EVERY LITTLE HELPS

AN INVESTIGATION OF FREQUENCY BIASES IN

COMPARATIVE JUDGMENTS

RICHARD LEWIS

A thesis submitted for the degree of Doctor of Philosophy to the

Department of Psychology at the University of Warwick

MARCH 2010

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

jnd = just-noticeable difference

LCJ = Law of Comparative Judgment

SDT = Signal Detection Theory

ROC = Receiver Operating Characteristic

IIT = Information Integration Theory

ERP = External Reference Price

IRP = Internal Reference Price

EDLP = Every Day Low Pricing

AIC = Akaike Information Criterion

PW = Purchase-Weighting

TW = Time-Weighting

BC = Basket Cost Comparison

ACKNOWLEDGEMENTS

I am grateful to the Decision Technology Research Group for funding this work and supplying the necessary resources. In particular I would like to thank my supervisor Nick Chater, for his cheerful encouragement, support and patience. I would also like to thank other Dectech colleagues past and present – especially Henry Stott, Greg Davies, Benny Cheung, Ivo Vlaev and Christoph Ungemach – for their intellectually-stimulating ideas, debate, comments and challenges.

Many other people have contributed their personal support and intellectual encouragement during the writing of this thesis, especially: Derrick Watson, Stian Reimers, Neil Stewart, Gaynor Foster, Gary Foster, Kim Wade, Al Galloway, Koen Lamberts, James Adelman, Liz Blagrove, Nilufa Ali, Duncan Guest, Caroline Morin, Rod Freeman, Lucy Bruce, Khalilah Kahn and Nicola Grant. I would like to express my gratitude to everyone who has helped or contributed in any way, as well as all the people who participated in the research.

I would like to dedicate this thesis to the two most important women in my life, without whom this thesis would never have been completed:

My mother - for love, encouragement and proof-reading.

My wife, Corinna - for everything.

DECLARATION

I hereby declare that the work reported in this thesis is my own work, unless stated otherwise. No part of this thesis has been submitted for a degree at another University. No part of this work has yet been presented at conferences and workshops nor been submitted for publication.

Richard Lewis

March 2010

ABSTRACT

Intuitive statistical inferential judgments involve the estimation of statistical properties of samples of information, such as the mean or variance. Prior research has shown that human judges are generally good at making unbiased estimates of sample properties. However, a series of recent applied consumer research experiments demonstrated a systematic bias in comparative judgments of item distributions in which the individual items are paired across those distributions, for example comparing the prices in two stores selling the same items. When the two distributions have the same mean, the distribution with the higher number of items that are smaller in magnitude than the equivalent item in the other distribution is typically judged to be the smaller of the two distributions: a frequency bias. In a series of experiments, the research in this thesis provides a robust demonstration of the frequency bias and explores possible explanations for the bias. A comparison between simultaneous and sequential presentation of information demonstrates that the frequency bias cannot solely be explained by the salience of the frequency cue. A novel web-based experiment, in which information was sampled incidentally from the environment and a naturalistic task was used to elicit comparative judgments, showed that the frequency effect persists in an ecologically-valid context. A systematic comparison between alternative cognitive models of the judgment process supports an explanation in which items are recalled from memory and compared in a pair-wise fashion, meaning the frequency bias may be found in a wide range of other judgment tasks and domains, which would have significant implications for our understanding of intuitive comparative judgments.

CHAPTER 1

STIMULUS DISCRIMINATION AND INTUITIVE STATISTICS LITERATURE REVIEW

1.1 Introduction

This thesis examines the ability of human subjects to discriminate between two complex stimuli in a naturalistic task and explores the possible cognitive processes involved in performing the required intuitive statistical judgment. The task adopted is that of determining the relative price level of two stores, an applied problem which has received increasing attention in the consumer research literature in recent years. This first chapter provides an overview of research into discrimination and intuitive statistical judgments, and related concepts. It will begin with a brief description of some intriguing (and conflicting) research findings from the consumer research literature concerning consumer price judgments. It will then present a brief history of psychological and psychophysical research into the detection and discrimination of stimuli and of human intuitive statistical judgments. It will focus particularly on the special role of frequency information in serially-experienced events, as this may be vital in understanding the cognitive processes underlying many common statistical judgments. In addition, this chapter will briefly review relevant concepts and findings concerning the architecture and performance of human memory, which will be important when considering discrimination tasks in which one of the experimental stimuli has to be recalled from memory. Some alternative schools of thought concerning the appropriate research methodology for exploring human performance in intuitive judgments will also be briefly presented and the relative merits of factorial experimental and naturalistic ecologically-representative research will be discussed. In subsequent sections, related price judgment research findings from the consumer research literature will be summarized and the implications for the research presented in this thesis will be drawn out. Finally, the chapter will conclude with an outline of the motivation for the thesis and the choice of research methodology, before briefly describing the structure of the following chapters.

1.2 The Relative Impact of Frequency and Magnitude Cues in Consumer Price Judgments

1.2.1 Comparative Price Data

A common tactic in grocery store advertising is to compare the prices of a selection of items against the prices of the same items in a competitor store. In a series of experiments reported in the Journal of Consumer Research in 1994, Joseph Alba and his colleagues systematically explored the relative impact of three factors that influence consumers' price judgments of such comparative price data (Alba, Broniarczyk, Shimp, & Urbany, 1994). The three factors were prior beliefs about the prices in each store, the number of items for which each store was cheaper than the other (the 'frequency cue') and the average size of those price advantages (the 'magnitude cue'). The experiments are described in some detail here as they are directly pertinent to the main subject of this thesis.

In Experiment 1 (p. 222), price judgments were operationalized as the difference between participants' estimates of the total cost of 60 items in each store, which participants were told lay between $100 and $130. In each case the total cost of the 60 items was identical in the two stores ($117.13) but one store was cheaper by about $0.07 on 40 of the 60 items (which had an average price of $1.91) and more expensive by about \$0.14 on the remaining 20 items (which had an average price of \$1.89). The prices were presented in a list format, with the price in each store presented side-by-side after each item description. The price lists were presented in a booklet with 15 items per page. Between-subjects manipulations of prior beliefs (using real store names) and processing time were used in a two-way factorial design. Contrary to the authors' expectations, price judgments in an initial exploratory experiment were not driven primarily by prior beliefs about the two stores but by the frequency cue: in every experimental condition in Experiment 1, the mean estimated cost in the frequency store was lower than in the magnitude store.

