In other words, the hypothesis was to look at what the affect of having both a landmark event and using a bounded recall technique would be on the accuracy of recall. This was achieved by collecting Juster probabilities in the first interview, followed by collecting recall data for the previous 8 weeks and then 4 weeks in the second interview.

As the Juster probabilities in this treatment were collected in an identical manner to those in Control B (T4), the Juster probabilities collected in this treatment could be considered a control group for the Bounded Juster treatment (T6).

There was some difference in the way in which the recall data was collected between this treatment and Control Group B (T4), but as the accuracy of the predictions is calculated using the actual data from the video store's computer records, the differences in the method of collecting the recall data generally does not have any affect on the results of this study (except H6 where the affect that using recall data as a surrogate for actual behavioural data has on calculated errors in predictions was being investigated).

### **Bounded Juster (T6)**

The respondents were asked what the chances were that they personally would hire at least one video, the number of videos they were most likely to hire and the chances of them personally hiring exactly that number of videos over the next eight weeks. They were then asked the chances of other people hiring at least one video, how many videos other people were most likely to hire and the chances of hiring exactly that number over the next eight weeks. The respondents were then asked the same questions, but with respect to a four week period. Eight weeks later, the respondents were reinterviewed and asked how many videos

they personally had hired and how many videos other people had hired, firstly in the previous four weeks, and then since they were last interviewed (eight weeks ago).

This treatment was specifically designed to test the hypothesis that using a bounded prediction question technique will result in smaller errors in predictions than predictions from using only a single prediction question (H4).

The data collected from the Bounded Juster (T6) group was also used to test the hypotheses that reduced prediction questioning procedures will not affect the errors in purchase level predictions (H5), that using recall as a surrogate for actual behaviour will not have an affect on the accuracy of predictions when compared to the accuracy of predictions calculated using actual behavioural data (H6), respondents with accurate recall will have less error in their predictions than those with inaccurate recall (H7), and that errors in predictions for low users will be greater than the errors in predictions for high users (H8).

## **5.6 Computer records**

For eight weeks prior to, and during the interviewing, the computer records were kept for the 1360 respondents in the sample. The computer store uses a code for each video it has. This code and the day that the video was hired were recorded for every video hired for every person in the sample by an employee of the video store. As only video codes were used, the privacy of the respondent was protected. This data was used to determine the actual number of videos hired, and against which the accuracy of both recall and predictions could be gauged.

## **5.7 Interviewers**

Each treatment was sorted by the respondent's surname to randomise the order (within each treatment) that the surveys were conducted. Questionnaires were then collated into sets for each of the interviewers, rotating the order of the treatments within each set to ensure that

any interviewer learning and fatigue affects, as well as any day or time of day affects were balanced across treatments. Interviewers were then assigned a set of questionnaires and were instructed that the first attempt at contacting each respondent must be in the order that the questionnaires were assigned.

In total, 13 interviewers were used to collect the data over the three interviewing phases. All of the interviewers had had some previous telephone interviewing experience. Each interviewer was given at least one hour of one-on-one training where the purpose of the questionnaire was explained to them. The questionnaire was read through with the interviewer by the researcher and then the interviewer conducted a practice interview with the researcher.

To ensure the interviewers were conducting the interviews to a high standard, ten percent of the respondents were telephoned and asked if they had been interviewed and if they had any complaints. Furthermore, in Phase 2, a dummy respondent was placed in each interviewers batch of questionnaires to check that the interviewers were asking the questions exactly as required. The dummy respondent had a copy of the questionnaire which they read while the interviewer was asking the question. The interviewers were told that they would be interviewing a dummy respondent who was going to check their performance, but they did not know who the dummy respondent was. The dummy respondent's questionnaire was placed approximately sixth in the interviewer's set of questionnaires so that the performance of the interviewer would not be judged while they were still experiencing any initial anxiety. The data from the dummy respondent was not used in the analysis.

Generally, it was found that the interviewers were performing to high standard. Two interviewers were retrained to ensure that they were asking the questions exactly as required, while one interviewer was removed from interviewing due to the quality of their interviewing being unacceptably low.

## 5.8 Instrument

The following is a description of the questionnaires used in the survey. Copies of the questionnaire and charts used can be seen in the appendix.

### 5.8.1 Juster questionnaires

Following is an explanation of the wording in the questionnaires used to collect the probabilities of hiring videos. The treatments referred to in this section can be seen in Table 5.

For all treatments, the interview started with the respondent being asked for by name (names were supplied by the video store). The respondent was then read the following introduction:

*Good evening. My name is <FULL NAME>, I'm doing a survey about videos and would like to ask you a few questions. This interview will only take about five minutes. May I interview you now?*

To reduce non-response, if a respondent was hesitant about participating, interviewers were instructed to stress the importance of the respondent's opinion and that the interview would only take about five minutes.

Having gained consent, all respondents were asked if there was a video player in their household, if they had ever rented a video tape and what video stores they were a member of. If the respondent claimed that they were not a member of the video store assisting with this study, the interviewer double checked with the respondent to ensure that they were not a member of the video store assisting with the research. If the respondent continued to claim they were not a member of the video store, the interview was terminated.

Those respondents in the Bounded Recall (T3) group were then asked:

*Thinking about the past eight weeks, that is, since <DAY AND DATE> SEE CHART A, how many videos have you personally hired from [VIDEO STORE NAME]?*

When a day and date was required, the interviewers were instructed to read from a chart which contained the appropriate day and date from the day of the interview.

The respondents in the Bounded Recall (T3) group who answered more than zero for the previous question, and all the respondents the Single Recall (T2) group were then asked:

*Now, thinking about the past four weeks, that is since <DAY AND DATE> SEE CHART B, how many videos have you personally hired from [VIDEO STORE NAME]?*

The respondents in the Bounded Recall (T3) and the Single Recall (T2) groups were then asked if anybody else had used their video cards in the previous eight or four weeks respectively. If somebody else had used their card, the respondents in the Bounded Recall (T3) group were asked to recall how many videos other people had hired in the previous eight weeks and then four weeks using the wording shown above, except with the references to "you personally" changed to "other people", while those in the Single Recall (T2) group were asked how many videos other people had hired in the previous four weeks.

All respondents were asked for the chances that they would use a new service if it was provided by the video store. The respondents were told:

*I would like you to answer on a scale of "zero" to "10". If you are certain, or practically certain that you will use this service, then you would choose the answer "10". If you think there is no chance, or almost no chance of using, the best answer would be "zero". If you are uncertain about the prospects, choose another answer as close to "0" or "10" as you think it should be. You can think of the numbers as chances out of "10". For example, 3 would mean 3 chances in 10 that you would use*

*the service, while a 7 would mean 7 chances in 10 that you would use the service, and so on.*

Having obtained the chances that the new service would be used, the chances of hiring videos were elicited. All respondents except those in the Bounded Juster (T6) group were asked:

*Now, taking everything into account, what are the chances of you personally hiring at least one video from [VIDEO STORE NAME], during the next four weeks, that is between now and <DAY AND DATE> SEE CHART C?*

If the chances were greater than zero, respondents were then asked:

*If you did hire any, how many videos are you personally most likely to hire from [VIDEO STORE NAME] in the next four weeks?*

*Using the scale from 0 to 10, what are the chances that you personally would hire exactly <MOST LIKELY NUMBER> videos from [VIDEO STORE NAME] in the next FOUR weeks?*

The respondents were then asked the chances that someone else would use their video card to hire at least one video in the next four weeks. If the chances were greater than zero, the respondent was also asked how many videos they thought other people would hire and the chances of other people hiring exactly that number.

For the Bounded Juster group (T6), the respondents were asked the Juster questions above, firstly for the next eight weeks, and then for the next four weeks for themselves personally, and then for the next eight and four weeks for other people using their video card.

The groups that were interviewed in Phase one (see Table 4), that is the Bounded Juster (T6) and Control Groups B and C (T4 and T5) were asked the following question:

*How many videos have you personally hired so far today from [VIDEO STORE NAME]?*

However, it was found that some respondents were incorrectly hearing the question, hearing “today” as “to date”. The question was then changed for the questionnaires in Phase two and three to:

*Thinking only about today, have you personally hired any videos from [VIDEO STORE NAME]?*

*IF YES ... How many have you personally hired just today?*

The same question was asked again for other people with the references to “you personally” changed to “other people”.

Finally, all respondents were asked for permission to call them back if needed.

### 5.8.2 Follow-up questionnaires

For each of the treatments (see Table 4), recall data was collected. This section outlines the question wording used to collect the recall data.

In the follow-up questionnaires, all of the respondents except those in the Bounded Juster (T6) group were asked for by name and were read a quick introduction.

*Good evening. My name is <FULL NAME>. Four weeks ago, we interviewed you about videos, and you said we could call you back to ask a few more questions. This interview will only take about two minutes. May I interview you now?*

For those respondents in the Bounded Juster (T6) group, the reference to “Four weeks ago” was changed to “Eight weeks ago”.

The interviewers were required to conduct the recall interviews exactly four weeks after the initial interview except for those in the Bounded Juster (T6) group where the interviews were to be completed exactly eight weeks after the initial interview.

If the respondent was interviewed on the correct day, they were asked the first set of questions presented below. If the respondent could not be interviewed on the correct day, they were asked a different set of recall questions (see Late Follow-up Interview, 5.8.3). It was necessary that different questions be used for those interviewed late so that the respondent's recall could be matched to exactly the same time period as the predictions.

Except for respondents in Control Group C (T5) and the Bounded Juster group (T6), the respondents were asked how many videos they had hired in the previous four weeks and how many videos other people had hired in the previous four weeks, as follows:

*Since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE CHART B, how many videos have you personally hired from [VIDEO STORE NAME]?*

*Has anyone else, other than yourself, used your [VIDEO STORE NAME] card to rent a video during the last four weeks?*

*Since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE CHART B, how many videos would you say other people have hired using your [VIDEO STORE NAME] card?*

For Control Group C (T5), the respondent was also asked about the number of videos they had hired in the previous eight weeks followed by the previous four weeks as follows:

*Thinking about the past eight weeks, that is, since <DAY AND DATE>, how many videos have you personally hired from [VIDEO STORE NAME]*

*And since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE CHART B, how many videos have you personally hired from [VIDEO STORE NAME]?*

The respondents were then asked about the number of videos other people had hired in the last eight weeks and then the previous four weeks.

For the Bounded Juster group (T6), the respondents were asked about the number of videos they had hired in the previous four weeks as follows:

*Thinking about the past four weeks, that is, since <DAY AND DATE>, how many videos have you personally hired from [VIDEO STORE NAME]*

*And since we interviewed you eight weeks ago, that is, since <DAY AND DATE> SEE CHART B, how many videos have you personally hired from [VIDEO STORE NAME]?*

The respondents were then asked if anybody else had used their video card in the last eight weeks. If someone had, the respondent was then asked the two questions above with the references to “you personally” changed to “other people”.

All respondents that were interviewed on the correct day were then asked if they personally, or other people had hired any videos that day, and if so how many.

### 5.8.3 Late follow-up interview

It was inevitable that some respondents would not be reinterviewed exactly four weeks after the initial interview (or eight weeks after the interview for the Bounded Juster group (T6)). If a respondent was interviewed late, they were not asked the recall questions in the previous section, but the following recall questions.

*Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos have you personally hired from [VIDEO STORE NAME]?*

*Does this number include any videos you personally have hired from [VIDEO STORE NAME]?*

*IF 1 DAY LATE .....just today*

*IF > 1 DAY LATE .....since last <ACTUAL DAY>?*

If the number the respondent recalled did include some videos hired since the day the interview was supposed to have been conducted on, they were then asked the following question.

*How many videos have you personally hired from [VIDEOS STORE NAME]?*

*IF 1 DAY LATE .....just today*

*IF > 1 DAY LATE .....since last <ACTUAL DAY>?*

The respondents were then asked if anybody else had used their video card in the last four weeks. If someone had, the respondent was then asked the three questions above with the reference to “you personally” changed to “other people” in the first and third questions and “anyone else” in the second question.

For those respondents in Control Group C (T5), the respondents interviewed late were first asked.

*Thinking about the past eight weeks, that is, since <DAY AND DATE> SEE CHART A to <ACTUAL DAY AND DATE>, how many videos have you personally hired from [VIDEO STORE NAME]?*

If the respondent reported that they had hired some videos in the previous eight weeks, they were then asked the same questions as the other respondents (outlined above)

The Bounded Juster Group (T6), were first asked about the number of videos hired in the four weeks since the date of the initial interview, and then the number of videos hired for the eight weeks from the time they were interviewed.

*Thinking about the four weeks, that is, since DAY AND DATE SEE CHART A to <ACTUAL DAY AND DATE>, how many videos have you personally hired from [VIDEO STORE NAME]?*

*Thinking about the eight weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos have you personally hired from [VIDEO STORE NAME]?*

*Does this number include any videos you personally have hired from [VIDEO STORE NAME]?*

*IF 1 DAY LATE .....just today*

*IF > 1 DAY LATE .....since last <ACTUAL DAY>?*

If the number the respondent recalled did include some videos hired since the day the interview was supposed to have been conducted on, they were then asked the following question.

*How many videos have you personally hired from [VIDEOS STORE NAME]?*

*IF 1 DAY LATE .....just today*

*IF > 1 DAY LATE .....since last <ACTUAL DAY>?*

The respondents were then asked if anyone else had used their video card in the last eight weeks. If someone had, they were asked the questions above with the references to "you personally" changed to "other people" for the first, second and fourth questions, and to

“anyone else” for the third question.

## **5.9 Limitations**

In previous studies (for example Seymour, Brennan and Esslemont 1994) respondents were only required to predict for themselves personally, and they could make the purchase from any store. For this study, it was critical that the predictions were made for the respondent themselves, as well as other people using the respondent's video card, from the specific video store, so that the predictions could be compared with the computer records. These conditions lead to two limitations of this study.

The first limitation of this study was that although this study was designed to accurately measure the number of videos hired from the video store assisting with study, it was not possible to prevent the respondents from hiring videos from other video stores where the video usage was not measured. As this study was dependant upon using the video stores computer records to provide the actual data, there is the chance the respondents may consider videos hired from stores other than that assisting with the study.

The second limitation was that the video store's computer records the videos hired using the video card, but not who hired the video. In other words, the computer only recorded that a video was rented, there was no information specifically about who the person who actually hired the video was. This means that other people could use the respondent's video card to hire a video, either with or without the respondent's knowledge, which could affect both the respondent's predictions and recall.

To overcome these potentially serious problems, the respondents were asked about the chances that they personally would hire using their video card, and the chances that anybody else would be using their video card to hire a video at the specific video store. Although the limitations were addressed in the design of the questionnaire, there is the potential for the respondents to ignore the parts of the question which limited their possible responses (for example, videos rented only from the specific video store, or videos rented by themselves personally). If a respondent did broaden the question to include hiring videos from other

video stores, then it would be expected that the predictions would over-estimate the number that will be hired, although if they neglected to fully consider the hiring behaviour of other users, then their predictions would under-estimate the number of videos that would be hired.

## 5.10 Variable Operationalisation

### 5.10.1 Error calculations

Predictions can be made at two levels. Purchase rate predictions estimate the proportion of the sample that will hire at least one video. Purchase level predictions estimate the number of videos that are expected to be hired.

Purchase rate errors are calculated as the chances that the respondent will hire at least one video during the prediction period (a value ranging from zero if there is no chance, to one if they were practically certain they would hire at least one video) less whether the respondent did, or did not actually hire at least one video during the prediction period (coded a zero if they did not, and one if they did). Hence:

$$\text{Purchase rate error} = E(a) - a$$

where  $E(a)$  = probability that at least one video will be hired

and  $a$  = whether or not at least one video was hired (0 = no, 1 = yes)

The purchase level error is calculated as the expected number of videos that would be hired during the prediction period less the number of videos actually hired during the prediction period.

$$\text{Purchase level error} = E(n) - n$$

where  $E(n)$  = the number of videos that are expected to be hired

$n$  = the number of videos actually hired

### 5.10.2 Expected value

Both the purchase rate and purchase level error calculation use an expected value, either the expected chance of hiring at least one video, or the number expected to be hired.