In a series of follow-up experiments, the authors explored the boundary conditions of this 'frequency effect' and some possible competing hypotheses for why the frequency cue dominated the price judgments. In Experiment 1A (p. 224) they found no moderating effect of either shortening or lengthening the time allocated to the task. In Experiment 1B (p. 225) they found that strengthening the prior belief manipulation by using fictional store descriptions rather than real store names also had no moderating effect. In Experiment 1C (p. 226) the prior beliefs were manipulated by using real store names but changing the credibility of the source of the price information. Again, even when prior beliefs were inconsistent with the frequency cue and the price information came from a highly credible source, the frequency store was consistently judged as the cheaper of the two stores.

In Experiment 2 (p. 226) the authors switched from presenting all price information in a booklet to presenting sets of six pairs of prices on a computer screen for 30 seconds at a time. Within each set of items, four items were cheaper in the frequency store and two items were cheaper in the magnitude store. Participants were able to choose how many sets of prices they viewed (between one and ten) before identifying the cheaper store. A $3 budget was decremented by $0.30 for each additional set of prices seen, but participants were told they would receive no payment if they failed to correctly identify the cheaper store. Information search was low across the experimental conditions, with an average of 2.2 sets of prices viewed. Again, the frequency cue dominated the price judgments, with 69% of participants judging the frequency store as cheaper when prior beliefs favoured the magnitude store and 88% of participants judging the frequency store as cheaper when prior beliefs also favoured the frequency store. In Experiment 3 (p. 227) the authors reduced the number of items used to just nine, presented in three sets of three. Within each set of three prices, two items were cheaper in the frequency store. Prior beliefs were not manipulated, but the salience of the magnitude cue was varied between subjects by presenting either three infrequently-purchased or three frequently-purchased items that were cheaper in the magnitude store. Participants were asked to identify the cheaper of the two stores, to estimate the total cost of the nine items in the magnitude store given that the total cost was $125 in the frequency store (bounded between $110 and $140), and to indicate the store they would personally choose in order to obtain good value. The proportion of participants identifying the frequency store as cheaper did not differ significantly between the low and high magnitude cue salience conditions (88% and 71% respectively). Similarly, the average basket cost estimate did not vary significantly between the low and high salience conditions ($129.44 and $128.61 respectively) although in both cases the total cost in the magnitude store was estimated to be significantly higher than in the frequency store. However, there was a significant shift in store preference toward the magnitude store in the high salience condition, although even in that case more subjects preferred the frequency store (64% in the high salience condition vs. 92% in the low salience condition).

In order to try and explore why the frequency cue dominated the price judgment tasks, the authors first examined informal descriptions of the price judgment strategies used, as written by participants at the end of the tasks. Of the rationales given, 78% clearly referred to strategies involving the frequency cue, magnitude cue or prior beliefs. Of those, 87% referred to the frequency cue, 8% to the magnitude cue and only 5% to prior beliefs about the two stores. This was followed by Experiment 4 (p. 229) in which the 60 items and prices from the initial experiments were presented for free inspection, followed by a numerical distracter task. Participants were then presented with a randomly-ordered list of the 30 middle items along with the price in the magnitude store and were asked to recall the price of each item in the frequency store. The results were noisy, but showed that when prior beliefs were consistent with the frequency cue, participants were reasonably accurate in their recall of the frequency cue (18.1 items were judged cheaper, when the correct answer was 20) but were insensitive to the magnitude cue with poor recall of item price differences. When prior beliefs were inconsistent with the frequency cue, sensitivity to (and accurate recall of) magnitude information increased but to a relatively small degree.

When the total cost is the same in two stores, then the frequency cue and magnitude cue are negatively correlated. In Experiment 5 (p. 230) the authors attempted to sensitize participants to this trade-off by explaining that stores could discount lots of items by a small amount or a few items by a large amount, making it difficult to determine which store is cheaper overall. A subset of the participants was also encouraged to try and keep a running total of the price differential between the stores as they studied the price lists. The same 60 items and prices from earlier experiments were used and again participants were asked to judge the total cost of those items in each store. There was no significant effect of either instruction format or processing time on the differences in estimated price between the two stores. Relative to Experiment 1, the mean estimated total cost advantage for the frequency store was smaller ($4.38 vs. $7.57) but still differed significantly from zero. Thus, despite attempting to sensitize participants to the trade-off between the frequency and magnitude cues, the frequency cue continued to dominate the price judgments. Finally, in Experiment 6 (p. 231) the salience of the magnitude cue was manipulated by replacing the item prices in the magnitude store with the signed difference in price between the frequency and magnitude stores. Hence, the list consisted of 40 positively-signed dollar differences and 20 negatively-signed dollar differences (as well as the 60 item prices in the frequency store). This format was intended to make the calculation of the overall price differential easier for participants, although the authors conceded that this format also heightens the salience of the frequency cue. Perhaps unsurprisingly then, the frequency store was again judged as cheaper, with a mean estimated total cost advantage of $7.82. The authors conclude that:

> "...these studies confirm our hypothesis regarding the inherent salience and persuasiveness of frequency cues. We found little evidence to support the alternative hypotheses that subjects are equally sensitive to frequency and magnitude but weigh the former more heavily or that the bias to favour the store with the frequency advantage is due to computational deficits." (Alba, et al., 1994)

However, they note that "additional research involving process measures may be needed" in order to explain the dominance of the frequency cue (p. 232).

Additionally, they recognize that "in most cases [...] the subject was compelled to process all the information [and hence] the salience of these cues, and therefore their influence on price perceptions, is unknown for contexts in which they compete for attention on unequal footing" (p. 233). In particular they suggest that "one approach would be to employ computer-based shopping methods, which would allow assessment of how consumers form price beliefs when direct store-to-store price comparisons are hindered by the natural sequential nature of shopping" (p. 233).

1.2.2 Sequentially-Sampled Price Data

In fact, a first attempt at just such an experiment had already been made eight years earlier by B. Kemal Buyyukurt and also published in the Journal of Consumer Research (Buyukkurt, 1986). The author was attempting to differentiate between different theoretical models of how a consumer might estimate and update the perceived value of a basket of items as they serially experience a sequence of item prices, analogous to serially sampling price information by selecting items to purchase during a shopping task. The exact specification of each model is not relevant to the current discussion, but each was based on the assumption that the overall judged value of a basket of items is based upon a weighted average of the perceived value of each individual item in the basket. The weight placed upon each item was hypothesized to depend upon serial-order effects such as primacy or recency effects. The perceived value of each item was hypothesized to depend on factors including the difference between the expected and observed price and the degree of certainty associated with each expected price. In particular, Buyukkurt proposed three different functions mapping the difference between the observed and expected price (expressed either as an absolute or percentage difference), D, and a component of the perceived value of each item, $V_1(D)$: an S-shaped, linear or exponential valuation function. The three functions are reproduced in Figure 1.1 below.

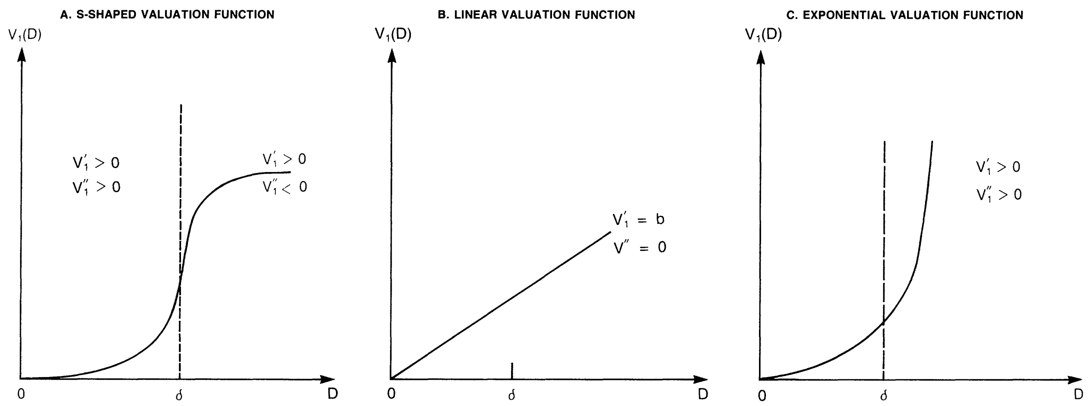

Figure 1.1: Alternative item valuation functions proposed by Buyukkurt (1986, Figure A).

The three valuation functions all include a threshold δ, below which it is assumed that differences between expected and observed prices are not noticed. No theoretical justification for the threshold is given, being described simply as a "psychological threshold" (p. 360). Similarly, no theoretical justification is given for the choice of the three alternative valuation functions, being described as "three mathematical forms [...] out of many theoretically possible" (p. 360). Nonetheless, Buyukkurt goes on to describe how each valuation function has different implications for the most highly-valued discount structure of a basket of items. Assuming that all discounts are greater than the threshold δ, the diminishing returns of the S-shaped valuation function imply that a basket with many small discounts would be perceived to offer the best value. On the contrary, the increasing returns of an exponential valuation function imply that a basket with a few large discounts would be perceived to offer the best value. A linear valuation function would imply no difference in perceived value between a frequency and a magnitude basket, provided the total cost was the same. In addition, the model predicts that if a frequency effect (implying an S-shaped valuation function) was observed, that it would be strongest when the basket contained more items and when uncertainty about the expected prices was low.

Buyukkurt also carried out an experiment to test a number of predictions generated by the theoretical models, including the discount structure predictions described above. The participants in the experiment were primary grocery shoppers for their household, intercepted in a shopping mall. Each participant selected 20 items from a list of 42: ten items where they felt quite certain about the usual selling price and ten items which they had previously purchased but for which they were uncertain about the usual selling price. The participant also gave a likely price and price range for each of the 20 items. The likely price estimate (or an average of estimates from a pilot study when the participant could not give a likely price) was used as the expected price of each item, with a small fluctuation of two percent randomly added or subtracted. A basket of items was then constructed for the participant, in which the number of items, the ordering of items and the discount structure were all varied as between-subjects factors. Small baskets contained ten items (five high certainty and five low certainty items) and large baskets contained all twenty items. In frequency baskets, 40% of the items were discounted by 12.5% while in magnitude baskets 20% of the items were discounted by 25%. The participants were then instructed to imagine themselves on a shopping trip in an unfamiliar store and to be examining the prices of a basket of items that they would normally purchase, with the intention of ultimately choosing between the new store and their usual grocery store. Information about each item (brand name, product name, package size and number of units purchased), along with the price paid, was displayed on a computer screen for 12 seconds each. Finally, the total cost of the basket was displayed to simulate the total bill paid at the checkout. The participant then rated the prices in the store relative to their usual store on three different seven-point Likert scales as well as estimating how much the same basket of items would cost in their usual store. The four response variables were highly correlated, so a factor analysis was used to create a linear composite of the four relative price judgments.

Because each basket of items and their expected prices had been tailored to the individual participant, the data were analysed using ANCOVA analysis with the total cost of the basket of items as a covariate. The predicted frequency effect in the discount structure was significant at $\alpha=0.05$, with frequency baskets being perceived to offer significantly better value than magnitude baskets, although the effect only accounted for three percent of the observed variance. However, the predicted interactions between discount structure and basket size and between discount structure and price certainty were not significant. In addition, no serial order effects (primacy or recency) were observed. Nonetheless, the frequency effect described by Alba et al was found when price information was sequentially sampled. Buyukkurt notes that the experimental procedure had a number of limitations, some of which would prevent strong conclusions being extrapolated from the experimental findings to real-world effects. Firstly, the information presentation rate was much higher than the acquisition rate in a store. Secondly, the price manipulation was one-sided, whereas in real stores it is highly unlikely that none of the items would have a higher-than-average selling price. Thirdly, there were no attention-distracting cues in the laboratory setting and participants were forced to pay attention to the price of each item. Fourthly, the non-random sampling method and relatively skewed demographic profile of the sample limits the generalizability of the findings. Finally, and perhaps most importantly, because the basket of items and prices was tailored to each individual, both the observed prices and total monetary discount varied between participants, introducing a source of additional variance which the ANCOVA analysis could only partially account for.