### 5.10.3 Expected value calculation for purchase rate

For purchase rate predictions, that is, the proportion of respondents that will hire at least one video, the expected value was a combination of the chances that the respondent would hire at least one video, and the chances that other people would hire at least one video. In previous studies, the predictions have only been made for the respondents themselves, and therefore, the expected value is simply the chances that the respondent would hire at least one video. However for this study, the predictions had to be made for both the respondents themselves hiring at least one video, and other people hiring at least one video so that the predictions could be compared to the computer records. To determine the chances that at least one video would be hired by either the respondents themselves or by other people using their video card, the chances that the respondent and the chances that others would make a purchase would need to be combined. This was calculated as follows:

$$E(a) = 1 - ((1 - P(\text{self})) * (1 - P(\text{others})))$$

where  $P(\text{self})$  = the probability that the respondent will make a purchase

and  $P(\text{others})$  = the probability that some other person will make a purchase

As  $P(\text{self})$  is the probability that the respondent will make a purchase,  $1 - p(\text{self})$  is the probability that the respondent will not make a purchase. The intersection between the probabilities that the respondent will not make a purchase and others will not make a purchase, that is  $(1 - p(\text{self})) * (1 - p(\text{others}))$ , is equal to the probability that a purchase is not made. One minus the probability that a purchase is not made then is equal to the probability that at least one purchase is made.

#### 5.10.4 Expected value calculation for purchase level

For purchase level predictions, that is, the number of videos that will be hired, there were four ways in which the expected number of videos were calculated, they were;

$$E(n) 1 = P(\text{any}) * N * P(N)$$

$$E(n) 2 = P(\text{any}) * N$$

$$E(n) 3 = N * P(N)$$

$$E(n) 4 = N$$

where:  $P(\text{any})$  = the probability of hiring at least one video in the time period

$N$  = the number of videos the respondent is most likely to hire

$P(N)$  = the probability that the respondent will hire exactly  $N$  videos

When purchase level predictions were first made, a somewhat cumbersome method of calculating the expected value was employed. The respondent was required to give the probability that they would purchase various quantities of the product. This meant that the interview was long and repetitive. However, Brennan, Esslemont and Hini (1995) found that by simply asking about the probability of purchasing any, the number most likely to purchase and the chances of purchasing exactly that number, predictions with accuracy comparable to that of the multiple question method was achieved. The calculation developed by Brennan et al is used in this study as  $E(n) 1$ ,  $P(\text{any}) * N * P(N)$ . The three other methods ( $P(\text{any}) * N$ ,  $N * P(N)$  and  $N$ ) are the three other possible combinations that may also give accurate predictions, and further reduce the questioning required.

To calculate the purchase rate, special manipulation of the probability that the respondent would hire are least one video and the probability that other people would hire at least one video was required to calculate the expected value. However, to calculate the expected purchase level, the calculations were much simpler. The expected value was the sum of the expected value for the respondent and the expected value for other people. In other words, the total number that is expected to be hired is the combination of number that the respondent

is expected to hire, plus the number that other people are expected to hire. This is shown in the following formula:

$$E(n) = E(n)_{\text{self}} + E(n)_{\text{others}}$$

where  $E(n)_{\text{self}}$  = the expected value for the respondent

and  $E(n)_{\text{others}}$  = the expected value for other people

#### 5.10.5 Actual value

In order to assess the accuracy of the prediction, the “actual” value (the number of videos hired), is subtracted from the expected value (the number of videos they predicted they would hire). However, there are two types of “actual” value available. In previous studies (for example see Brennan, Hini and Esslemont 1994) the “actual” value has been the number recalled by the respondent. However, respondent recall is subject to many biases. To avoid the recall biases, as well as collecting the respondent's recall, the video store's computer records were also collected as an accurate, unbiased record of the actual number of videos hired.

To assess the accuracy of the predictions, the computer records were used as the actual value. The number of videos the respondent recalled hiring was used primarily to assess the affect of using recall data as a surrogate for actual behaviour, in this study the computer records (H6).

As the respondent was making predictions from the time of the interview, it was necessary to have actual data from the time of the interview. However, the computer records only recorded the number of videos hired for the entire day. It did not display the time that the videos were hired. This is a problem, as the respondents were making predictions for the four weeks from the time of the initial interview. This would obviously include any videos hired later in the day, but exclude any already hired earlier that day.

Unfortunately, the computer records only recorded the total number of videos hired on the day, and did not show at what time the videos were hired. This means that a respondent

could hire a video, after which they may be interviewed and predict that they would not hire any videos for the rest of the day, yet the computer records would show that a video was hired on that day. To allow for this possibility, respondents were asked in the initial interview how many videos they had hired so far that day. This number was then subtracted from the number of videos hired on the computer records. This results in the number of videos hired on the day of the initial interview, but after the interview, based on the assumption that there was no error in the respondent's recall for that day. For example, if a respondent reported they had already hired two videos on that day, and the computer records showed four videos were hired on the day of the interview, it was assumed that two videos (four hired on the day from the computer records minus the two reported by the respondent) were hired after the interview.

One of the hypotheses (H6) was that the error in predictions calculated using recall as a surrogate for actual behaviour would not be different from the error in predictions calculated using actual behavioural data. This requires recall data and actual data to be collected for comparison. The recall data was collected in the second telephone interview, while the actual data was collected from the video store's computer which keeps a record of all video tapes hired.

A problem arises when the respondent is recalling how many videos they had hired since they were interviewed four weeks previously, as they could still hire videos later in the day after the recall interviews had been conducted and so the computer records would show a greater number of tapes for that day than were hired at the time of the interview. The respondents were required to recall how many videos they and others had hired over the survey period from the time that they were last interviewed. As the interviewing was conducted throughout the day, there is a chance that after someone had been interviewed, they could still hire some videos later that day. The videos hired after the recall interview would still be recorded on the computer, but the respondent would not recall them as being hired. In this case, the respondent may have correctly recalled that they had not hired any videos when asked, even though the computer records show some videos were hired that day.

There are two ways that this problem could be avoided. One is to omit the day of the

recall interview from the study period. The respondents could simply be asked to recall how many videos they had hired up until the day before the recall interview. However, although every attempt was made to conduct the recall interview on the correct day, inevitably, some would be reinterviewed late. If the person was reinterviewed late, it would be possible to ask them to recall the number of videos hired from the time of the first interview until the day that the recall interview was supposed to have been conducted. However, this approach would then be open to both forward and backwards telescoping. The respondents could recall videos hired from before the first interview, as well as videos hired after the time when the recall interview should have been conducted as being hired within the survey period. The longer the delay from when the interview was supposed to have been conducted until it actually was, the greater the chances of forward telescoping. Furthermore, as the predictions were made for the entire period, omitting the final day of the prediction period would introduce bias into the results as the prediction period would not match the recall period.

The second way to avoid the problems, and the method that was used, is as follows. It was decided that the actual data would need to be taken for the entire four week period after the initial interview. However, this then resulted in the problem that the respondent could still hire videos later in the day after the recall interview. The computer records would show the number of videos hired for the entire day, but the respondent would only be recalling for part of the day. To avoid any apparent anomaly that would result if a video was hired after the respondent had been interviewed, the following calculation was used to calculate how many videos the respondent recalled as hiring in the four week period:

Number recalled for the entire period - number hired on the last day + computer records for the number hired on the last day.

The respondents were asked to recall how many videos they had hired for the entire period, and how many they had hired so far that day. The number that they claimed to have hired that day was then subtracted from the total they claimed they had hired for the entire period. The effect of this calculation, is to give the number of videos the respondents claimed they had hired up until the day before the recall interview. The computer records were then used to find out how many videos were hired on the day of the recall interview. This number was

then added to the number the respondent recalled they had hired for the period up until the day before the recall interview.

The above calculations give a value that could be compared with the computer records for the entire period from the time of the first interview until the day of the recall interview.

If the respondents were interviewed late, that is, the interviewer was not able to interview the respondent exactly four weeks after the initial interview, the respondents were simply asked to recall the number of videos they had hired since the initial interview, and how many videos they had hired since the day when the interview was supposed to have been conducted. This second number was then subtracted from the first to give the number recalled as being hired in the four weeks after the initial interview.

## 6. RESULTS

### 6.1 Introduction

The Method section outlined the procedures to test the hypotheses stated in the Hypotheses section. This section described the results of this study.

This section begins with an examination of the effect of recall prior to predictions (Single Recall and Bounded Recall) and the effect of the bounded prediction procedure for purchase rate predictions and then purchase level predictions.

The effect that the four methods of calculating purchase level estimates has on the purchase level error in predictions is examined. Following this is an appraisal of the effect of recall prior to predictions, and the effect of the bounded prediction procedure for purchase level predictions.

In section 6.4, the effect of using recall data to calculate errors compared with using actual data is examined for both purchase rate and purchase level error calculations.

Section 6.5 looks at the effect of accuracy of recall immediately prior to making predictions on the accuracy of predictions (H7), firstly for purchase rate and then purchase level predictions.

Finally, section 6.6 details the effect that usage has on the accuracy of purchase rate and purchase level predictions. This analysis shows that non-users had larger errors in their purchase rates and purchase level predictions than those who had used their video card. This finding prompts a new method of calculating the purchase level, based on identifying who are users, and then by assuming non-users have a zero purchase probability, the purchase level of the market is equal to the purchase level of users.

This new method of estimating purchase levels is then tested using the computer records to identify the users and non-users. Having established the new method improved purchase level predictions, the analysis was performed again, but used the respondent recall to identify users and non-users.

## 6.2 Effect of Recall Assisted and Bounded Prediction Questions on Predictions

### 6.2.1 Introduction

In the Beach et al model of subjective probability, recall is a central component in making predictions. To make a prediction, a person needs to recall either their personal knowledge about a situation, or their statistical knowledge about the frequency with which an event occurs. If the accuracy of the recall could be improved, then it would follow that the resulting predictions should also be improved. Therefore, it is hypothesised that (H1) asking a single recall question prior to the prediction questions will result in smaller errors in predictions than asking the prediction questions without any previous recall questions and that (H2) asking a bounded recall question prior to the prediction questions will result in smaller errors in predictions than asking the prediction questions without any previous recall questions. As it is expected that bounded recall would be effective at reducing telescoping in recall, it was hypothesised that (H3) asking a bounded recall question prior to the prediction questions will result in smaller errors in predictions than asking a single recall question prior to the prediction questions.

It has been suggested that bounded recall is effective at reducing recall error because the technique provides cues to the respondent to provide accurate answers, in other words, bounding increases the respondent's motivation to provide accurate answers. Motivation to provide accurate answers is also the third component in Beach et al's model. Therefore, it is hypothesised that (H4) using a bounded prediction question technique will result in smaller errors in predictions than predictions from using only a single prediction question.

This part of the results section examines these four hypotheses, firstly with respect to purchase rate predictions (that is, the proportion that will hire at least one video), and then purchase level predictions (that is, the total number of videos that will be hired).

One tailed Student t-tests were used to test the significance of any differences between group

errors. This was performed using SPSS-PC+'s two tailed Student t-test. The probabilities (level of significance) computed by SPSS-PC+ were then divided by 2 to give the one tailed probability (SPSS Base System Syntax Reference Guide Release 6.0, 1993). It is this one tailed probability that is reported in the significance columns of Tables 6 and 7.

As described in the Method section, the data for the predictions were collected in two phases, with Control Group A, the Single Recall Juster and the Bounded Recall Juster groups all interviewed in Phase Two, while the predictions from the control groups B and C, and the Bounded Juster group were collected in Phase One of interviewing. There is the possibility that uncontrollable stimuli could have affected the accuracy of the predictions of the treatments in Phase One differently to the treatments in Phase Two. Therefore, it is necessary to partition the table into the two groups of treatments that were interviewed in the same phase.

#### 6.2.2 Effect of recall assisted and bounded prediction questions on purchase rate predictions

The accuracy of the purchase rate predictions for the six treatments are shown in Table 6. It should be noted that the actual purchase rates are identical for all treatment groups, as the sample was post-stratified by the number of videos hired (see the Method section). Therefore, the number of respondents who hired on at least one occasion was equal for all groups.

Table 6. Effect of recall assisted and bounded prediction questions on the predictive accuracy of purchase rate estimates.

	n	Purchase Rate			Sig
		Predicted <sup>1</sup>	Actual <sup>2</sup>	Error <sup>3</sup>	
T1 Control A	100	0.56	0.31	0.25	-
T2 Single Recall	100	0.67	0.31	0.36	.081 <sup>a</sup>
T3 Bounded Recall	100	0.57	0.31	0.26	.440 <sup>b</sup> .086 <sup>c</sup>
T4 Control B	100	0.74	0.31	0.43	-
T5 Control C	100	0.73	0.31	0.42	-
T6 Bounded Juster	100	0.68	0.31	0.38	.246 <sup>c</sup> .278 <sup>d</sup>

- Note
- 1 Predicted Purchase Rate = The proportion of the sample who said they would hire at least one video
  - 2 Actual Purchase Rate = The proportion of the sample who did hire at least one video (measured from the computer records)
  - 3 Error = Predicted - Actual
  - a Significance of difference between T2 and T1,  $t = 1.41$ ,  $df = 198$
  - b Significance of difference between T3 and T1,  $t = 0.15$ ,  $df = 198$
  - c Significance of difference between T3 and T2,  $t = 1.37$ ,  $df = 198$
  - d Significance of difference between T6 and T4,  $t = 0.69$ ,  $df = 198$
  - e Significance of difference between T6 and T5,  $t = 0.59$ ,  $df = 198$

As expected, the errors in purchase rates (predicted - actual) are generally quite small, although none of the errors of the experimental groups are significantly smaller than those of the control groups.

It was hypothesised that asking a bounded prediction question (the Bounded Recall group, T3) prior to the prediction questions would result in smaller errors in predictions than asking a single recall question (the Single Recall group, T2) prior to the prediction question. Table 6 shows that the error in purchase rate predictions was smaller for the Bounded Recall group (T3) than the Single Recall group (T2). However, the difference was only significant at the 90% level, thus there is only weak evidence to accept H3. It was also hypothesised that asking a single recall question prior to the prediction questions would result in smaller errors in predictions than asking the prediction questions without any previous recall questions (H1) and that (H2) asking a bounded recall question prior to the prediction questions would result in smaller errors in predictions than asking the prediction questions without any previous

recall questions.

However, as Table 6 shows, both the Single and Bounded Recall groups (T2 and T3) had larger errors than their control group (T1). Although the error for the Bounded Recall group is virtually identical to that of the control group, the error for the Single Recall group was significantly larger than the Control Group error at the 90% level. Thus there is no evidence to accept H1 or H2.

Thus H1 for purchase rate predictions is rejected

H2 for purchase rate predictions is rejected

H3 for purchase rate predictions is tentatively accepted

It was hypothesised that using a bounded prediction question technique would result in smaller errors in predictions than predictions from using only a single prediction question. The Bounded Juster group (T6), achieved a smaller error in predictions than both of its control groups, although though not significantly so. This provides weak evidence supporting the hypothesis that a Bounded Juster technique would result in more accurate predictions being made.

Thus H4 for purchase rate predictions is tentatively accepted

### 6.2.3 Effect of recall assisted predictions and the effect of a bounded prediction question on purchase level predictions.

As well as obtaining purchase rate estimates, purchase level data was also collected. As detailed in the methodology section four purchase level estimates were calculated. These four methods of estimating the purchase level are shown in the "Predicted" column for each of the treatment groups. Again the "Actual" number hired is the same across each group as the sample was post-stratified. The "Error" is the predicted (using the four error calculation methods) minus the actual from the computer records. The hypotheses regarding the effect of the treatments on purchase level estimates are the same as the hypotheses for purchase rate.