1.2.3 Temporal Distributions of Item Prices

Joseph Alba published a follow-up paper five years after his original frequency effect findings (Alba, Mela, Shimp, & Urbany, 1999) in which he and his colleagues studied another complex price judgment task, closely related to the earlier task of comparing a selection of item prices between two stores. This time the research focused upon an alternative strategy for determining which of two stores is cheaper: observing the price of a single item in two different stores over a period of time in order to judge which store has the lower average price. This is analogous to the previous basket comparison in that it involves comparing two distributions of prices. It also involves sequential-sampling of price information, similar to the study described above (Buyukkurt, 1986). However, there are two important differences. Firstly, the term ‘frequency’ changes from describing the number of items on which each store is cheaper to describing the number of times that a single item is discounted from its usual selling price, *regardless of when those discounts are applied*. This is subtly different to counting the number of occasions on which the item is cheaper in one store than the other, which would be a closer analogue of the previously described paired-price frequency cue. Secondly, the potential for strategic purchasing behaviour to influence price judgments is much greater in this second case. For non-perishable items, purchasing the item when it is discounted and stock-piling inventory for future consumption would mean that the average *purchase* price would be lowest when a single deep discount was applied, relative to a series of frequent shallow discounts, even if the average *presented* price were the same or higher over the time period considered. As will be discussed in more detail later in this chapter, this strategic purchasing tactic is more difficult when consumers shop across a range of items in a basket, particularly for consumers who shop infrequently for large baskets of items (Bell & Lattin, 1998). It is not clear whether consumers would use the judged average presented price or purchase price when comparing between two stores, nor do the authors offer any evidence that this brand-over-time comparison strategy is employed by consumers in the real world. Nonetheless, the studies offer a close enough parallel to the earlier studies to merit comparison and the authors began the second series of studies expecting to observe a similar frequency effect in the new price comparison task.

In a pilot study, participants took part in a 'buying game' in which they observed the prices of three brands of shampoo over 36 simulated months. In each month they chose whether or not to purchase, which brand to choose, and how many units to purchase. The participants were provided with a simple formula for calculating inventory costs ($0.10 per bottle stored and not consumed that month) and told that they consumed one bottle of shampoo each month. Their task was to minimize their combined inventory and purchase costs over the 36 months, whilst ensuring sufficient inventory for their required consumption each month. At the end of the 36 months participants provided estimates of each brand's average price, sale price, regular price and promotional frequency. The three brands consisted of a control brand (constant price of $2.39), a frequency brand (usual selling price of $2.49 for 18 months, discounted price of $2.29 for 18 months) and a magnitude brand (usual selling price of $2.49 for 33 months, discounted price of $1.29 for 3 months). The discounts were distributed uniformly throughout the 36 month period, so that the frequency brand had three sale months randomly assigned in each six month period and the magnitude brand had one sale month randomly assigned in each 12 month period. The magnitude brand also had no discount applied in the final five months to avoid any recency effects. It should be noted that the difference between the frequency and magnitude brands (reduced six times more often vs. six times greater discount depth) is three times larger here than the difference between the frequency and magnitude stores in the 1994 studies (reduced twice as often vs. twice the discount depth).

Repeated measures ANOVA analysis showed significant differences in average price estimates and discount frequency estimates between the three brands. The estimated average price for the magnitude brand was much lower than for the frequency brand ($2.18 vs. $2.33). Participants also underestimated the discount frequency of the frequency brand (9.33 instead of 18) and overestimated the discount frequency of the magnitude brand (4.22 instead of 3). This average price difference counteracts the previously observed frequency effect, so a series of follow-up studies were again conducted to explore possible explanations for the discrepancy. Three specific explanations were proposed. Explanation one: that the average price estimates were correctly calculated as a weighted average of the usual and sale prices, but participants systematically underestimated high frequencies and overestimated low frequencies. A possible cause for underestimation of the high discount frequency is that the small discounts for the frequency brand may have been subsumed into a latitude of acceptable prices (Monroe, 1971a) or may have been small enough to fall into a region of perceptual indifference ($\delta > 0.08$ using the terminology of Buyukkurt above). Explanation two: that the deep sale price of the magnitude brand was systematically over-weighted in calculating the numerical average of the prices. A possible cause for over-weighting the deep discount price would be high availability in memory caused by the extremity of the observation (Tversky & Kahneman, 1973) or simply that more attention was paid to the sale price because the depth brand was usually purchased when it was on sale. Explanation three: that the dichotomous price distributions in the pilot study were much less complex than the basket price distributions from the 1994 study, and that participants were more likely to fall back on a frequency heuristic when their cognitive resources are stretched by complex stimuli. A magnitude effect would be especially likely in a dichotomous price distribution if participants used an 'anchor-and-adjust' strategy to estimate the average price, with the salient low sale price used as the anchor.

Study 1 (p. 103) replicated the pilot study, but introduced an additional two-level between-subjects factor by flagging a discounted price with the word 'Sale' for half the participants. Although discount frequency estimates for the frequency brand were much closer to the correct value (13.1 flagged vs. 6.1 un-flagged), the average price estimates for the magnitude and frequency brands did not differ significantly between the flagged and un-flagged groups. Hence, the authors reject the first potential explanation, that of systematic biases in frequency estimation. Study 2 (p. 105) reduced the extremity of the discount for the magnitude brand by a factor of three (and therefore increased the discount frequency by a factor of three) in order to bring the difference between the frequency and depth brands in line with the 1994 study. The magnitude effect was reduced, but the magnitude brand was still perceived to have a lower average price than the frequency brand ($2.31 vs. $2.35) and discount frequency estimates for each brand were consistent with the prior studies. Hence, the authors reject the explanation that the extremity of the sale price for the magnitude brand in the pilot study was sufficient to reverse the frequency effect. Study 3 (p. 105) removed the potential attention effect caused by purchasing the magnitude brand in the buying game by reverting to the simple paired-price list format used in the 1994 study and eliminating the constant price brand. The discount depth advantage of the magnitude brand was introduced as a two-level between-subjects factor (2X vs. 6X) and a 'basket cost' measure (estimated total cost if the shampoo was purchased once each month over the 36 months) was also collected. The difference between discount depths was not significant and the magnitude effect persisted for both average price and basket cost estimates. Hence, the authors reject the explanation that the depth effect was caused by increased attention being paid to the magnitude brand's sale price in the buying game paradigm.