Table 7. Effect of recall assisted and bounded prediction questions on the predictive accuracy of purchase level estimates.

	n	Purchase Level				Actual <sup>2</sup>	Error (Predicted - Actual)							
		Predicted <sup>1</sup>					P(A).N.P(N)	Sig	N.P(N)	Sig	P(A).N	Sig	N	Sig
		P(A).N.P(N)	N.P(N)	P(A).N	N									
T1 Control A	100	1.19	1.67	1.75	2.78	0.69	0.50	-	0.98	-	1.06	-	2.09	-
T2 Single Recall	100	1.56	2.01	2.16	3.11	0.69	0.87	.096a	1.32	.128a	1.47	.091a	2.42	.162a
T3 Bounded Recall	100	1.20	1.76	1.80	2.90	0.69	0.51	.483b .094c	1.07	.374b .194c	1.11	.431b .114c	2.21	.361b .268c
T4 Control B	100	2.05	2.61	2.94	3.90	0.69	1.36	-	1.92	-	2.25	-	3.21	-
T5 Control C	100	1.91	2.37	2.81	3.73	0.69	1.22	-	1.68	-	2.12	-	3.04	-
T6 Bounded Juster	100	1.35	1.77	2.04	2.88	0.69	0.66	.020d .057e	1.08	.012d .056e	1.35	.016d .028e	2.19	.015d .031e

- Note. 1 Predicted Purchase Level = the predicted number of videos that will hired  
 2 Actual Purchase Level = the actual number of videos hired (from computer records)  
 a Significance of difference between T2 and T1,  $t(P(A).N.P(N)) = 1.31$ ;  $t(N.P(N)) = 1.14$ ;  $t(P(A).N) = 1.34$ ;  $t(N) = 0.99$ ,  $df = 198$   
 b Significance of difference between T3 and T1,  $t(P(A).N.P(N)) = 0.04$ ;  $t(N.P(N)) = 0.32$ ,  $t(P(A).N) = 0.15$ ,  $t(N) = 0.36$ ,  $df = 198$   
 c Significance of difference between T3 and T2,  $t(P(A).N.P(N)) = 1.32$ ;  $t(N.P(N)) = 0.87$ ;  $t(P(A).N) = 1.21$ ;  $t(N) = 0.62$ ;  $df = 198$   
 d Significance of difference between T6 and T4,  $t(P(A).N.P(N)) = 2.07$ ;  $t(N.P(N)) = 2.29$ ;  $t(P(A).N) = 2.17$ ;  $t(N) = 2.19$ ;  $df = 198$   
 e Significance of difference between T6 and T5,  $t(P(A).N.P(N)) = 1.59$ ;  $t(N.P(N)) = 1.60$ ;  $t(P(A).N) = 1.93$ ;  $t(N) = 1.88$ ;  $df = 198$

It was hypothesised that asking a single recall question prior to the prediction questions would result in smaller errors in predictions than asking the prediction questions without any previous recall questions (H1). For each of the purchase level calculation methods (Table 7), the error in predictions from the Single Recall (T2) treatment were larger than the error in the control group (T1) irrespective of the calculation method used. Although none of the differences were significant at the 95%, two were significant at the 90% level, thus there is no evidence to accept H1.

It was also hypothesised that asking a bounded recall question prior to the prediction questions would result in smaller errors in predictions than asking the prediction questions without any previous recall questions (H2). The errors in predictions from the Bounded Recall groups (T3) were all consistently larger than the control group's (T1) irrespective of the calculation method used. However, the difference between the errors were marginal and none of the differences in errors were significant, thus there is no evidence to accept H2.

H3 stated that asking a bounded recall question prior to the prediction questions would result in smaller errors in predictions than asking a single recall question prior to the prediction questions. Looking at the differences in the errors between the Bounded Recall (T3) and the Single Recall (T2) groups, it can be seen that the errors for the Bounded Recall group were consistently smaller than the Single Recall group for each of the error calculation methods used. However, none of the differences in errors were statistically significant at the 95% level. One (the calculation method using the formula  $P(a).N.P(N)$ ) was significantly smaller for the Bounded Recall (T3) group than the Single Recall (T2) group at the 90% level, thus there is weak evidence to tentatively accept H3.

Thus H1 for purchase level predictions is rejected

H2 for purchase level predictions is rejected

H3 for purchase level predictions is tentatively accepted

It was hypothesised that using a bounded prediction question technique would result in smaller errors in predictions than predictions from using only a single prediction question (H4). The Bounded Juster (T6) consistently produced smaller errors than either of its control

groups (T4 and T5), with six of the differences in errors significant at the 95% level. This provides further evidence to that found for purchase rate errors that bounding the prediction question will improve the accuracy of the predictions.

Thus H4 for purchase level predictions is accepted

#### 6.2.4 Summary

This part of the Results section examine four hypotheses. It was found that there was no evidence to provide support for hypotheses 1 or 2. Thus, these hypotheses were rejected for both purchase rate predictions and purchase level predictions.

There was weak evidence in support of H3.

Of the hypotheses examined in this part of the Results section, Hypothesis 4 had the most evidence in support of it. Whilst there was only weak evidence supporting H4 with respect to purchase rate predictions, the evidence was much more compelling for purchase level predictions where significant differences between the Bounded Juster and control groups was seen.

### 6.3 Effect of Calculation Method of Purchase Level Predictions

#### 6.3.1 Introduction

Brennan, Esslemont and Hini (1995) moved away from the logically complete method of calculating purchase level predictions which used the probability of purchasing exactly  $n$  units, for all values of  $n$  with a non-zero purchase probability, to just using  $P(A).N.P(N)$ . Thus it appears that purchase level calculations do not need to be logically correct to predict accurately. Therefore, it was hypothesised that the questioning required could be reduced without logical reasons for doing so without adversely affecting the accuracy of the predictions. Thus it was hypothesised that reduced prediction questioning procedures will not affect the accuracy of purchase level prediction (H5).

This part of the Results section examines the results of tests of H5. SPSS-PC+'s two tailed paired Student t-test was used to analyse the difference in errors calculated between the four purchase level calculation methods.

#### 6.3.2 Effect of calculation method on purchase level errors.

The effect of the four methods of calculating the predicted purchase level on the purchase level error can be seen in Table 8. Table 8 shows the error, t value and significance of the t between the errors of the calculation methods.

Rather than displaying the significance of the differences between every pairing of purchase level estimate calculation methods, the estimate calculation methods have been ranked in order of accuracy (i.e. from smallest error to largest error) for each treatment and the significant between consecutive calculation methods have been displayed. The significances that have been calculated are the difference between  $P(A).N.P(N)$  and  $N.P(N)$ , between  $N.P(N)$  and  $P(A).N$  and between  $P(A).N$  and  $N$ . As the error calculated using the

P(A).N.P(N) method is significantly smaller than the second most accurate method, it is therefore significantly more accurate than all of the other methods. Conversely, as the N method is significantly less accurate than the P(A).N method, it is therefore significantly less accurate than all of the other methods.

Table 8. Effect of purchase level calculation method on calculated error in purchase level

	Purchase Level Calculation Method										
	df	P(A).N.P(N)			N.P(N)			P(A).N			N
		Error <sup>1</sup>	t (from N.P(N))	Sig (from N.P(N))	Error <sup>1</sup>	t (from P(A).N)	Sig (from P(A).N)	Error <sup>1</sup>	t (from N)	Sig (from N)	Error <sup>1</sup>
T1 Control A	99	0.50	7.95	.000	0.98	0.92	.361	1.06	9.45	.000	2.09
T2 Single Recall	99	0.87	7.66	.000	1.32	1.80	.075	1.47	8.68	.000	2.42
T3 Bounded Recall	99	0.51	8.18	.000	1.07	0.36	.718	1.11	11.13	.000	2.21
T4 Control B	99	1.36	6.74	.000	1.92	2.31	.023	2.25	8.31	.000	3.21
T5 Control C	99	1.22	8.02	.000	1.68	3.48	.001	2.12	8.20	.000	3.04
T6 Bounded Juster	99	0.66	5.66	.000	1.08	3.04	.003	1.35	7.15	.000	2.19

Note: The error reported is the error in purchase level prediction using the specified purchase level calculation method. These errors are reproductions of those found in Table 7.

It can be seen that the P(A).N.P(N) method of calculating the number that will be hired was consistently the most accurate method for all treatments. Across all treatments, the N.P(N) method was second, then the P(A).N method and the least accurate method was simply using N as a means of calculating purchase levels. Furthermore, the P(A).N.P(N) method results in errors that are significantly smaller than the next most accurate method N.P(N) irrespective of the treatment. The difference in the error between the N.P(N) method and the P(A).N tended to be closer, with the difference between the calculation methods for Control A (T1), Single Recall (T2) and Bounded Recall (T3) not significant at the 95% level. Using just N as a estimate of purchase level was significantly less accurate than using P(A).N across all treatments.

Thus, there is strong evidence that all three questions, that is,

*Now, taking everything into account, what are the chances of you personally hiring at least one video from [VIDEO STORE NAME], during the next four weeks, that is between now and <DAY AND DATE>?*

*If you did hire any, how many videos are you personally most likely to hire from [VIDEO STORE NAME] in the next four weeks?*

*Using the scale from 0 to 10, what are the chances that you personally would hire exactly <MOST LIKELY NUMBER> videos from [VIDEO STORE NAME] in the next FOUR weeks?*

need to be asked to calculate a prediction that maximises the predictive accuracy.

Thus H5 is rejected

### 6.3.3 Summary

In this part of the Results section whether the questioning could be reduced without adversely affecting the accuracy of the purchase level predictions was examined. It was found that

there was significant differences in the errors depending upon which purchase level calculation method was used with the calculation method P(A).N.P(N) significantly more accurate than the second most accurate calculation method N.P(N). Thus there was strong evidence to reject the hypothesis.

## **6.4 Effect of using recall as a surrogate for actual behaviour for testing the accuracy of predictions.**

### 6.4.1 Introduction

The formula of  $(\text{predicted} - \text{actual})/\text{actual}$  has often been used to assess how accurate the predictions were. However, the actual data has usually been the number the respondent recalled they had purchased, not how many were actually purchased. As there are several biases that have been shown to affect the accuracy of respondent recall, there is good reason to question the validity of using recall data as a surrogate for actual data.

Of the three main recall biases studied, telescoping is considered to be the main affect, resulting in over-reporting. However, it has been shown that using a landmark event, that is, some event that acts as a clear boundary of the recall period, is effective at reducing the amount of over-reporting in recall question.

As the initial interview was used to collect purchase probabilities, it was expected that, that interview would act as a personal landmark event for the respondent, providing a boundary to aid the respondent's recall, reducing the occurrence of telescoping, and ultimately leading to more accurate recall.

As the number the respondent recalled they had hired would be more accurate than recall without a landmark event, and therefore more similar to the number that they had actually hired (as determined from the computer records) there would not be a significant difference in the calculated error in predictions when recall data was used as a surrogate for actual data compared to the calculated error when using the actual data (the computer records). Thus it hypothesised (H6) that the accuracy of predictions calculated using recall as a surrogate for actual behaviour will not be different from the accuracy of predictions calculated using actual behavioural data.

SPSS-PC+'s two tailed paired Student t-test was used to test this hypothesis by measuring the significance of the difference between the error in predictions calculated using computers (that is, predicted - computer records) and the error in predictions calculated using recall (that is, predicted - recall)

#### 6.4.2 Effect of using recall as a surrogate for actual behaviour for testing the accuracy of purchase rate predictions

Table 9 shows the effect of using recall as a surrogate for actual purchase data on calculated purchase rate errors. The column headed Predicted - Computer Record is the mean purchase rate errors using the formula number predicted minus number hired from computer records. The next column is the mean purchase rate errors using the formula number predicted minus number respondent recalled hiring.

Table 9. Purchase rate error using computer records and recall as a surrogate for computer records in the error calculations

	n	Purchase Rate Error		t	Sig	df
		Predicted - Computer Record	Predicted - Recall			
T1 Control A	100	0.25	0.18	1.41	.163	99
T2 Single Recall	100	0.36	0.26	2.17	.032	99
T3 Bounded Recall	100	0.26	0.22	0.85	.396	99
T4 Control B	100	0.43	0.25	3.75	.000	99
T5 Control C	100	0.42	0.31	2.24	.027	99
T6 Bounded Juster	100	0.38	0.26	1.87	.064	99

The results in Table 9 show that when recall is used as a surrogate for actual behaviour, the calculated errors in predictions are smaller than the errors calculated using the data collected from the computer records with four of the six errors significantly smaller.

Thus H6 for purchase rate predictions is rejected

#### 6.4.3 Effect of using recall as a surrogate for actual behaviour for testing the accuracy of purchase level predictions

Table 10 compares the errors in purchase level predictions when computers records are used as the actual data, compared with when the respondent's recall is used in place of actual data. The results are displayed for the four methods of calculating the purchase level. The Sig column is the significance of the difference between the error calculated using computer records as the actual value compared to the error in predictions calculated when recall data is used as the actual number of videos hired in a two tailed paired Student t-test.

Table 10. Effect of using recall as a surrogate for actual behaviour for purchase level estimates

	n	Error in purchase level prediction <sup>1</sup>								t	Sig <sup>2</sup>	df
		P(A).N.P(N)		P(A).N		N.P(N)		N				
		Computer	Recall	Computer	Recall	Computer	Recall	Computer	Recall			
T1 Control A	100	0.50	0.34	1.06	0.90	0.98	0.82	2.09	1.93	1.13	.260	99
T2 Single Recall	100	0.87	0.49	1.47	1.09	1.32	0.94	2.42	2.04	2.76	.007	99
T3 Bounded Recall	100	0.51	0.34	1.11	0.94	1.07	0.90	2.21	2.04	1.36	.176	99
T4 Control B	100	1.36	0.61	2.25	1.50	1.92	1.17	3.21	2.46	4.38	.000	99
T5 Control C	100	1.22	0.74	2.12	1.64	1.68	1.20	3.04	2.56	2.54	.013	99
T6 Bounded Juster	100	0.66	-0.07	1.35	0.62	1.08	0.35	2.19	1.46	3.08	.003	99

Note 1: The errors reported are based on the formula (Predicted - Actual). The errors in the columns headed Computer have used the computer records as the "Actual" value whereas the columns headed Recall have used the respondent's recall as the "Actual" value.

2: The significance measured with the significance of the difference in the errors calculated using computer records and using recall, within each of the four calculation methods. The significance of the differences is the same across the four calculation methods as only the predicted amount changes with each calculation method, not the actual.

As was found with purchase rate predictions, in all treatments, the error in predictions calculated using recall as a surrogate for actual data (in this case, the computer records) was less than the actual error in predictions, based on computer records (calculated as the predicted - the number recorded on the computer). Four of the treatments had significantly smaller errors calculated when recall data was used as opposed to the computer record data as the actual data, providing further evidence to reject H6.

Thus H6 for purchase level predictions is rejected

#### 6.4.4 Summary

This part of the Results section examined the effect that using respondent recall has on the calculated error in predictions. For both purchase rate and purchase level predictions, using respondent recall as a surrogate for actual behaviour resulted in significantly smaller errors being calculated, in other words, there was a difference in the calculated error depending upon whether actual data, or recall data was used in the error calculation. As such, H6 which stated that the accuracy of predictions calculated using recall as a surrogate for actual behaviour would not be different from the accuracy of predictions calculated using actual behavioural data was rejected.

## 6.5 Effect of accuracy of prior recall on predictions

### 6.5.1 Introduction

In Beach et al's model of subjective probability, recall is a central component. For a respondent to make a judgement concerning the chances that an event will occur, they need to use either their recall of the specific circumstances concerning the event, or they need to recall the frequency with which the event occurs to base their judgement on. As recall is a central part in the model of subjective probability formulation, it was hypothesised (H7) that respondents with accurate recall at the time of making a prediction will have less error in their predictions than those with inaccurate recall at the time of making the prediction. This part of the Results section reports the findings from tests of this hypothesis.

A scree test to ensure that the groups (accurate versus inaccurate recall) were reasonably balanced, thus ensuring statistically large samples so that comparison between the groups could be made with some statistical reliability. The respondents were categorised as "None" if they made no error in their recall, or "Some" if they made some error in their recall of their previous four week hiring behaviour compared to the computer records. A one tailed Student t-test was then used to analyse the significance of the difference in errors between those with error in their recall and those without. To do this, SPSS-PC+'s two tailed Student t-test was used, with the two tailed probability being divided by two to calculate the one tailed probability.

Only two groups were asked to recall their previous hiring behaviour prior to the prediction questions; the Single Recall (T2) and Bounded Recall (T3) groups. Therefore, the accuracy of the recall prior to the prediction questions of only these two groups can be assessed, and hence, assessing the effect of the accuracy of the respondent's recall can only be performed on the two groups.

### 6.5.2 Effect of accuracy of prior recall on purchase rate predictions

Table 11 shows the result of analysis by the level of error in the respondent's recall of their previous hiring.

Table 11. Effect of accuracy of recall on predictive accuracy of purchase rate

	Recall error	n	Purchase Rate			t	Sig	df
			Predicted	Actual	Error			
T2 Single Recall	None	53	0.60	0.26	0.34	0.40	.345	98
	Some	47	0.75	0.36	0.38			
T3 Bounded Recall	None	65	0.48	0.22	0.26	0.10	.460	98
	Some	35	0.74	0.49	0.25			

For purchase rate estimates, whether the respondent had any error in their recall of previous behaviour prior to giving their future purchase probabilities did not have any significant effect on the subsequent predictions. The errors in predictions were very similar between those with and without error in their recall and for neither of the treatments was the difference statistically significant. This suggests that for purchase rate estimates, the accuracy with which a respondent can recall their previous purchasing does not affect the accuracy of their purchase rate predictions for these items.

Thus H7 for purchase rate predictions is rejected

### 6.5.3 Effect of accuracy of prior recall on purchase level predictions

Table 12 shows the result of analysis of purchase level predictions, by the level of error in the respondent's recall prior to being asked the prediction questions.



For both of the treatment groups, those with no error in their recall had more accurate purchase level predictions than those who had some error in their recall, with these differences in predictive errors statistically significant for all four methods of calculating the error, providing evidence to accept H7.

Thus H7 for purchase level predictions is accepted.

#### 6.5.4 Summary

This section examined the effect that recall has on predictions. It was found that the accuracy of recall prior to making purchase rate predictions does not appear to affect the accuracy of the prediction. However, the accuracy of recall prior to making a purchase level prediction does affect the accuracy of the prediction. Thus, H7 that states respondents with accurate recall at the time of making a prediction will have less error in their predictions than those with inaccurate recall at the time of making the prediction was rejected for purchase rate predictions, but accepted for purchase level predictions.

## 6.6 Effect of usage

### 6.6.1 Introduction

In a study by U (1991) it was found that when the usage level was low, the predictions were very inaccurate. Parfitt (1967) found that recall was less accurate for infrequently purchased products. It would follow then, that as recall is less accurate for infrequently purchased products, through H7 which states that respondents with accurate recall at the time of making a prediction will have less error in their predictions than those with inaccurate recall at the time of making the prediction, there will be more error in predictions for infrequently purchased products. Thus H8 states errors in predictions for low users will be greater than the errors in predictions for high users.

This section presents the findings from the analysis investigating Hypothesis 9. A scree test was used to define high and low users. As almost two thirds of the sample had not hired a video over the four week prediction period, low users were defined as those who had not used their video card (that is, those who had not hired any videos) whilst high users were defined as those who had used their video card to hire a video. To more accurately describe the groups, in Tables 13 and 14, the low users group was renamed as "Not" (for Not Used) and high users were named "Used".

Based on the findings testing Hypothesis 8, a new method of calculating purchase levels estimates is developed. This method is based on identifying users, and then calculating purchase level estimates of users.

A one tailed Student t-test was used test the significance of the difference in purchase rate error between users and non users. This was done by dividing SPSS-PC+'s two tailed Student t-test probabilities by two. Significance testing was only performed on the purchase rate predictions. The significance measures was the difference between the absolute error in predictions of those who had used their video card, against the absolute error in predictions of

those who had not used their video card. In other words, the difference being measured for Control A for instance, is the significance of the difference between 0.55 and 0.42 (see Table 13), a difference of 0.13, not the total difference between 0.55 and -0.42 of 0.97. The difference between the absolute errors was used for analysis in this table because those who had not used their video card must make a error in their prediction ranging from 0 (if they predicted they would not hire) through to 1 (if they were certain they would hire but did not), while those that did hire could only make errors ranging from 0 (if they were certain that they would hire) through to -1 if they predicted that they would not hire but did. As those that did not hire could only make positive errors and those who did hire could only make negative errors in their predictions, it was necessary to use the absolute error to meaningfully compare the two errors.

Significance testing was not performed for the purchase level estimates as it was not possible to compute the significance of a meaningful difference in errors of purchase level predictions.

For purchase rate calculations the error must be zero or positive for those who had not hired and zero or negative for those who had hired, thus taking the absolute error creates a meaningful comparison. However, for purchase level calculations the error for those who had hired could be positive or negative. The respondents who did hire could predict a number that was either too high, giving a positive error, or a number too low, giving a negative error. The respondents who did not hire, obviously actually hired zero videos, and therefore could still only have errors that were zero or positive. As those who did hire could have positive or negative errors, taking the mean of the absolute individual errors results in an error different to the absolute mean of the individual errors (the errors shown in Table 8).

Taking the mean of the absolute individual errors negates the under-estimations that cancel out the over-estimations which is vital to get accurate predictions. For example, the mean of  $\{|-4|, |-3|, |3|, |4|\} = 3.5$  whereas the absolute mean of  $\{-4, -3, 3, 4\} = |0| = 0$ .

It would be possible to calculate the absolute difference between the two aggregate errors (for example, 1.12 and .88 for Control A in Table 14) then find the standard errors for each group and calculate the significance of the difference manually. However, the standard errors were only by SPSS-PC+ to two decimal places, which, for such small standard errors, was considered to be too coarse to be used.

### 6.6.2 Effect of usage on purchase rate predictions

Table 13 shows the error in purchase rate for those who had used their video card to hire a video and those who had not used their video card to hire a video.

Table 13. Effect of being a user or non-user on predictive accuracy of purchase rate

	Usage	n	Purchase Rate			t	Sig	df
			Predicted	Actual	Error			
T1 Control A	Not	69	.55	.00	0.55	1.73	.044	98
	Used	31	.58	1.00	-0.42			
T2 Single Recall	Not	69	.64	.00	0.64	5.10	.000	98
	Used	31	.75	1.00	-0.25			
T3 Bounded Recall	Not	69	.50	.00	0.50	3.31	.001	98
	Used	31	.73	1.00	-0.27			
T4 Control B	Not	69	.72	.00	0.72	7.38	.000	98
	Used	31	.77	1.00	-0.23			
T5 Control C	Not	69	.68	.00	0.68	7.42	.000	98
	Used	31	.84	1.00	-0.16			
T6 Bounded Juster	Not	69	.65	.00	0.65	5.91	.000	98
	Used	31	.77	1.00	-0.23			

As hypothesised, those who had used their video card had smaller errors than those who had not. In fact, for each treatment, those who had used their video card had significantly less error in their purchase rate predictions than those who had not used their video card.

Thus H8 for purchase rate predictions is accepted.

### 6.6.3 Effect of usage on purchase level predictions

Table 14 shows the error in purchase level predictions for those who had used their video card to hire a video and those who had not used their video card to hire a video.

Table 14. Effect of being a user or non-user on predictive accuracy of purchase level

	Usage	n	Predicted				Actual	Error			
			P(ANY)* N*P(N)	P(ANY)*N	N*P(N)	N		P(ANY)* N*P(N)	P(ANY)*N	N*P(N)	N
T1 Control A	Not	69	1.12	1.67	1.63	2.68	0.00	1.12	1.67	1.63	2.68
	Used	31	1.35	1.95	1.77	3.00	2.23	-0.88	-0.28	-0.46	0.78
T2 Single Recall	Not	69	1.46	1.98	1.91	2.95	0.00	1.46	1.98	1.91	2.96
	Used	31	1.78	2.55	2.23	3.45	2.23	-0.45	0.33	0.01	1.23
T3 Bounded Recall	Not	69	0.79	1.33	1.32	2.48	0.00	.79	1.33	1.32	2.48
	Used	31	2.12	2.83	2.76	3.84	2.23	-0.11	0.61	0.53	1.61
T4 Control B	Not	69	2.16	3.18	2.72	4.20	0.00	2.16	3.18	2.72	4.20
	Used	31	1.81	2.40	2.39	3.23	2.23	-0.41	0.18	0.16	1.00
T5 Control C	Not	69	2.02	2.84	2.50	3.78	0.00	2.02	2.84	2.50	3.78
	Used	31	1.66	2.75	2.09	3.61	2.23	-0.57	0.52	-0.14	1.39
T6 Bounded Juster	Not	69	1.32	1.93	1.73	2.78	0.00	1.32	1.93	1.73	2.78
	Used	31	1.42	2.31	1.87	3.10	2.23	-0.81	0.08	-0.36	0.87

For every treatment, those who had used their video cards had smaller absolute errors in their purchase level predictions than the non-users irrespective of the method used to calculate the purchase level providing evidence to accept Hypothesis 8.

Thus H8 for purchase level predictions is accepted

Traditionally, predictions have been made based on calculating the average number of purchases that will be made for all of the respondents in the sample. An alternative way of viewing the total number of purchases that will be made is that, by definition, non-purchasers will make no purchases, while the purchasers must make all of the purchases. Therefore, the estimated total number of purchases that will be made is equal to the estimated total number of purchases made by purchasers. For example, if total sales = 50, sales to those who make purchases = 50 and sales to those who do not make any purchases = 0. Table 14 showed that much of the error in predictions was made by the non-users, but as the preceding discussion has highlighted, logically, non-users make no purchases, and therefore it can be assumed that non-users will have a purchase probability of zero. In other words, the purchase level for the sample can be estimated as the purchase level for the users in the sample.

By assuming that those who were determined to be non-users had a purchase probability of zero, then the purchase rate would be equal to the purchase rate of the users, multiplied by the proportion of the market that are users. For example, if users had a purchase level of 2.5 products, and users made up 10% of the entire market, then the average purchase level would be  $2.5 * .1 = .25$  products per person in the entire market. The results of this type of analysis can be seen in the following table.

Table 15 shows the predicted purchase rate for the users of their video cards (users were defined as those people who hired at least one video during the prediction period), followed by the proportion of the sample for each treatment that were users (.31). The next column is the predicted purchase level for the entire market calculated by multiplying the purchase level for users by the proportion of the market that are users. The "Act" column is the actual purchase level for the entire market, while the error columns are results of the predicted purchase level for the entire market - the actual purchase level for the entire market.

Table 15. Alternative method of calculating purchase level errors using computer data to determine users and non-users

	n	Predicted for users <sup>1</sup>				Prop <sup>2</sup>	Predicted for market				Act <sup>3</sup>	Error			
		P(ANY)* N*P(N)	P(ANY)*N	N* P(N)	N		P(ANY)* N*P(N)	P(ANY)*N	N* P(N)	N		P(ANY) *N*P(N)	P(ANY)*N	N* P(N)	N
T1 Control A	31	1.42	1.95	1.77	3.00	.31	0.44	0.61	0.55	0.94	.69	-0.25	-0.08	-0.14	0.25
T2 Single Recall	31	1.69	2.55	2.23	3.45	.31	0.53	0.80	0.70	1.08	.69	-0.16	0.11	0.01	0.39
T3 Bounded Recall	31	1.81	2.83	2.76	3.84	.31	0.57	0.89	0.86	1.20	.69	-0.12	0.20	0.17	0.51
T4 Control B	31	1.78	2.40	2.39	3.23	.31	0.56	0.75	0.75	1.01	.69	-0.13	0.06	0.06	0.32
T5 Control C	31	2.12	2.75	2.09	3.62	.31	0.66	0.86	0.65	1.13	.69	-0.03	0.17	-0.04	0.44
T6 Bounded Juster	31	1.66	2.31	1.87	3.10	.31	0.52	0.72	0.58	0.97	.69	-0.17	0.03	-0.11	0.28

Note 1: Users were defined as those people who hired at least one video during the prediction period.  
 2: Prop is the proportion of the sample that were users.  
 3: Act is the actual number of videos hired by very respondent in the sample.

Comparing the errors achieved using this method of purchase rate calculations with the traditional method used in Table 7, shows that for this product at least, predictions were greatly improved. The largest error in purchase level was only 0.51, compared with a range in purchase level errors of 0.50 to 3.21 when the traditional method was used (see Table 7).

For this analysis, the way in which the predicted purchase level is calculated does make a difference. It was found in Table 7, that using  $P(\text{ANY}) * N * P(N)$  consistently resulted in the smallest errors in predictions (compared to the other methods of calculating the purchase level). However, for this analysis, it can be seen that the lowest error is achieved using the  $P(\text{ANY}) * N * P(N)$  for the Bounded Recall and Control C groups, while the  $P(\text{ANY}) * N$  results in the lowest error for Control A, Control B and the Bounded Juster groups. The Single Recall group and Control Group B had their lowest errors using  $N * P(N)$  to calculate the predicted purchase level. This would suggest the need to use the appropriate prediction calculation, depending upon the treatment used, although, apart from simply using  $N$ , that is, the number most likely to be purchased, any of the prediction calculations would result in predictions that were reasonably close to the actual purchase level.

In most circumstances, actual usage data, such as computer records, would not be available. Furthermore, the preceding analysis used usage over the prediction period for defining users and non-users. Obviously, the definition of users and non-users must be able to be made at the time of collecting the probability data if it is to be used to made predictions. Therefore, respondent recall is used to define users and non-users.

Two groups, the Single Recall (T2) and the Bounded Recall (T3) groups required the respondents to recall the number of videos they had hired in the last four weeks. This recall data was used to identify users and non-users. User were defined as those respondents who recalled hiring at least one video while non-users were defined as those respondent who recalled hiring no videos in the previous four weeks.

Table 16 shows the results of performing the alternative method of predicting purchase levels using the respondent's recall to classify users and non-users.

As only two of the treatments obtained recall data prior to asking the prediction questions (the Single Recall and the Bounded Recall groups), only the results from these two treatments can be analysed and reported.

Table 16. Alternative error calculation method using recall data to determine users and non-users

	n	Predicted for users <sup>1</sup>				Prop <sup>2</sup>	Predicted for market				Act <sup>3</sup>	Error			
		P(ANY)* N*P(N)	P(ANY)*N	N*P(N)	N		P(ANY)* N*P(N)	P(ANY)*N	N*P(N)	N		P(ANY)* N*P(N)	P(ANY)*N	N*P(N)	N
T2 Single Recall	40	1.82	2.74	2.39	3.75	.40	0.73	1.10	0.96	1.50	.69	0.04	0.41	0.27	0.81
T3 Bounded Recall	38	1.89	2.76	2.47	3.87	.38	0.72	1.05	0.94	1.47	.69	0.03	0.36	0.25	0.78

- Note
- 1: Users were defined as those people who recalled hired at least one video in the four weeks prior to the initial interview.
  - 2: Prop is the proportion of the sample that were users.
  - 3: Act is the actual number of videos hired by very respondent in the sample.

Firstly, it does appear that using the respondent's recall to classify users and non-users does effect the classification. When the computer records are used, the proportion of the sample that were users was .31 (see Table 16) of the sample, however, when the respondent's recall is used, .4 of respondents in the Single Recall group and .38 of respondents in the Bounded Recall group were classified as users. The increase in the proportion defined as users is not surprising as it would be expected that due to telescoping, some respondents who had not hired in the four weeks would telescope and report that they had hired leading to the increase in the reported number of users.

Despite the misclassification due to using the respondent's recall, this did not greatly affect the errors in predictions. The errors in prediction for the Bounded Recall group were slightly less than for the Single Recall group, but the errors for both groups were again very small ranging from 0.03 to 0.81. In this analysis though, the  $P(\text{ANY}) * N * P(N)$  method of calculating the predicted purchase level was the most accurate method for both treatments, followed by the  $N * P(N)$  method then the  $P(\text{ANY}) * N$  and least accurate was using just the number most likely to hire ( $N$ ).

Despite the concerns that using the respondent's recall to classify users and non-users may lead to misclassification which would in turn lead to increased inaccuracies in purchase level predictions, the results shown in Table 16 show that the concerns were unjustified. Given this encouraging result, there is evidence to justify pursuing this type of calculations further. Analysing the errors of users and non-users for a variety of product types needs to be undertaken to ensure that the results found here are not limited to the product type (video tape hire).

#### 6.6.4 Summary

This section examined the effect that usage level has on predictive accuracy. It was found that while usage did not appear to have any affect on purchase rate estimates, there were significant differences in the purchase level errors between those who had used their video card, and those who had not, with users having significantly smaller purchase level errors. This finding lead to an alternative method of calculating the purchase level. The new method

was based upon the assumption that non-user have a purchase probability of zero, and thus the purchase level of the market must be equal to the purchase level of users. The new method identified users, and then calculated the purchase level for users. Using the computer records to identify those who had used their video card over the four week prediction period, it was found that the new calculation method greatly improved the predictive accuracy. The data was reanalysed using the respondent's recall to identify users and non-user, users were defined as those who recalled hiring at least one video in the four weeks prior to day on which the purchase probability data was collected. It was gain found that the new method of calculating purchase levels greatly reduced the error in prediction.

## **7. DISCUSSION**

### **7.1 Introduction**

The focus of this section is to evaluate the findings of the Results section in light of the theoretical grounds upon which the hypotheses set out in section three were based. The implication and recommendations for future research are also discussed. Each of the hypotheses are discussed in turn.

### **7.2 Effect of Recall Assisted Predictions**

It was hypothesised that asking a bounded recall question prior to the prediction question would result in smaller errors in predictions than asking a single recall question prior to the prediction questions, which in turn would have smaller errors in predictions than just asking the prediction questions without any recall questions.

The rationale behind these hypotheses was that as brand choice is related to current brand usage, and recall is implicit in Beach, Barnes and Christensen-Szalanski (1986) model of subjective probability, recalling current usage would act as a guide to future purchasing. As bounded recall reduces telescoping, it would be expected that the predictions from the group who were asked bounded recall would have more accurate predictions.