Study 4 (p. 106) increased the similarity to the 1994 study by replacing the two brands with a single brand's prices observed in two different stores. The price distributions were created by taking the first 36 items from the 1994 study and assigning a price of $2.49 to whichever was the higher priced brand that month, and using the price differential from the 1994 study to determine the price of the discounted brand. As a result, the frequency brand was discounted in 24 months and the magnitude brand was discounted in 12 months, with the two brands never discounted in the same month. The price distributions were non-dichotomous and overlapping, in that some of the discounts for the magnitude store ($0.06 - $0.18) were smaller than some of the discounts for the frequency store ($0.03 - $0.13). A second set of non-overlapping prices was created by subtracting a constant $0.18 from the magnitude brand's sale prices and a constant $0.09 from the frequency brand's sale prices. The level of complexity (overlapping vs. non-overlapping) was used as a between-subjects factor. For the first time in this series of studies a frequency effect was observed, with the mean basket price for 36 months of shampoo being lower in the frequency store than in the magnitude store ($84.05 vs. $86.69). There was, however, no difference between store estimates in the overlapping and non-overlapping groups suggesting that the frequency effect was caused by the adoption of non-dichotomous price distributions. Study 5 (p. 108) explicitly tested this hypothesis by pitting the dichotomous distributions from Study 2 against the non-dichotomous, non-overlapping distributions from Study 4 in a between-subjects manipulation. Repeated-measures ANOVA showed a significant interaction between the store (frequency or magnitude) and the complexity of the distributions (dichotomous or non-dichotomous) with the magnitude store being perceived as cheaper for dichotomous price distributions and the frequency store being perceived as cheaper for non-dichotomous price distributions. The authors also note that estimates of discount frequency and regular selling price for the two stores did not vary significantly between the dichotomous and non-dichotomous cases, but the estimates of the sale price were much more accurate for dichotomous price distributions. They argue that this supports the idea that participants used an anchor-and-adjust strategy using the perceived sale price as the anchor, although they admit that increasing the complexity of the price distribution may also have led participants to switch from making within-store estimates of the average price to using a between-stores comparison strategy of counting the number of months in which each store had the lower price.

A recent follow-up study by Lalwani and Monroe argued that the switch from a frequency effect to a magnitude effect could better be explained by the relative salience of the frequency and magnitude cues, rather than the complexity of the stimuli (Lalwani & Monroe, 2005). The authors first replicated the findings of Study

5 above (Alba, et al., 1999, pp. 108-110) using a simple buying game with 36 months and two brands of shampoo, with either dichotomous or non-dichotomous price distributions. They then repeated the experiment, but this time switched the product to a desktop computer, ranging in price from $520 to $740 in the non-dichotomous condition and from $530 to $740 in the dichotomous condition. Thus the magnitude of discounts and the regular selling price were about 580 times larger, but all other aspects of the price distributions were unchanged. The results showed that the magnitude brand was perceived as having the lower average price not only in the dichotomous condition ($710.23 vs. $714.00) but also in the non-dichotomous condition ($706.43 vs. $717.93). Lalwani and Monroe argue that increasing the magnitude of the discounts made the magnitude cue more salient in the second study, causing the switch from a frequency effect in their first study to a magnitude effect in their second study, for non-dichotomous price distributions. They support this conclusion with a third experiment using the same stimuli as their first study, but increasing the salience of the frequency cue by discounting the frequency brand in 20 of the 36 months and the magnitude brand in just 2 of the 36 months (a ratio of 10:1 instead of 2:1). In this case the frequency brand was perceived to have the lower average price for both dichotomous ($1.93 vs. $2.04) and non-dichotomous price distributions ($1.95 vs. $2.02). In this case, they argue that the frequency cue was made more salient than the magnitude cue, hence causing a switch to using the frequency cue. However, they are silent on why Alba et al. (1994) first observed a magnitude effect with a ratio of 6:1, nor do they address the fact that the frequency and magnitude cues are coupled when the average price of the two brands is held constant. Increasing the salience of the frequency cue also increases the salience of

the magnitude cue: in this case the change from a 2:1 ratio to a 10:1 ratio increases the depth of the discount for the magnitude brand from $0.30 to $1.00.

1.2.4 Conclusions from Experimental Findings

The frequency effect first observed by Buyukkurt (1986) and Alba et al. (1994) appears to be less robust than first thought, with relatively subtle changes in experimental task and price distributions causing it to either shrink or even disappear entirely. Although a series of experiments have explored the boundary conditions under which the frequency or magnitude cue dominates price judgments, no firm conclusions have been reached nor have convincing explanations for the observed pattern of results been put forward. Worse, the definition of 'frequency' varies arbitrarily between the frequency of within-store/brand discounts over time and the frequency of between-store/brand price advantages. Finally, the methodological limitations described by Alba et al. (1994), namely (i) forcing participants to pay attention to all price information and (ii) not presenting prices in a format which mirrors the sequential nature of between store price comparisons, have not yet been addressed. Nonetheless, the frequency effect offers an intriguing insight into the ability of consumers to discriminate between complex real-world stimuli, as well as having important practical implications for retailers, consumers and policy-makers. As will be shown in the following sections, understanding discrimination and judgments has been a central problem in mainstream psychology for a very long time, although judgment tasks analogous to those already described have so far received relatively little attention. I shall first describe some of the key theories and findings from research into psychophysical judgments, intuitive statistics and memory in the psychology literature before returning to other related findings from applied price research in the consumer research literature.