The results from tests of Hypotheses 1 and 2 were such that they were rejected. This would initially suggest that the rationale used to justify the hypotheses was incorrect. There would also be reason to question the validity of applying the model of subjective probability to this situation, and that previous purchase behaviour may not be a good guide for predicting future behaviour.

However, the results from the test of Hypothesis 3 prevents the unequivocal acceptance of the aforementioned criticisms. By accepting Hypothesis 3, the rationale used to develop the hypothesis gains some validity. To further validate the rationale, an alternative explanation of why Hypotheses 1 and 2 were rejected is that the behaviour in question is one for which recall is inaccurate. The rationale for the hypotheses is built around accurate recall. If the performance of the behaviour is difficult to recall, then it must be expected that the predictions would be less accurate than for a behaviour that is accurately recalled. One reason why the control group was more accurate is that the respondent was not basing their prediction on inaccurate recall. This highlights a limitation of the method tested in Hypotheses 1 and 2, that is, if a behaviour is one which tends to inaccurately recalled, obtaining recall information from the respondent prior to making prediction may make the resulting predictions even less accurate.

As mentioned earlier, the product itself may be one that is difficult to recall, therefore, it may be that if an alternative product was used, the results may be such that the hypothesis may be accepted. In this study it was found that usage levels were low. Only around a third of the respondents had hired at least one video in the four week period. Parfitt (1967) found that the recall of infrequently purchased products was less accurate than that of frequently purchased products. Furthermore, U (1991) found that for products with low usage levels, predictions were less accurate than products with higher usage levels. Thus a test of the hypotheses on more frequently purchased products may give different results to those found here.

As well as some direction for future research (as just discussed) the implication of these results to research is that at this stage of its development, obtaining recall data immediately prior to purchase probabilities being collected is unlikely to improve predictions. It also suggests that when collecting purchase probability data, the researcher must consider question order effects carefully. Recall questions placed before the probability question in the questionnaire may affect the prediction.

### 7.3 Effect of the Bounded Prediction Procedure

It was hypothesised that using a bounded prediction question technique would result in smaller errors in predictions than predictions using only a single prediction question. The rationale behind the hypothesis was that one of the factors in Beach, Barnes and Christensen-Szalanski (1986) contingency model of subjective probability judgements is the motivation to provide accurate answers. It has been claimed that motivation to provide accurate answers is the reason why bounding recall questions is effective at reducing over-reporting in recall questions. Thus bounding the prediction question may increase the respondent's motivation to accurate answers to the prediction questions.

The results showed that bounded the prediction question was effective at reducing predictive errors.

This result suggests that the problem of over-estimation in predictions and over-reporting in retrospective questioning are the manifestations of the same affect. That is, telescoping is the name given to a motivational effect found in recall that leads to over-reporting, but the motivational effect that leads to over-reporting in recall is the same motivational effect that leads to over-estimation in predictions.

The implication of this is that if telescoping is a motivational effect, then it would follow if there are other situations that are affected by the same motivational effect, then applying a bounding technique may be effective at reducing over-statements.

Identification of situations where over-statement is commonly found, and then testing to establish if techniques such as single interview bounded recall can be adapted to the situation and reduce the level of over-statement would have two benefits. Firstly, it would provide additional evidence to support the existence of a general motivational effect that leads to over-statement. Secondly, by virtue of the method needed in establishing that it is the motivational effect that is causing the over-statement, that is, using techniques such as bounded interviewing adapted from other situations where over-statement had been identified and reduced, a method of improving the accuracy of the data gathered in that situation is

established. For example, if it was found that socially desirable topics such as the amount of money donated to charities tended to be over-stated, and an adaptation of the single interview bounded recall technique was found to be effective at reducing the level of over-reporting, then this result would provide further evidence to suggest that there is a common cause which leads to over-statement, and it would also identify a method for reducing the overstatement.

#### **7.4 Effect of Purchase Level Estimate Calculations**

It was hypothesised that reduced prediction questioning procedures would not affect the error in purchase level predictions. The rationale behind the hypothesis was the Brennan, Esslemont and Hini (1995) were able to reduce the questioning needed to calculate purchase level estimates, without adversely affecting the accuracy of the prediction. In doing so, it showed that the purchase level calculation does not need to be logically complete to accurately predict. Thus it was hypothesised that the questioning required could be reduced further without being logically complete, but still provide accurate predictions.

The results showed that any reduction in the questioning significantly reduced the accuracy of the prediction.

The implication of this result is that the three questions, that is, the chances of purchasing any, the number most likely to be purchased and the chances of purchasing exactly that number, must be asked and used in the purchase level calculation, otherwise significant reductions in the predictive accuracy are likely.

#### **7.5 Effect of using Respondent Recall to Calculated Error in Predictions**

It was hypothesised that the error in predictions calculated using recall as a surrogate for actual behaviour would not be different from the error in predictions calculated using actual

behavioural data. The rationale behind the hypothesised was that as the recall data used to test the accuracy of predictions is collected after the interview where the predictions were made, the initial interview would act as a landmark event and therefore the recall would not be so greatly affected by recall biases such as telescoping. With the reduction in recall biases, it would be expected that there would not be significant differences between using recall as a surrogate for actual data in the error calculation, and the error calculated using actual data.

When the errors in predictions using the formula (Predicted - recall) was compared to (Predicted - computer records) it was found that using recall understated the actual error, in some cases, the level of under-statement was statistically significant.

In discussing this result, there are two issues that need to be considered. Firstly, there is the absolute accuracy of the prediction, that is, what is the level of under-statement when using recall data as a surrogate for actual data. The second issue is the relative accuracy of the treatments. In other words, although the absolute accuracy may be under-stated, is the under-statement consistent across experimental treatments.

The results clearly shows that the traditional method of collecting purchase probability data and then reinterviewing the respondent at the end of the prediction period to collect retrospective data which is then used to test the accuracy of the predictions is likely to under-report the actual error level. From these results, it would be likely that conclusions concerning the absolute accuracy of the predictions drawn would be misleading. That is, when calculating the error in purchase rate or purchase level predictions, if recall data was used as a surrogate for actual behaviour, it is likely that any conclusions drawn concerning the absolute accuracy of the predictions would tended to under-state the actual error.

For two of the control groups (B and C), using recall data in the error calculations resulted in errors in predictions that were significantly smaller than what the actual errors were. The control groups used the procedures that had initially been developed for the collection of purchase probabilities via telephone interviews (see Brennan, Hini and Esslemont 1994; Brennan, Esslemont and Hini 1995). As there was significant difference between the actual error in predictions and the calculated error in predictions using recall data, there would be

some reason to question the conclusions drawn from previous studies investigating the Verbal Purchase Probability scale regarding the absolute accuracy.

The second issue of consistency of under-statement across treatments can be addressed by examining the rank order of the size of the error of the treatments (see tables 10 and 11). Looking at the Control A, Single Recall and Bounded Recall groups, in terms of purchase rate errors, using the formula (Predicted - Computer Records) Control A had the smallest error, followed by the Bounded Recall and then the Single Recall group. When recall is used to in the formula (Predicted - Recall), Control A had the smallest error, followed by the Bounded Recall and then the Single Recall group. Thus, when looking at the relative accuracy, the under-statement in predictive error caused by recall error was consistent across the three treatments. Conclusions regarding the relative superiority of the three treatments would be the same irrespective of whether computer records or respondent recall was used. Looking at the treatments, Control B, Control C and Bounded Juster, in terms of purchase rate errors the Bounded Juster group had the smallest predictive error when calculated using either the computer records, or the respondent recall. However, using the formula (Predicted - Computer records), Control C was marginally more accurate than Control B, but when the formula (Predicted - Recall) was used to calculate predictive error, Control B was slightly more accurate than Control C. Therefore, irrespective of the formula used (that is, either (Predicted - Actual) or (Predicted - Recall)) it would be concluded that the Bounded Juster group produced smaller predictive errors than the control group.

The results for purchase level were similar to those for purchase rate. Out of the Control A, Single Recall and Bounded Recall groups, the errors in predictions for Control A were consistently smaller than the Single Recall and Bounded Recall group irrespective of whether the computer records or recall was used to calculate the predictive error (except for when P(A).N.P(N) was used to calculate the purchase level and respondent recall was used in the error calculation, in which case, the predictive accuracy of Control A was matched by the Bounded Recall group).

For Control B, Control C and the Bounded Juster groups, the Bounded Juster had the smallest errors in predictions, irrespective of whether the computer records or recall was used

in the error calculation. As was found with purchase rates, the relative accuracy of Control B and Control C alternated depending on whether the computer records or recall was used. Nonetheless, irrespective of whether the computer records or recall was used, it would be concluded that the Bounded Juster group produced more accurate predictions than the control groups.

These results have both practical and theoretical implications.

It was hypothesised that there would not be a significant difference between the calculated error in predictions because the initial interview where the purchase probability data is collected would act as a landmark event and thus provide clear boundaries to the respondent to aid their recall (Loftus and Marburger, 1983). The results show that this was not correct. In Loftus et al's study, the personal landmarks were events such as birthdays, which should be more memorable than an interview. Thus the implication is that a personal landmark event must be more significant to the respondent than being interviewed.

Practically, these results suggest that the method of calculating errors in prediction is understating the actual error level and as such questions the claims regarding the absolute accuracy of studies that have used this method (for example see Brennan, Hini and Esslemont 1994; Brennan, Esslemont and Hini 1995; U 1991). In light of these findings new methods of collecting data to test the accuracy the predictions needs to be used. Obviously using actual behavioural data would be ideal, but often this is neither practicable nor possible. Therefore, methods of improving respondent recall need to be used. Two techniques were examined in the Literature Review, single interview bounded recall and landmark events. As discussed previously, the initial interview was not sufficient to act as a landmark event. Therefore, single interview bounded recall techniques need to be tested as a method of collecting data to test the accuracy of predictions.

## **7.6 Effect of Recall**

It was hypothesised that respondents with accurate recall at the time of making a prediction would have less error in their predictions than those with inaccurate recall at the time of making the prediction. The rationale behind this hypothesis was that implicit in Beach et al's (1986) model of subjective probability was that recall is part of the judgement making process, and if recall was inaccurate, it would follow that the judgement would also be inaccurate.

The results were such that the hypothesis was accepted. This confirms the importance of recall in the subjective probability model.

This result gives a generalisable result, that is, there is evidence to suggest that purchases that can be accurately recalled can be accurately predicted. Thus if a research knows if a product is accurately recalled then it is likely that predictions would be more accurate than if the product is not accurately recalled. The obvious limitation of this is that the researcher must know if the product can be recalled accurately, thus the result would only be generalisable to products for which the accuracy of recall has been tested. However, Parfitt (1967) has shown that frequency of purchase affects recall, and that recall is more accurate for frequently purchased products. Frequency of purchasing can be easily estimated, and then used as an estimator of the likely accuracy of purchase recall for a product.

## **7.7 Effect of Usage**

It was hypothesised that errors in predictions for low users (non-users) would be greater than the errors in predictions for high users (users). The reason behind this was that Parfitt (1967) had found that recall tended to be less accurate for infrequently purchased items, it was then assumed that infrequent purchasing would indicate infrequent usage. Thus, as recall would tend to be less accurate for infrequently used products (by assuming that infrequent purchasing indicates infrequent usage), through H7, (which states that inaccurate recall would

lead to inaccurate predictions), infrequent usage would lead to inaccurate predictions.

The results were such that the hypothesis was accepted. When the data was reanalysed calculating the purchase level based on users, it was found that the errors in predictions were greatly reduced.

The new method of calculating purchase level estimates appears to have great potential. It has certainly shown itself to be effective at reducing predictive errors in this study where usage levels were low. However, the success of the alternative technique raises more questions and thus has many implications for future research.

Firstly, can the alternative method be used for products where usage levels are higher. Predictions using the Juster Scale and Verbal Purchase Probability scale have been much more accurate when usage was higher (for example see U, 1991). If the alternative technique was used on products with high usage levels, would it still be more accurate? If it was more accurate, could it then be used as the default method of calculating purchase level estimates, that is, by collecting more evidence comparing the alternative method with the traditional method for a variety of brands, product classes and product types (i.e. high usage levels and so on), does the alternative method consistently produce more accurate purchase level estimates. Respondent recall was used to define users and non-users. Using respondent recall requires previous and future purchasing to be very similar. It is quite possible that using recall to define users and non-user would not work for different products. Therefore, is there a better method of identifying users and non-users?

The practical implications of these finding are that if a research is trying to predict a the purchase level of a product that has low levels of usage, then using the new method of calculating purchase level estimates is likely to greatly improve the accuracy of the predictions. Thus in designing the questionnaire, recall questions need to be included.

## 8. CONCLUSIONS

Whilst asking respondents to recall their previous behaviour prior to making the prediction failed to improve the accuracy of predictions, asking the prediction question in a bounded fashion, that is, asking the respondent to predict for a long time period and then the time period of interest was found to improve predictive accuracy. Thus in order to improve predictions it is recommended that a bounded prediction technique is used, and not a recall assisted technique. Furthermore, the presence of recall questions prior to prediction questions in a questionnaire is cautioned against.

It was found that any reduction in the prediction questioning from that which was tested by Brennan, Esslemont and Hini (1995) (that is, the chances of buy any, the number mostly to be bought, and the chances of buying exactly that number) resulted in a significant reduction in predictive accuracy. Thus all three questions need to be used to estimate purchase levels, otherwise predictive accuracy is likely to reduce significantly.

Using respondent recall to test the accuracy of predictions was found to understate the actual error. However, although the error calculated using respondent recall is understated, the effect is consistent across treatments, and thus claims regarding the relative performance of experimental treatments investigating the accuracy of the Juster Scale or Verbal Purchase Probability scale are unlikely to be affected. Thus it is recommended that an alternative to the respondent recall that has previously been used to test the accuracy of predictions be used. The alternative may still be respondent recall, but collected in a way that will reduce retrospective questioning biases. The alternative method of collecting the recall data suggested is to use bounded recall data.

It was found that those with accurate recall at the time of making predictions made more accurate predictions. This suggests that by knowing if recall of purchasing a product is accurate, then it can be expected that predictions would tend to be accurate.

Finally, it was found that non-users had significantly more error in purchase level predictions than those who had used their video card to hire a video. This led to the development of a new method of calculating purchase level estimates based on assuming non-users have a zero purchase probability and then calculating the purchase level based on that of users. This new method was found to greatly reduce the predictive error. Thus it was concluded that, particularly for products that have low usage levels, the alternative method of calculating purchase level estimates is likely to provide more accurate predictions than using the traditional method.

## REFERENCES

- Assendelft, E.W. & Lewis, T. (1991). *An experimental approach to estimating the demand curve for raw wool using the Juster Scale*. Work in Progress Paper presented to the Marketing Educators Conference, Massey University.
- Baddeley, A. (1979). The limitations of human memory: Implications for the design of retrospective surveys. In Moss, M. & Goldstein, G. (Eds.), *The recall method in social surveys*. London: University of London Institute of Education.
- Baldwin, G., Gendall, P. & Hoek, J. (1995). Secondary students' perceptions of debt and its effects on tertiary enrolment intentions. *Marketing Bulletin*, **6**, 1-11.
- Beach, L., Barnes, V. & Christensen-Szalanski, J. (1986). Beyond heuristics and biases: A contingency model of judgmental forecasting. *Journal of Forecasting*, **5**, 143-157.
- Beach, L.R. & Braun, G.P. (1994). Laboratory studies of subjective probability: A status report. In Wright, G. & Ayton, P. (Eds.), *Subjective probability*. Chichester: John Wiley & Sons.
- Belson, W.A. (1985). Respondent understanding of survey questions. *Polls*, **3**(4), 1-13.
- Bird, M. & Ehrenberg, A.S.C. (1966). Intentions-to-buy and claimed brand usage. *Operational Research Quarterly*, **17**(1), 27-46.
- Bradburn, N., Rips, L. & Shevell, S. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science*, **236**, 157-161.
- Brennan, M. (1995). Constructing Demand Curves from Purchase Probability Data: An application of the Juster Scale. *Marketing Bulletin*, **6**, 51-58.

- Brennan, M. & Esslemont, D. (1994). The accuracy of the Juster Scale for predicting purchase rates of branded, fast-moving consumer goods. *Marketing Bulletin*, **5**, 47-53.
- Brennan, M., Esslemont, D. & Hini, D. (1995). Obtaining purchase predictions via telephone interviews. *Journal of the Market Research Society*, **37**(3), 241-250.
- Brennan, M., Hini, D. & Esslemont, D. (1994). Obtaining purchase probability data via telephone surveys: A preliminary test of two techniques. *Marketing Bulletin*, **5**, 64-70.
- Byrnes, J.C. (1964). Consumer intentions to buy. *Journal of Advertising Research*, **4**(3), 49-51.
- Cassells, A. (1991). *Remembering and forgetting*. United Kingdom: The British Psychological Society.
- Charlett, D., Garland, R. & Marr, N. (1995). How damaging is negative word of mouth?. *Marketing Bulletin*, **6**, 42-50.
- Clark, J. & Tiffit, L. (1966). Polygraph and interview validation of self-reported deviant behavior. *American Sociological Review*, **31**, 516-523.
- Cook, W. (1987). Telescoping and memory's other tricks. *Journal of Advertising Research*, **27**, RC5-RC8.
- Day, D., Gan, B., Gendall, P. & Esslemont, D. (1991). Predicting Purchase Behaviour. *Marketing Bulletin*, **2**, 18-30.
- Ferber, R. & Piskie, R.A. (1965). Subjective probabilities and buying intentions. *The Review of Economic and Statistics*, **47**(8), 322-325.

- Fisher, R.J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, **20**, 303-315.
- Gendall, P., Esslemont, E. & Day, D. (1991). A comparison of two versions of the Juster Scale using self-completion questionnaire. *Journal of the Market Research Society*, **33**(3), 257-263.
- Gruber, A. (1970). Purchase intent and purchase probability. *Journal of Advertising Research*, **10**(1), 23-27.
- Hamilton-Gibbs, D. (1986). *Predicting the demand for frequently purchased items*. Unpublished Student Report, Massey University, Palmerston North, New Zealand.
- Hawkins, D.I. & Coney, K.A. (1981). Uninformed response error in survey research. *Journal of Marketing Research*, **18**, 370-374.
- Hornik, J., Cherian, J. & Zakay, D. (1994). The influence of prototypic values on the validity of studies using time estimates. *Journal of the Market Research Society*, **36**(2), 145-147.
- Huttenlocher, J., Hedges, L. & Bradburn, N. (1990). Reports of elapsed time: Bounding and rounding processes in estimation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16**(2), 196-213.
- Huttenlocher, J., Hedges, L. & Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review*, **95**(4), 471-484.
- Johnston, V.E. & Kaplan, S.E. (1991). Experimental evidence on the effects of accountability on auditor judgements. *Auditing*, **10**, 96-107.
- Juster, F. (1966). *Consumer Buying Intentions and Purchase Probabilities*. Occasional Paper 99, National Bureau of Economic Research, Columbia University Press.

- Kemsley, W.F.F. & Nicholson, J.L. (1960). Some experiments in methods of conducting family expenditure surveys. *Journal of the Royal Statistical Society*, **123**, 307-328.
- Kuehn, A. (1962). Consumer brand choice as a learning process. *Journal of Advertising Research*, **2**(4), 10-17.
- Loftus, E., Klinger, M., Smith, K. & Fiedler, J. (1990). A tale of two questions: Benefits of asking more than one question. *Public Opinion Quarterly*, **54**, 330-345.
- Loftus, E. & Marburger, W. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition*, **11**(2), 114-120.
- Mahalanobis, P. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, **109**, 326-370.
- Neter, J. & Waksberg, J. (1964). A study of responses errors in expenditures data from household interviews. *Journal of the American Statistical Association*, **59**, 18-55.
- Parfitt, J.H. (1967). A comparison of purchase recall with diary panel records. *Journal of Advertising Research*, **7**, 16-31.
- Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording and context*. New York: Academic Press.
- Scott Armstrong, J., Denniston, Jr. W. & Gordon, M. (1975). The use of the decomposition principle in making judgements. *Organisational Behaviour and Human Performance*, **14**, 257-263.
- Schwarz, N., Strack, F. & Mai, H. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, **55**, 3-23.

- Seymour, P., Brennan, M. & Esslemont, D. (1994). Predicting purchase quantities: Further investigation of the Juster Scale. *Marketing Bulletin*, **5**, 21-36.
- SPSS Inc. (1993). *SPSS Base system syntax reference guide release 6.0*. Illinois: SPSS Inc.
- Sudman, S. (1964). On the accuracy of recording of consumer panels: I. *Journal of Marketing Research*, **1**, 14-20.
- Sudman, S. & Bradburn, N. (1973). Effects of time and memory factors on response in surveys. *Journal of the American Statistical Association*, **68**(344), 805-815.
- Sudman, S. & Bradburn, N. (1974). *Response effects in surveys. A review and synthesis*. Chicago: Adline Publishing Company.
- Sudman, S. & Ferber, R. (1971). Experiments in obtaining consumer expenditures by diary methods. *Journal of the American Statistical Association*, **66**(336), 725-735.
- Sudman, S. & Ferber, R. (1974). A comparison of alternative procedures for collecting consumer expenditure data for frequently purchased products. *Journal of Marketing Research*, **11**, 128-35.
- Sudman, S., Finn, A. & Lannom, L. (1984). The use of bounded recall procedures in single interviews. *Public Opinion Quarterly*, **48**, 520-524.
- Thompson, C.P., Skowronski, J.J. & John Lee, D. (1988). Telescoping in dating naturally occurring events. *Memory & Cognition*, **16**(5), 461-468.
- U, C. (1991). *Predicting the demand for branded products using the Juster Scale*. Unpublished Student Report, Massey University, Palmerston North, New Zealand.
- Warrington, E.K. & Sanders, H.I. (1971). The fate of old memories. *Quarterly Journal of Experimental Psychology*, **23**, 432-442.

Wind, Y., and Lerner, D. (1979). On the measurement of purchase data: Surveys versus purchase diaries. *Journal of Marketing Research*, **16**, 39-47.

Wright, M. & Esslemont, D. (1994). The logical limitations of target marketing. *Marketing Bulletin*, **5**, 13-20.

## **BIBLIOGRAPHY**

Aaker, D.A. & Day, G.S. (1990). *Marketing research* (4th ed.). Singapore: John Wiley & Sons.

Emory, C. & Cooper, D. (1991). *Business research methods* (4th ed.). Boston: Irwin.

Wright, G. & Ayton, P. (Eds.). (1994). *Subjective probability*. Chichester: John Wiley & Sons.

## APPENDICES

## QUESTIONNAIRES

## Control A (T1) Questionnaires

Good evening. My name is <FULL NAME>. I'm from Massey University and I'm doing a short survey about videos and would like to ask you a few questions. This interview will only take about five minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW     16-

Before we start, I'd just like to assure you that everything you say is completely confidential.

Is there a video player in this household?

- Yes ..... 1  20
- No ..... 2

Have you personally ever rented a video tape?

- Yes ..... 1  21
- No ..... 2

Which of the following Palmerston North video stores are you a member of?

**READ LIST, CIRCLE ALL THAT APPLY**

- Selwyn Plaza Video ..... 1  22
- Video Concepts Albert Street ... 1  23
- Video Concepts Featherston St . 1  24
- Movie World ..... 1  25
- Other ..... 1  26

**IF SELWYN PLAZA NOT CIRCLED THEN:** I'm sorry, but the rest of the questions relate to Selwyn Plaza Video, and so I don't need to ask you any more questions. Thank you for your time.

Selwyn Plaza Video is thinking about providing access to the internet. Customers would be able to use this services for \$5 per 15 minutes. We would like to know what the chances are of you using this service at Selwyn Plaza Video if it was available.

I would like you to answer on a scale of "zero" to "10". If you are certain, or practically certain that you will use this service, then you would choose the answer "10". If you think there is no chance, or almost no chance of using, the best answer would be "zero". If you are uncertain about the prospects, choose another answer as close to "0" or "10" as you think it should be. You can think of the numbers as chances out of "10". For example, 3 would mean 3 chances in 10 that you would use the service, while a 7 would mean 7 chances in 10 that you would use the service, and so on

So, taking everything into account, what would be the chances of you personally using the internet service at Selwyn Plaza Video if it was available?

39-4

**IF THE PROBABILITY = 0 , THEN GO TO Q6**

How often do you think you would use this service in a four week period, if it was available?

41-4

I'm now going to ask you some questions about the chances of you hiring videos. I'd like you to use the same 0 to 10 scale again.

Now, taking everything into account, what are the chances of you personally hiring at least one video from Selwyn Plaza Video, during the next four weeks, that is between now and <DAY AND DATE> **SEE CHART C?**

50-4

**IF THE PROBABILITY = 0 , THEN GO TO Q9**

If you did hire any, how many videos are you personally most likely to hire from Selwyn Plaza Video in the next four weeks?

52-4

Using the scale from 0 to 10, what are the chances that you personally would hire exactly <MOST LIKELY NUMBER> videos from Selwyn Plaza Video in the next four weeks.

55-4

Now, using the 0 to 10 scale again, what are the chances that someone else will use your Selwyn Plaza Video card to hire at least one video during the next four weeks, that is between now and <DAY AND DATE> **SEE CHART C?**

64

**IF THE PROBABILITY = 0 , THEN GO TO Q12.**

If they did hire any, how many videos are other people most likely to hire using your Selwyn Plaza Video card in the next four weeks?

66

Using the scale from 0 to 10, what are the chances that other people would hire exactly <MOST LIKELY NUMBER> videos with your Selwyn Plaza Video card in the next four weeks?

69

2. Thinking only about today, have you personally hired any videos from Selwyn Plaza Video?

**IF YES** . . . . How many have you personally hired just today?

7

**IF NO** . . . . **RECORD 0 IN THE BOX & GO TO Q13**

3. Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?

**IF YES** . . . . How many have other people hired just today?

73

**IF NO** . . . . **RECORD 0 IN THE BOX & GO TO Q14**

4. If there is anything else that we need to know, can we call you back?

Yes . . . . . 1

No . . . . . 2

75

That's all I need to ask. Thank you for helping us with this survey. Bye.

Audited \_\_\_\_\_

RECORD THE EXACT TIME NOW

76-7

Callback details \_\_\_\_\_

ID

Good evening. My name is <FULL NAME>. Four weeks ago, we interviewed you about videos, and you said we could call you back to ask a few more questions. This interview will only take about two minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW

**IF INTERVIEWED ON TIME, ASK ONLY Q1 - Q6.  
IF INTERVIEWED LATE, START FROM Q7.**

Since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE **CHART B**, how many videos have you personally hired from Selwyn Plaza Video?

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?  
Yes . . . . . 1, **GO TO Q3**   
No . . . . . 2 **GO TO Q4**

Since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE **CHART B**, how many videos would you say other people have hired using your Selwyn Plaza Video card?

When was the last time anybody including yourself hired a video using your Selwyn Plaza Video card?  
\_\_\_\_\_ Days Ago      \_\_\_\_\_ Weeks Ago      \_\_\_\_\_ Months Ago  
\_\_\_\_\_ Date              \_\_\_\_\_ Yesterday      \_\_\_\_\_ Others

Thinking only about today, have you personally hired any videos from Selwyn Plaza Video?  
**IF YES** . . . How many have you personally hired just today?    
**IF NO** . . . . **RECORD 0 IN THE BOX & GO TO Q6**

Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?  
**IF YES** . . . How many have other people hired just today?    
**IF NO** . . . . **RECORD 0 IN THE BOX**

# IF INTERVIEWED LATE

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos have you personally hired from Selwyn Plaza Video?

Does this number include any videos you personally have hired from Selwyn Plaza Video....

IF 1 DAY LATE ... just today?

IF >1 DAY LATE .. since last <ACTUAL DAY>?

Yes ..... 1 GO TO Q9

No ..... 2 GO TO Q10

How many videos have you personally hired from Selwyn Plaza Video....

IF 1 DAY LATE ... just today?

IF >1 DAY LATE .. since last <ACTUAL DAY>?

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?

Yes ..... 1 GO TO Q11

No ..... 2 GO TO Q14

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos would you say other people have hired using your Selwyn Plaza Video card?

Does this number include any videos anyone else has hired from Selwyn Plaza Video....

IF 1 DAY LATE ... just today?

IF >1 DAY LATE .. since last <ACTUAL DAY>?

Yes ..... 1 GO TO Q13

No ..... 2 GO TO Q14

How many videos have other people hired using your Selwyn Plaza Video card....

IF 1 DAY LATE ... just today?

IF >1 DAY LATE .. since last <ACTUAL DAY>?

When was the last time anybody including yourself hired a video using your Selwyn Plaza Video card?

\_\_\_\_\_ Days Ago

\_\_\_\_\_ Weeks Ago

\_\_\_\_\_ Months Ago

\_\_\_\_\_ Date

\_\_\_\_\_ Yesterday

\_\_\_\_\_ Others

**END** That's all I need to ask. Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW

Audited \_\_\_\_\_

## Single Recall (T2) Questionnaires

Version	<input type="text" value="1"/>	1			
Phase	<input type="text" value="2"/>	2			
Outcome	<input type="text"/>	3			
No. of Calls	<input type="text"/>	<input type="text"/>	4-5		
Day	<input type="text"/>	<input type="text"/>	6-7		
Month	<input type="text"/>	<input type="text"/>	8-9		
Interviewer ID	<input type="text"/>	<input type="text"/>	10-11		
ID	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	12-15

Callback details \_\_\_\_\_

Good evening. My name is <FULL NAME>. I'm from Massey University and I'm doing a short survey about videos and would like to ask you a few questions. This interview will only take about five minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW  16-1

Before we start, I'd just like to assure you that everything you say is completely confidential.

Is there a video player in your household?

Yes ..... 1  20  
 No ..... 2

Have you personally ever rented a video tape?

Yes ..... 1  21  
 No ..... 2

Which of the following Palmerston North video stores are you a member of?

**READ LIST, CIRCLE ALL THAT APPLY**

Selwyn Plaza Video ..... 1  22  
 Video Concepts Albert Street . . . 1  23  
 Video Concepts Featherston St . . 1  24  
 Movie World ..... 1  25  
 Other ..... 1  26

**IF SELWYN PLAZA NOT CIRCLED THEN:** I'm sorry, but the rest of the questions relate to Selwyn Plaza Video, and so I don't need to ask you any more questions. Thank you for your time.

Thinking about the past four weeks, that is, since <DAY AND DATE> **SEE CHART B**, how many videos have you personally hired from Selwyn Plaza Video?  29-3

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?  35

Yes ..... 1 **GO TO Q6**  
 No ..... 2 **GO TO Q7**

Thinking about the past four weeks, that is, since <DAY AND DATE> **SEE CHART B**, how many videos would you say other people have hired using your Selwyn Plaza Video card?  36-3

Selwyn Plaza Video is thinking about providing access to the internet. Customers would be able to use this services for \$5 per 15 minutes. We would like to know what the chances are of you using this service at Selwyn Plaza Video if it was available

I would like you to answer on a scale of "zero" to "10". If you are certain, or practically certain that you will use this service, then you would choose the answer "10". If you

think there is no chance, or almost no chance of using, the best answer would be "zero". If you are uncertain about the prospects, choose another answer as close to "0" or "10" as you think it should be. You can think of the numbers as chances out of "10". For example, 3 would mean 3 chances in 10 that you would use the service, while a 7 would mean 7 chances in 10 that you would use the service, and so on.

So, taking everything into account, what would be the chances of you personally using the internet service at Selwyn Plaza Video if it was available?

39-

**IF THE PROBABILITY = 0 , THEN GO TO Q9.**

How often do you think you would use this service in a four week period, if it was available?

41-

I'm now going to ask you some questions about the chances of you hiring videos. I'd like you to use the same 0 to 10 scale again.

Now, taking everything into account, what are the chances of you personally hiring at least one video from Selwyn Plaza Video, during the next four weeks, that is between now and <DAY AND DATE>SEE CHART C?

50-

**IF THE PROBABILITY = 0 , THEN GO TO Q12.**

If you did hire any, how many videos are you personally most likely to hire from Selwyn Plaza Video in the next four weeks?

52-

Using the scale from 0 to 10, what are the chances that you personally would hire exactly <MOST LIKELY NUMBER> videos from Selwyn Plaza Video in the next four weeks.

55-

Now, using the 0 to 10 scale again, what are the chances that someone else will use your Selwyn Plaza Video card to hire at least one video during the next four weeks, that is between now and <DAY AND DATE> SEE CHART C?

64-

**IF THE PROBABILITY = 0 , THEN GO TO Q15.**

If they did hire any, how many videos are other people most likely to hire using your Selwyn Plaza Video card in the next four weeks?

66-

Using the scale from 0 to 10, what are the chances that other people would hire exactly <MOST LIKELY NUMBER> videos with your Selwyn Plaza Video card in the next four weeks?