1.3 A Brief History of Judgment Research

1.3.1 Detection and Discrimination

1.3.1.1 Weber's Law and Fechner's Law

One of the most elementary cognitive processes is that of detecting an object (or sensation) or discriminating between two objects (or sensations), i.e. detecting a difference. The first answer to the question of whether or not a difference can be detected was given in the 19th century by Weber's Law:

$$\frac{\Delta I}{I} = k$$

Weber derived this relationship from a series of psychophysical experiments. In one example, a blindfolded participant would be given a reference weight to hold in one hand and then given a series of weights to hold in their other hand. The participant had to decide whether the second weight was lighter, heavier or the same as the reference weight. Weber found that the smaller the difference between the reference weight and the second stimulus, the smaller the proportion of tests in which the participant correctly identified the second weight as being heavier. Weber assumed that there was a fixed threshold or 'just-noticeable difference' (jnd) below which differences could not be detected. The jnd was often taken to be the threshold at which the difference in weight was correctly identified on 75% of trials (Gigerenzer & Murray, 1987). Weber found that the jnd was a constant proportion of the starting stimulus intensity and his law is a mathematical expression of that finding. For example, if the jnd for a 100g weight was found to be 1g then the jnd for a 1kg weight would be 10g.

Fechner developed Weber's work with a mathematical derivation of a relationship between the stimulus intensity and a unit of sensation, S. He assumed that if a sensation difference of ΔS were to be just noticeable then it would be a constant multiple of the Weber fraction:

$$\Delta S = k\frac{\Delta I}{I}$$

Rearranging the terms and solving as a differential equation yields Fechner's Law:

$$S = k \ln I + C$$

where C is a constant of integration to be determined experimentally. Fechner's Law states that sensation is logarithmically related to stimulus intensity, so that if stimulus intensity is increased in a geometric progression (multiplied by a fixed constant) then perceived sensation increases as an arithmetic progression (adding a fixed amount).

Fechner's work was later criticized by Stevens who developed an alternative methodology for obtaining psychophysical judgments of stimulus intensity, known as magnitude estimation (Gigerenzer & Murray, 1987; Stevens, 1957). This methodology involved asking participants to report numbers in proportion to the sensation arising from different signals, a methodology which Stevens described as "direct" measurement and that he contrasted with the "indirect" methods of Weber and Fechner (Stevens, 1957). Stevens claimed that there are two classes of sensation, prothetic and metathetic, which approximately correspond to quantity (e.g. loudness) and quality (e.g. pitch). Furthermore, he claimed that prothetic sensations did not follow the logarithmic relationship proposed by Fechner, but were in fact better modelled by a power law:

$S = a\, I^{k}$

This form of relationship is equivalent to assuming that the relative difference in perceived sensation (ΔS/S) is proportional to the Weber fraction, and not the absolute difference (ΔS) as assumed by Fechner. Stevens also applied his direct techniques to measuring other forms of judgment, such as opinions and attitudes (Stevens, 1966).

Stevens' psychophysical work has itself been criticized on a number of fronts. Firstly, he obtained ratings from a number of subjects and then averaged them before fitting a power function, which fails to account for individual differences (Gigerenzer & Murray, 1987; Stevens, 1966). Secondly, the methodology of magnitude estimation does not allow for separate testing of participants' psychophysical sensation and any distortion that might occur in transforming that sensation into a numerical response (Luce, 2002). Thirdly, and most fundamentally, the very idea of fixed thresholds in stimulus discrimination has been strongly criticized by many psychophysical researchers, beginning with Thurstone in developing his 'Law of Comparative Judgment'.

1.3.1.2 Thurstone's Law of Comparative Judgment

The key concept in Thurstone's work was what he referred to as *discriminal dispersion*: the idea that there is ambiguity or variation in how one stimulus is perceived by a single observer on different occasions (Thurstone, 1927). If an observer is shown two stimuli, A and B, then on any single occasion i the observer will perceive the magnitude of A as a_i and the magnitude of B as b_i. Because of discriminal dispersion, on some occasions a_i will be perceived as greater than b_i and on other occasions a_i will be perceived as less than b_i. If the observer is asked on each occasion to judge whether A is larger or smaller than B, then after a number of trials the frequency of each response can be used to generate the probability that A is judged to be greater than B, $p_{A>B}$. If the magnitude of stimulus B is held constant and the magnitude of stimulus A is gradually increased then $p_{A>B}$ rises, just as Weber and Fechner observed. Thurstone himself made no claims about the source of this discriminal dispersion, stating simply that "as we inspect two or more specimens for the task of comparison there must be some kind of process in us by which we react differently to the several specimens" but that "you may suit your own predilections in calling this process physical, neural, chemical or electrical" (Thurstone, 1927, p. 274). However, what is quite clear is that he totally rejected the idea of fixed thresholds in discrimination: "Everyone who works at all seriously in psychophysics knows that just noticeable differences have never been found, and that it is necessary to specify quite arbitrarily a stipulated frequency of discrimination in order to put any sense in the jnd" (Gigerenzer & Murray, 1987).

Thurstone also made no particular claims about the distribution of perceived stimulus intensities (or "discriminal processes") arising from discriminal dispersion, but rather claimed that the psychological scale should be defined such that the frequencies of those stimulus intensities are normally distributed. In this way, his theory is as much about construction of appropriate scales as it is about the cognitive processes involved in discrimination. Indeed, much of his work was concerned with attitudes and opinions rather than psychophysical stimuli, and how to appropriately scale a collection of stimuli based on a series of pairwise comparisons (Stevens, 1966). Discriminal dispersion is specifically defined as the standard deviation of discriminal processes on the psychological scale for a particular stimulus. Discriminal deviation is defined as the distance between the *modal* discriminal process and the discriminal process on any particular occasion. Finally, the separation between the modal discrimination processes for two stimuli is the distance assigned to those stimuli on the psychological scale. This leads to the full form of Thurstone's Law of Comparative Judgment (LCJ):

$$S_A - S_B = x_{AB} \cdot \sqrt{\sigma_A^2 + \sigma_B^2 - 2r\sigma_A\sigma_B}$$

where S_A and S_B are the psychological scale values of the two compared stimuli; x_{AB} is the sigma value from the cumulative normal distribution corresponding to the proportion of judgments $p_{A>B}$; σ_A and σ_B are the discriminal dispersions of the two stimuli; and r is the correlation between discriminal deviations for the two stimuli in the same judgment. Thurstone considers a number of cases in which various simplifying assumptions are made, of which Case V is the most commonly used. It is assumed that (i) the perceived relative values for two stimuli are normally distributed for a group of observers as well as a single observer, (ii) that the correlation between discriminal deviations for the two stimuli in the same judgment is zero (i.e. no contrast effects), and (iii) that all the discriminal dispersions are equal. This can be represented graphically as in Figure 1.2, reproduced from Gigerenzer and Murray (1987).