69-

Thinking only about today, have you personally hired any videos from Selwyn Plaza Video?

**IF YES** . . . How many have you personally hired just today?

71-

**IF NO** . . . . **RECORD 0 IN THE BOX & GO TO Q16**

Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?

**IF YES** . . . How many have other people hired just today?

73-

**IF NO** . . . . **RECORD 0 IN THE BOX & GO TO Q17**

If there is anything else that we need to know, can we call you back?

Yes . . . . . 1  
No . . . . . 2

75

That's all I need to ask. Thank you for helping us with this survey. Bye

RECORD THE EXACT TIME NOW

76-

Audited \_\_\_\_\_

Callback details \_\_\_\_\_

ID

Good evening. My name is <FULL NAME>. Four weeks ago, we interviewed you about videos, and you said we could call you back to ask a few more questions. This interview will only take about two minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW

**IF INTERVIEWED ON TIME, ASK ONLY Q1 - Q6.  
IF INTERVIEWED LATE, START FROM Q7.**

1. Since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE **CHART B**, how many videos have you personally hired from Selwyn Plaza Video?

2. Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?  
Yes ..... 1 GO TO Q3   
No ..... 2 GO TO Q4

3. Since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE **CHART B**, how many videos would you say other people have hired using your Selwyn Plaza Video card?

4. When was the last time anybody including yourself hired a video using your Selwyn Plaza Video card?  
\_\_\_\_\_ Days Ago      \_\_\_\_\_ Weeks Ago      \_\_\_\_\_ Months Ago  
\_\_\_\_\_ Date              \_\_\_\_\_ Yesterday      \_\_\_\_\_ Others

5. Thinking only about today, have you personally hired any videos from Selwyn Plaza Video?  
**IF YES** ... How many have you personally hired just today?    
**IF NO** ... RECORD 0 IN THE BOX & GO TO Q6

6. Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?  
**IF YES** ... How many have other people hired just today?    
**IF NO** ... RECORD 0 IN THE BOX

# IF INTERVIEWED LATE

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos have you personally hired from Selwyn Plaza Video?

Does this number include any videos you personally have hired from Selwyn Plaza Video....  
IF 1 DAY LATE ... just today?

IF >1 DAY LATE ... since last <ACTUAL DAY>?

Yes ..... 1 GO TO Q9

No ..... 2 GO TO Q10

How many videos have you personally hired from Selwyn Plaza Video....

IF 1 DAY LATE ... just today?

IF >1 DAY LATE ... since last <ACTUAL DAY>?

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?

Yes ..... 1 GO TO Q11

No ..... 2 GO TO Q14

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos would you say other people have hired using your Selwyn Plaza Video card?

Does this number include any videos anyone else has hired from Selwyn Plaza Video....  
IF 1 DAY LATE ... just today?

IF >1 DAY LATE ... since last <ACTUAL DAY>?

Yes ..... 1 GO TO Q13

No ..... 2 GO TO Q14

How many videos have other people hired using your Selwyn Plaza Video card....  
IF 1 DAY LATE ... just today?

IF >1 DAY LATE ... since last <ACTUAL DAY>?

When was the last time anybody including yourself hired a video using your Selwyn Plaza Video card?

\_\_\_\_\_ Days Ago

\_\_\_\_\_ Weeks Ago

\_\_\_\_\_ Months Ago

\_\_\_\_\_ Date

\_\_\_\_\_ Yesterday

\_\_\_\_\_ Others

**END** That's all I need to ask. Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW

Audited \_\_\_\_\_

## Bounded Recall (T3) Questionnaires

Version	<input type="text" value="3"/>			
Phase	<input type="text" value="2"/>			
Outcome	<input type="text"/>			
No. of Calls	<input type="text"/>	<input type="text"/>	4	
Day	<input type="text"/>	<input type="text"/>	6	
Month	<input type="text"/>	<input type="text"/>	8	
Interviewer ID	<input type="text"/>	<input type="text"/>	10	
ID	<input type="text"/>	<input type="text"/>	<input type="text"/>	12

Callback details \_\_\_\_\_

Good evening. My name is <FULL NAME>. I'm from Massey University and I'm doing a short survey about videos and would like to ask you a few questions. This interview will only take about five minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW  16

Before we start, I'd just like to assure you that everything you say is completely confidential.

Is there a video player in your household?

- Yes ..... 1
- No ..... 2

20

Have you personally ever rented a video tape?

- Yes ..... 1
- No ..... 2

21

Which of the following Palmerston North video stores are you a member of?

**READ LIST, CIRCLE ALL THAT APPLY**

- Selwyn Plaza Video ..... 1
- Video Concepts Albert Street ..... 1
- Video Concepts Featherston St. .... 1
- Movie World ..... 1
- Other ..... 1

22

23

24

25

26

**IF SELWYN PLAZA NOT CIRCLED THEN:** I'm sorry, but the rest of the questions relate to Selwyn Plaza Video, and so I don't need to ask you any more questions. Thank you for your time.

Thinking about the past eight weeks, that is, since <DAY AND DATE> **SEE CHART A**, how many videos have you personally hired from Selwyn Plaza Video?

27

**IF THE NUMBER = 0 , THEN GO TO Q6.**

Now, thinking about the past four weeks, that is, since <DAY AND DATE> **SEE CHART B**, how many videos have you personally hired from Selwyn Plaza Video?

28

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last eight weeks?

- Yes ..... 1 **GO TO Q7**
- No ..... 2 **GO TO Q9**

Thinking about the past eight weeks, that is, since <DAY AND DATE> **SEE CHART A**, how many videos would you say other people have hired using your Selwyn Plaza Video card?

**IF THE NUMBER = 0 , THEN GO TO Q9.**

Thinking about the past four weeks, that is, since <DAY AND DATE> **SEE CHART B**, how many videos would you say other people have hired using your Selwyn Plaza Video card?

Selwyn Plaza Video is thinking about providing access to the internet. Customers would be able to use this services for \$5 per 15 minutes. We would like to know what the chances are of you using this service at Selwyn Plaza Video if it was available.

I would like you to answer on a scale of "zero" to "10". If you are certain, or practically certain that you will use this service, then you would choose the answer "10". If you think there is no chance, or almost no chance of using, the best answer would be "zero". If you are uncertain about the prospects, choose another answer as close to "0" or "10" as you think it should be. You can think of the numbers as chances out of "10". For example, 3 would mean 3 chances in 10 that you would use the service, while a 7 would mean 7 chances in 10 that you would use the service, and so on.

So, taking everything into account, what would be the chances of you personally using the internet service at Selwyn Plaza Video if it was available?   39

**IF THE PROBABILITY = 0 , THEN GO TO Q11.**

How often do you think you would use this service in a four week period, if it was available?   41

I'm now going to ask you some questions about the chances of you hiring videos. I'd like you to use the same 0 to 10 scale again.

Now, taking everything into account, what are the chances of you personally hiring at least one video from Selwyn Plaza Video, during the next four weeks, that is between now and <DAY AND DATE> SEE CHART C?   50-5

**IF THE PROBABILITY = 0 , THEN GO TO Q14.**

If you did hire any, how many videos are you personally most likely to hire from Selwyn Plaza Video in the next four weeks?    51

Using the scale from 0 to 10, what are the chances that you personally would hire exactly <MOST LIKELY NUMBER> videos from Selwyn Plaza Video in the next four weeks.   51

Now, using the 0 to 10 scale again, what are the chances that someone else will use your Selwyn Plaza Video card to hire at least one video during the next four weeks, that is between now and <DAY AND DATE> SEE CHART C?   64

**IF THE PROBABILITY = 0 , THEN GO TO Q17.**

If they did hire any, how many videos are other people most likely to hire using your Selwyn Plaza Video card in the next four weeks?    61

Using the scale from 0 to 10, what are the chances that other people would hire exactly <MOST LIKELY NUMBER> videos with your Selwyn Plaza Video card in the next four weeks?   61

Thinking only about today, have you personally hired any videos from Selwyn Plaza Video?

**IF YES** . . . . How many have you personally hired just today?   7

**IF NO** . . . . RECORD 0 IN THE BOX & GO TO Q19

Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?

**IF YES** . . . . How many have other people hired just today?   7

**IF NO** . . . . RECORD 0 IN THE BOX & GO TO Q20

If there is anything else that we need to know, can we call you back?

Yes . . . . . 1  71  
No . . . . . 2

That's all I need to ask Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW     71

Audited \_\_\_\_\_

Good evening. My name is <FULL NAME>. Four weeks ago, we interviewed you about videos, and you said we could call you back to ask a few more questions. This interview will only take about two minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW

**IF INTERVIEWED ON TIME, ASK ONLY Q1 - Q6.  
IF INTERVIEWED LATE, START FROM Q7.**

Since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE **CHART B**, how many videos have you personally hired from Selwyn Plaza Video?

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?  
Yes ..... 1 **GO TO Q3**   
No ..... 2 **GO TO Q4**

Since we interviewed you four weeks ago, that is, since <DAY AND DATE> SEE **CHART B**, how many videos would you say other people have hired using your Selwyn Plaza Video card?

When was the last time anybody including yourself hired a video using your Selwyn Plaza Video card?  
\_\_\_\_\_ Days Ago      \_\_\_\_\_ Weeks Ago      \_\_\_\_\_ Months Ago  
\_\_\_\_\_ Date              \_\_\_\_\_ Yesterday      \_\_\_\_\_ Others

Thinking only about today, have you personally hired any videos from Selwyn Plaza Video?  
**IF YES** .... How many have you personally hired just today?    
**IF NO** .... **RECORD 0 IN THE BOX & GO TO Q6**

Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?  
**IF YES** .... How many have other people hired just today?    
**IF NO** .... **RECORD 0 IN THE BOX**

# IF INTERVIEWED LATE

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos have you personally hired from Selwyn Plaza Video?

Does this number include any videos you personally have hired from Selwyn Plaza Video....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

Yes . . . . . 1 GO TO Q9

No . . . . . 2 GO TO Q10

How many videos have you personally hired from Selwyn Plaza Video....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

0 Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?

Yes . . . . . 1 GO TO Q11

No . . . . . 2 GO TO Q14

1. Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos would you say other people have hired using your Selwyn Plaza Video card?

2. Does this number include any videos anyone else has hired from Selwyn Plaza Video....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

Yes . . . . . 1 GO TO Q13

No . . . . . 2 GO TO Q14

13. How many videos have other people hired using your Selwyn Plaza Video card....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

14. When was the last time anybody including yourself hired a video using your Selwyn Plaza Video card?

\_\_\_\_\_ Days Ago      \_\_\_\_\_ Weeks Ago      \_\_\_\_\_ Months Ago  
\_\_\_\_\_ Date              \_\_\_\_\_ Yesterday      \_\_\_\_\_ Others

**END** That's all I need to ask. Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW

Audited \_\_\_\_\_

## Control B (T4) Questionnaires

Version  1  
Phase  2  
Outcome  3  
No. of Calls   4-5  
Day   6-7  
Month   8-9  
Interviewer ID   10-

Callback details \_\_\_\_\_ ID     12-

Good evening. My name is <FULL NAME>. I'm doing a short survey about videos and would like to ask you a few questions. This interview will only take about five minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW     16-

Before we start, I'd just like to assure you that everything you say is completely confidential.

Is there a video player in this household?

Yes ..... 1  20  
No ..... 2

Have you personally ever rented a video tape?

Yes ..... 1  21  
No ..... 2

Which of the following Palmerston North video stores are you a member of?

**READ LIST, CIRCLE ALL THAT APPLY**

Selwyn Plaza Video ..... 1  22  
Video Concepts Albert Street . . . 1  23  
Video Concepts Featherston St . 1  24  
Movie World ..... 1  25  
Other ..... 1  26

**IF SELWYN PLAZA NOT CIRCLED THEN:** I'm sorry, but the rest of the questions relate to Selwyn Plaza Video, and so I don't need to ask you any more questions. Thank you for your time.

Selwyn Plaza Video is thinking about providing access to the internet. Customers would be able to use this services for \$5 per 15 minutes. We would like to know what the chances are of you using this service at Selwyn Plaza Video if it was available.

I would like you to answer on a scale of "zero" to "10". If you are certain, or practically certain that you will use this service, then you would choose the answer "10". If you think there is no chance, or almost no chance of using, the best answer would be "zero". If you are uncertain about the prospects, choose another answer as close to "0" or "10" as you think it should be. You can think of the numbers as chances out of "10". For example, 3 would mean 3 chances in 10 that you would use the service, while a 7 would mean 7 chances in 10 that you would use the service, and so on.

So, taking everything into account, what would be the chances of you personally using the internet service at Selwyn Plaza Video if it was available?

40-41

How often do you think you would use this service in a four week period, if it was available?

42-43

I'm now going to ask you some questions about the chances of you hiring videos. I'd like you to use the same 0 to 10 scale again.

Now, taking everything into account, what are the chances of you personally hiring at least one video from Selwyn Plaza Video, during the next four weeks, that is between now and <DAY AND DATE> SEE CHART C?

51-52

**IF THE PROBABILITY IS GREATER THAN ZERO, THEN ASK Q7, OTHERWISE GO TO Q9.**

If you did hire any, how many videos are you personally most likely to hire from Selwyn Plaza Video in the next four weeks?

53-55

Using the scale from 0 to 10, what are the chances that you personally would hire exactly <MOST LIKELY NUMBER> videos from Selwyn Plaza Video in the next four weeks.

56-57

Now, using the 0 to 10 scale again, what are the chances that someone else will use your Selwyn Plaza Video card to hire at least one video during the next four weeks, that is between now and <DAY AND DATE> SEE CHART C?

58-60

**IF THE PROBABILITY IS GREATER THAN ZERO, THEN ASK Q10, OTHERWISE GO TO Q12.**

If they did hire any, how many videos are other people most likely to hire using your Selwyn Plaza Video card in the next four weeks?

61-63

Using the scale from 0 to 10, what are the chances that other people would hire exactly <MOST LIKELY NUMBER> videos with your Selwyn Plaza Video card in the next four weeks?

64-66

How many videos have you personally hired so far today from Selwyn Plaza Video?

67-68

How many videos have other people hired so far today using your Selwyn Plaza Video card?

69-71

If there is anything else that we need to know, can we call you back?

- Yes . . . . . 1
- No . . . . . 2

72

That's all I need to ask. Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW

73

Audited \_\_\_\_\_

Version	<input type="text" value="4"/>	1					
Phase	<input type="text" value="2"/>	2					
Outcome	<input type="text"/>	3					
No. of Calls	<input type="text"/>	<input type="text"/>	4-5				
Actual Day	<input type="text"/>	<input type="text"/>	6-7	Month	<input type="text"/>	<input type="text"/>	8-9
Day	<input type="text"/>	<input type="text"/>	10-11	Month	<input type="text"/>	<input type="text"/>	12-13
Interviewer ID	<input type="text"/>	<input type="text"/>	14-15				
ID	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	16-17		

Callback details \_\_\_\_\_

Good evening. My name is <FULL NAME>. Four weeks ago, we interviewed you about videos, and you said we could call you back to ask a few more questions. This interview will only take about two minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW  20-21

**IF INTERVIEWED ON TIME, ASK FROM Q1 - Q7.**  
**IF INTERVIEWED LATE, START FROM Q8.**

Thinking about the past eight weeks, that is, since <DAY AND DATE> **SEE CHART A**, how many videos have you personally hired from Selwyn Plaza Video?  24-25

**IF THE NUMBER = 0, THEN GO TO Q3.**

And, since we interviewed you four weeks ago, that is, since <DAY AND DATE> **SEE CHART B**, how many videos have you personally hired from Selwyn Plaza Video?  26-27

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last eight weeks?  28  
Yes ..... 1 **GO TO Q4**  
No ..... 2 **GO TO Q6**

Thinking about the past eight weeks, that is since <DAY AND DATE> **SEE CHART A**, how many videos would you say other people have hired using your Selwyn Plaza Video card?  29-31

**IF THE NUMBER = 0, THEN GO TO Q6.**

Now, since we interviewed you four weeks ago, that is, since <DAY AND DATE> **SEE CHART B**, how many videos would you say other people have hired using your Selwyn Plaza Video card?  33-35

Thinking only about today, have you personally hired any videos from Selwyn Plaza Video?  
**IF YES** ... How many have you personally hired just today?  36  
**IF NO** ... **RECORD 0 IN THE BOX & GO TO Q7**

Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?  
**IF YES** ... How many have other people hired just today?  38  
**IF NO** ... **RECORD 0 IN THE BOX**

# IF INTERVIEWED LATE

Thinking about the past eight weeks, that is, since <DAY AND DATE> SEE CHART A to <ACTUAL DAY AND DATE>, how many videos have you personally hired from Selwyn Plaza Video?   40-

IF THE NUMBER = 0, THEN GO TO Q12.

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos have you personally hired from Selwyn Plaza Video?   42-

Does this number include any videos you personally have hired from Selwyn Plaza Video....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

Yes . . . . . 1 GO TO Q11

No . . . . . 2 GO TO Q12

44

How many videos have you personally hired from Selwyn Plaza Video....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

45-

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last eight weeks?

Yes . . . . . 1 GO TO Q13

No . . . . . 2 GO TO END

47

Thinking about the past eight weeks, that is, since <DAY AND DATE> SEE CHART A to <ACTUAL DAY AND DATE>, how many videos would you say other people have hired using your Selwyn Plaza Video card?    48-

IF THE NUMBER = 0, THEN GO TO END

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos would you say other people have hired using your Selwyn Plaza Video card?    52-

Does this number include any videos anyone else has hired from Selwyn Plaza Video....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

Yes . . . . . 1 GO TO Q16

No . . . . . 2 GO TO END

55

How many videos have other people hired using your Selwyn Plaza Video card....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

56

**ND** That's all I need to ask. Thank you for helping us with this survey. Bye

RECORD THE EXACT TIME NOW     58

Audited \_\_\_\_\_

## Control C (T5) Questionnaires

Version  1

Phase  2

Outcome  3

No. of Calls   4

Day   6

Month   8

Interviewer ID

Callback details \_\_\_\_\_ ID

Good evening. My name is <FULL NAME>. I'm doing a short survey about videos and would like to ask you a few questions. This interview will only take about five minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW

Before we start, I'd just like to assure you that everything you say is completely confidential.

Is there a video player in this household?

Yes ..... 1   
No ..... 2

Have you personally ever rented a video tape?

Yes ..... 1   
No ..... 2

Which of the following Palmerston North video stores are you a member of?

**READ LIST, CIRCLE ALL THAT APPLY**  
Selwyn Plaza Video ..... 1   
Video Concepts Albert Street . . . 1   
Video Concepts Featherston St . 1   
Movie World ..... 1   
Other ..... 1

**IF SELWYN PLAZA NOT CIRCLED THEN:** I'm sorry, but the rest of the questions relate to Selwyn Plaza Video, and so I don't need to ask you any more questions. Thank you for your time.

Selwyn Plaza Video is thinking about providing access to the internet. Customers would be able to use this services for \$5 per 15 minutes. We would like to know what the chances are of you using this service at Selwyn Plaza Video if it was available.

I would like you to answer on a scale of "zero" to "10". If you are certain, or practically certain that you will use this service, then you would choose the answer "10". If you think there is no chance, or almost no chance of using, the best answer would be "zero". If you are uncertain about the prospects, choose another answer as close to "0" or "10" as you think it should be. You can think of the numbers as chances out of "10". For example, 3 would mean 3 chances in 10 that you would use the service, while a 7 would mean 7 chances in 10 that you would use the service, and so on.

So, taking everything into account, what would be the chances of you personally using the internet service at Selwyn Plaza Video if it was available?

40-

How often do you think you would use this service in a four week period, if it was available?

42-

I'm now going to ask you some questions about the chances of you hiring videos. I'd like you to use the same 0 to 10 scale again.

Now, taking everything into account, what are the chances of you personally hiring at least one video from Selwyn Plaza Video, during the next four weeks, that is between now and <DAY AND DATE> **SEE CHART C?**

51-5

**IF THE PROBABILITY IS GREATER THAN ZERO, THEN ASK Q7, OTHERWISE GO TO Q9.**

If you did hire any, how many videos are you personally most likely to hire from Selwyn Plaza Video in the next four weeks?

53-

Using the scale from 0 to 10, what are the chances that you personally would hire exactly <**MOST LIKELY NUMBER**> videos from Selwyn Plaza Video in the next four weeks.

56-

Now, using the 0 to 10 scale again, what are the chances that someone else will use your Selwyn Plaza Video card to hire at least one video during the next four weeks, that is between now and <DAY AND DATE> **SEE CHART C?**

65-

**IF THE PROBABILITY IS GREATER THAN ZERO, THEN ASK Q10, OTHERWISE GO TO Q12.**

0. If they did hire any, how many videos are other people most likely to hire using your Selwyn Plaza Video card in the next four weeks?

67

1. Using the scale from 0 to 10, what are the chances that other people would hire exactly <**MOST LIKELY NUMBER**> videos with your Selwyn Plaza Video card in the next four weeks?

70

2. How many videos have you personally hired so far today from Selwyn Plaza Video?

72

3. How many videos have other people hired so far today using your Selwyn Plaza Video card?

74

4. If there is anything else that we need to know, can we call you back?

Yes ..... 1

No ..... 2

76

That's all I need to ask. Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW

77

Audited \_\_\_\_\_

Group 2 (Phase 2)

Questionnaire (Recall)

Version  1

Phase  2

Outcome  3

No. of Calls   4-5

Actual Day   6-7 Month   8-9

Day   10-11 Month   12-

Interviewer ID   1-4

Callback details \_\_\_\_\_ ID     1-6

Good evening. My name is <FULL NAME>. Four weeks ago, we interviewed you about videos, and you said we could call you back to ask a few more questions. This interview will only take about two minutes. May I interview you now? **GET CALLBACK DETAILS IF NECESSARY**

RECORD THE EXACT TIME NOW     2-0

**IF INTERVIEWED ON TIME, ASK ONLY Q1 - Q5.  
IF INTERVIEWED LATE, START FROM Q6.**

Since we interviewed you four weeks ago, that is, since <DAY AND DATE> **SEE CHART B**, how many videos have you personally hired from Selwyn Plaza Video?   2-6

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?  3-2

- Yes ..... 1 GO TO Q3
- No ..... 2 GO TO Q4

Since we interviewed you four weeks ago, that is, since <DAY AND DATE> **SEE CHART B**, how many videos would you say other people have hired using your Selwyn Plaza Video card?    3-3

Thinking about only today, have you personally hired any videos from Selwyn Plaza Video?  
**IF YES** . . . . How many have you personally hired just today?   3-6  
**IF NO** . . . . **RECORD 0 IN THE BOX & GO TO Q5**

Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?  
**IF YES** . . . . How many have other people hired just today?   3-8  
**IF NO** . . . . **RECORD 0 IN THE BOX**

**GO TO END ON PAGE 2**

# IF INTERVIEWED LATE

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos have you personally hired from Selwyn Plaza Video?

 42

Does this number include any videos you personally have hired from Selwyn Plaza Video....  
IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

Yes . . . . . 1 GO TO Q8

No . . . . . 2 GO TO Q9

 44

How many videos have you personally hired from Selwyn Plaza Video....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

 45-

Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last four weeks?

Yes . . . . . 1 GO TO Q10

No . . . . . 2 GO TO END

 51

Thinking about the four weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos would you say other people have hired using your Selwyn Plaza Video card?

 52-5

Does this number include any videos anyone else has hired from Selwyn Plaza Video....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

Yes . . . . . 1 GO TO Q12

No . . . . . 2 GO TO END

 55

How many videos have other people hired using your Selwyn Plaza Video card....

IF 1 DAY LATE . . . just today?

IF >1 DAY LATE . . . since last <ACTUAL DAY>?

 56

**END** That's all I need to ask. Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW

 58

Audited \_\_\_\_\_

## Bounded Juster (T6) Questionnaires

Callback details \_\_\_\_\_

ID

Grid for ID and other variables

Good evening. My name is <FULL NAME>. I'm doing a short survey about videos and would like to ask you a few questions. This interview will only take about five minutes. May I interview you now? GET CALLBACK DETAILS IF NECESSARY

RECORD THE EXACT TIME NOW

Time recording grid

Before we start, I'd just like to assure you that everything you say is completely confidential.

1. Is there a video player in this household?

- Yes ..... 1
No ..... 2

Response grid for Q1

2. Have you personally ever rented a video tape?

- Yes ..... 1
No ..... 2

Response grid for Q2

3. Which of the following Palmerston North video stores are you a member of?

READ LIST, CIRCLE ALL THAT APPLY

- Selwyn Plaza Video ..... 1
Video Concepts Albert Street . . . 1
Video Concepts Featherston St . 1
Movie World ..... 1
Other ..... 1

Response grid for Q3

IF SELWYN PLAZA NOT CIRCLED THEN: I'm sorry, but the rest of the questions relate to Selwyn Plaza Video, and so I don't need to ask you any more questions. Thank you for your time.

4. Selwyn Plaza Video is thinking about providing access to the internet. Customers would be able to use this services for \$5 per 15 minutes. We would like to know what the chances are of you using this service at Selwyn Plaza Video if it was available.

I would like you to answer on a scale of "zero" to "10". If you are certain, or practically certain that you will use this service, then you would choose the answer "10". If you think there is no chance, or almost no chance of using, the best answer would be "zero". If you are uncertain about the prospects, choose another answer as close to "0" or "10" as you think it should be. You can think of the numbers as chances out of "10". For example, 3 would mean 3 chances in 10 that you would use the service, while a 7 would mean 7 chances in 10 that you would use the service, and so on.

So, taking everything into account, what would be the chances of you personally using the internet service at Selwyn Plaza Video if it was available?

Response grid for Q4

5. How often do you think you would use this service in a four week period, if it was available?

Response grid for Q5

I'm now going to ask you some questions about the chances of you hiring videos. I'd like you to use the same 0 to 10 scale again.

6. Now, taking everything into account, what are the chances of you personally hiring at least one video from Selwyn Plaza Video, during the next eight weeks, that is between now and <DAY AND DATE> SEE CHART D?

Response grid for Q6

IF THE PROBABILITY IS GREATER THAN ZERO, THEN ASK Q7, OTHERWISE GO TO Q9.

If you did hire any, how many videos are you personally most likely to hire from Selwyn Plaza Video in the next eight weeks?

Using the scale from 0 to 10, what are the chances that you personally would hire exactly <MOST LIKELY NUMBER> videos from Selwyn Plaza Video in the next eight weeks.

Now, taking everything into account, what are the chances of you personally hiring at least one video from Selwyn Plaza Video, during the next four weeks, that is between now and <DAY AND DATE> SEE CHART C?

IF THE PROBABILITY IS GREATER THAN ZERO, THEN ASK Q10, OTHERWISE GO TO Q12.

If you did hire any, how many videos are you personally most likely to hire from Selwyn Plaza Video in the next four weeks?

Using the scale from 0 to 10, what are the chances that you personally would hire exactly <MOST LIKELY NUMBER> videos from Selwyn Plaza Video in the next four weeks.

Now, using the 0 to 10 scale again, what are the chances that someone else will use your Selwyn Plaza Video card to hire at least one video during the next eight weeks, that is between now and <DAY AND DATE> SEE CHART D?

IF THE PROBABILITY IS GREATER THAN ZERO, THEN ASK Q13, OTHERWISE GO TO Q15.

If they did hire any, how many videos are other people most likely to hire using your Selwyn Plaza Video card in the next eight weeks?

Using the scale from 0 to 10, what are the chances that other people would hire exactly <MOST LIKELY NUMBER> videos with your Selwyn Plaza Video card in the next eight weeks?

Now, using the 0 to 10 scale again, what are the chances that someone else will use your Selwyn Plaza Video card to hire at least one video during the next four weeks, that is between now and <DAY AND DATE> SEE CHART C?

IF THE PROBABILITY IS GREATER THAN ZERO, THEN ASK Q16, OTHERWISE GO TO Q18.

If they did hire any, how many videos are other people most likely to hire using your Selwyn Plaza Video card in the next four weeks?

Using the scale from 0 to 10, what are the chances that other people would hire exactly <MOST LIKELY NUMBER> videos with your Selwyn Plaza Video card in the next four weeks?

How many videos have you personally hired so far today from Selwyn Plaza Video?

How many videos have other people hired so far today using your Selwyn Plaza Video card?

If there is anything else that we need to know, can we call you back?

Yes . . . . . 1  
No . . . . . 2

That's all I need to ask. Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW

Audited \_\_\_\_\_

Version	5
Phase	3
Outcome	
No. of Calls	
Month	
Month	
ID	

Actual Day

Day

Interviewer ID

Callback details \_\_\_\_\_ ID

Good evening. My name is <FULL NAME>. Eight weeks ago, we interviewed you about videos, and you said we could call you back to ask a few more questions. This interview will only take about two minutes. May I interview you now?

GET CALLBACK DETAILS IF NECESSARY

RECORD THE EXACT TIME NOW

IF INTERVIEWED ON TIME, ASK FROM Q1 - Q8.

IF INTERVIEWED LATE, START FROM Q9.

- Thinking about the past four weeks, that is, since <DAY AND DATE> SEE CHART A, how many videos have you personally hired from Selwyn Plaza Video?
- And, since we interviewed you eight weeks ago, that is, since <DAY AND DATE> SEE CHART B, how many videos have you personally hired from Selwyn Plaza Video?
- Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last eight weeks?  
 Yes ..... 1 GO TO Q4   
 No ..... 2 GO TO Q6
- Thinking about the past four weeks, that is since <DAY AND DATE> SEE CHART A, how many videos would you say other people have hired using your Selwyn Plaza Video card?
- Now, since we interviewed you eight weeks ago, that is, since <DAY AND DATE> SEE CHART B, how many videos would you say other people have hired using your Selwyn Plaza Video card?
- When was the last time anybody including yourself hired a video using your Selwyn Plaza Video card?  
 \_\_\_\_\_ Days Ago      \_\_\_\_\_ Weeks Ago      \_\_\_\_\_ Months Ago  
 \_\_\_\_\_ Date              \_\_\_\_\_ Yesterday      \_\_\_\_\_ Others
- Thinking only about today, have you personally hired any videos from Selwyn Plaza Video?  
 IF YES ... How many have you personally hired just today?    
 IF NO ... RECORD 0 IN THE BOX & GO TO Q8
- Thinking only about today, has anyone else hired any videos from Selwyn Plaza Video using your Selwyn Plaza Video card?  
 IF YES ... How many have other people hired just today?    
 IF NO ... RECORD 0 IN THE BOX

# IF INTERVIEWED LATE

1. Thinking about the past four weeks, that is, since <DAY AND DATE> SEE CHART A to <ACTUAL DAY AND DATE> , how many videos have you personally hired from Selwyn Plaza Video?
2. Thinking about the eight weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos have you personally hired from Selwyn Plaza Video?
11. Does this number include any videos you personally have hired from Selwyn Plaza Video....  
**IF 1 DAY LATE** ... just today?  
**IF >1 DAY LATE** ... since last <ACTUAL DAY>?  
 Yes ..... 1 GO TO Q12   
 No ..... 2 GO TO Q13
12. How many videos have you personally hired from Selwyn Plaza Video....  
**IF 1 DAY LATE** ... just today?  
**IF >1 DAY LATE** ... since last <ACTUAL DAY>?
13. Has anyone else, other than yourself, used your Selwyn Plaza Video card to rent a video during the last eight weeks?  
 Yes ..... 1 GO TO Q14   
 No ..... 2 GO TO Q18
14. Thinking about the past four weeks, that is, since <DAY AND DATE> SEE CHART A to <ACTUAL DAY AND DATE>, how many videos would you say other people have hired using your Selwyn Plaza Video card?
15. Thinking about the eight weeks up to last <ACTUAL DAY>, that is from when we first interviewed you on <DAY AND DATE> SEE CHART B to <ACTUAL DAY AND DATE>, how many videos would you say other people have hired using your Selwyn Plaza Video card?
16. Does this number include any videos anyone else has hired from Selwyn Plaza Video....  
**IF 1 DAY LATE** ... just today?  
**IF >1 DAY LATE** ... since last <ACTUAL DAY>?  
 Yes ..... 1 GO TO Q17   
 No ..... 2 GO TO Q18
17. How many videos have other people hired using your Selwyn Plaza Video card....  
**IF 1 DAY LATE** ... just today?  
**IF >1 DAY LATE** ... since last <ACTUAL DAY>?
18. When was the last time anybody including yourself hired a video using your Selwyn Plaza Video card?  
 \_\_\_\_\_ Days Ago      \_\_\_\_\_ Weeks Ago      \_\_\_\_\_ Months Ago  
 \_\_\_\_\_ Date              \_\_\_\_\_ Yesterday      \_\_\_\_\_ Others

**END** That's all I need to ask. Thank you for helping us with this survey. Bye.

RECORD THE EXACT TIME NOW

Audited \_\_\_\_\_

## Phase 1 Chart



## Phase 2 Chart



## Phase 3 Chart