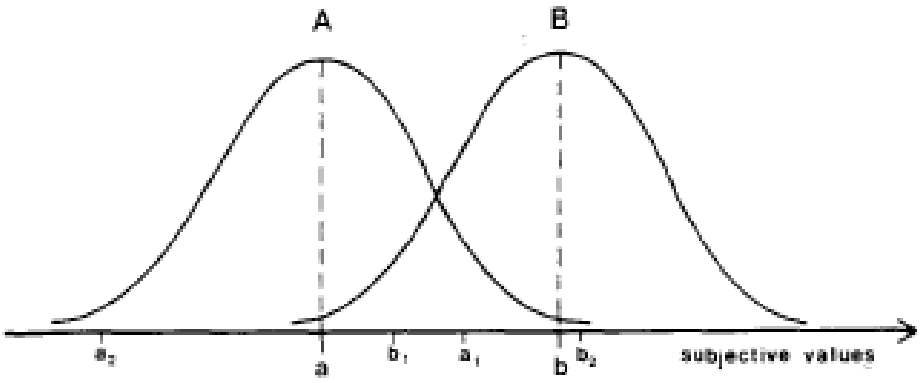

Figure 1.2: Illustration of Case V of Thurstone's Law of Comparative Judgment (Gigerenzer & Murray, 1987, p. 37, Figure 2.1).

Under these simplifying assumptions, and by making the appropriate (arbitrary) choice of unit scale, the LCJ simplifies to:

$S_A - S_B = x_{AB}$

It was the assumption of equal discriminal dispersion in particular which Stevens later rejected, although he noted that if Thurstone had conceived of a Case VI in which dispersion increased proportionally with intensity then he too would have arrived at a power function relating sensation magnitude to stimulus intensity (Stevens, 1966)

1.3.1.3 Signal Detection Theory

Both the "fixed threshold" and the "discriminal dispersion" theories view discrimination as an essentially passive response of the human judge to external stimulus differences or internal variation, whereas the emergence of Signal Detection Theory (SDT) in the 1950's reflected a broader trend of viewing the mind as making active inferential statistical judgments (Gigerenzer & Murray, 1987). Like the LCJ, the basis of SDT is two overlapping distributions as the internal representation of two signals, but it goes further by introducing the concept of a subjective decision criterion (Tanner & Swets, 1954). The two distributions are often assumed to be equal variance normal distributions, just as in Thurstone's Case V of the LCJ, but SDT can be applied to almost any distribution, for example a logistic probability distribution (Swets, 1986). The mathematics of SDT was originally developed during World War II to determine the optimal behaviour of radar operators and the language of SDT bears witness to this heritage: the two distributions are usually referred to as "noise" (*n*) and "signal plus noise" (*sn*) and the decision maker is assumed to be attempting to detect whether or not a signal is present against a background of random noise. The ideas are, however, just as applicable to the task of discrimination between two stimuli in the pairwise comparison tasks described earlier. If the reference stimulus A is labelled *n* and the second stimulus B is labelled *sn*, then the task of discrimination is analogous to detecting the presence of a signal. In order to decide whether or not a signal is present, the decision-maker must set a criterion value above which they will judge that a signal was present (the two stimuli were different) and below which they will judge that no signal was present (the two stimuli were the same). This can be represented graphically as in Figure 1.3, reproduced from Gigerenzer and Murray (1987).

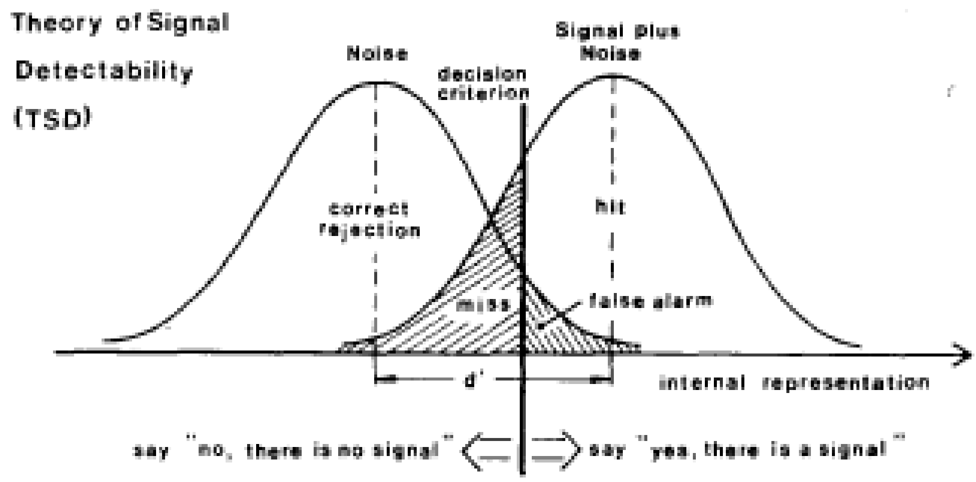

Figure 1.3: Illustration of Signal Detection Theory (Gigerenzer & Murray, 1987, p. 43, Figure 2.2).

In setting the decision criterion the decision-maker is making a trade-off between the hit rate (the probability that a signal is correctly identified as being present) and the false alarm rate (the probability that a signal is incorrectly identified as being present). If the criterion is set very high then very few false alarms will occur, but the hit rate will also be low. If the criterion is set very low then the hit rate will be high but the false alarm rate will also be high. The optimal decision criterion depends upon the relative utility of hits, misses, false alarms and correct rejections, i.e. it is both subjective and context dependent. One implication of SDT is that the hit rate and the false alarm rate are coupled. This is identical to the inferential statistics of Neyman and Pearson, in which the power of an experiment falls (the chance of Type II errors increases) as the probability of a Type I error is reduced (Gigerenzer & Murray, 1987). The relationship between the hit rate and the false alarm rate can be represented graphically in the form of a Receiver Operating Characteristic (ROC) curve. For a given pair of stimuli, an ROC curve represents the different combinations of hit rates (h) and false alarm rates (f) that occur as the decision criterion is varied. As the inter-stimulus distance (d') is increased or decreased then families of ROC curve are obtained. An example family of ROC curves for two equal-variance normal distributions is shown in Figure 1.4, reproduced from Swets (1986).

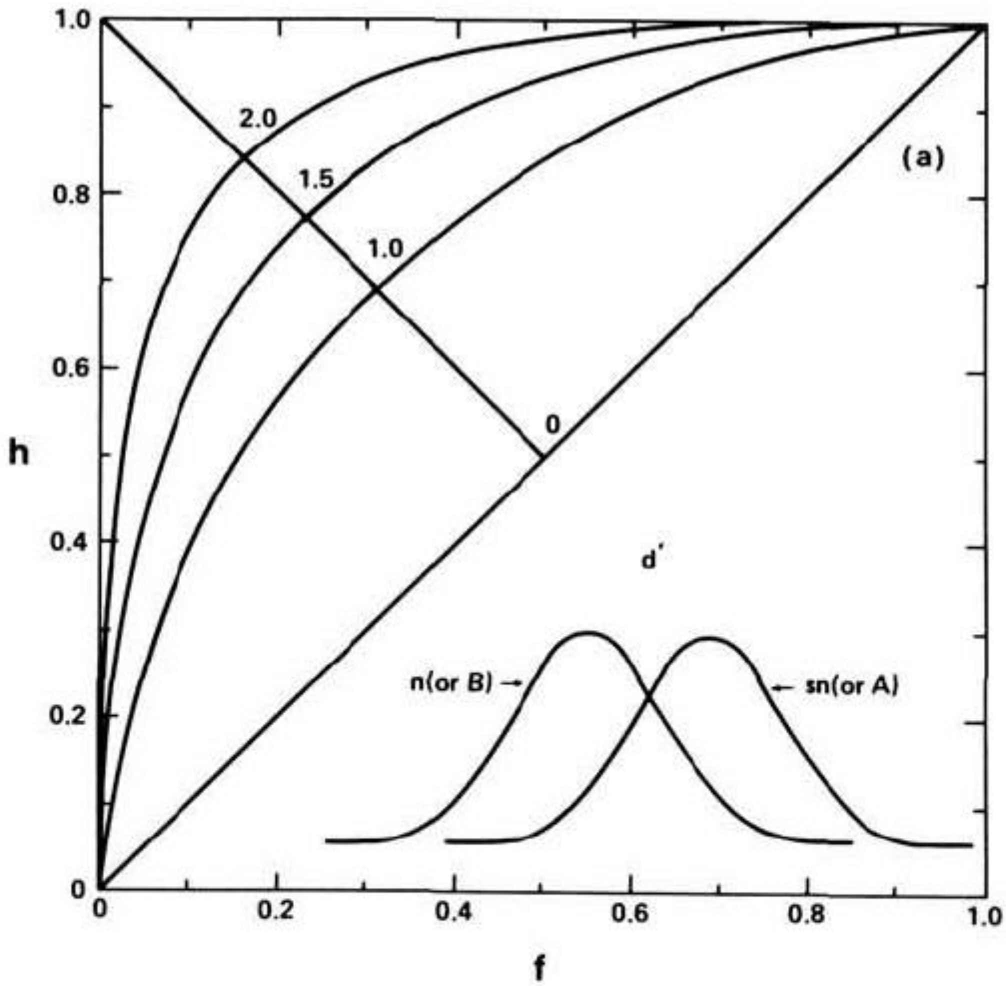

Figure 1.4: ROC curves for equal variance normal distributions as a function of the standardized inter-stimulus distance *d′* (Swets, 1986, p. 106, Figure 3).

One of the major contributions of SDT to the understanding of human discrimination performance is that it introduced the need to consider not only the observer's subjective sensitivity (*d′*) but also the decision criterion. Ideally, experiments should be designed to vary both *d′* and the decision criterion, and also to explore the factors that influence the setting of the decision criterion. In order to determine the sensitivity and the decision criterion, both hit rates and false alarm rates have to be collected as data, and ROC curves used to derive the required values. Most importantly for the topic of this thesis, SDT emphasizes that changes in decision context can cause changes in discrimination rates between two identical stimuli, because the decision criterion has changed. Specifically, if the relative utility of hits and false alarms differs, or is perceived to differ, between two different decision contexts then the probability of discriminating a difference between two constant stimuli will also vary. The importance of context for human judgments is a topic that I shall return to in more detail later in this chapter.

1.3.1.4 Attribution Theory

Judgment and discrimination is of course not restricted to the domain of sensory perception and psychophysics. Humans exist in a complex world of social interactions, requiring intuitive judgments concerning the motives and actions of other people. One of the dominant theories in social psychology – certainly the widest researched – is that of Attribution Theory, first developed in the late 1950s and early 1960s (reviewed by Kelley, 1973; Kelley & Michela, 1980). Attribution Theory is primarily a theory of causal judgments that are made in order to interpret other people's behaviour, but it is closely related to other theories of judgment and decision-making:

> "...Attribution Theory is related to a more general field that might be called *psychological epistemology*. This has to do with the processes by which man "knows" his world and, more importantly, *knows that he knows*, that is has a sense that his beliefs and judgments are veridical. The ascription of an attribute to an entity amounts to a

particular causal explanation of effects associated with that entity – reactions or responses to it, judgments and evaluations of it, etc. So all judgments of the type "Property X characterizes Entity Y" are viewed as causal attributions." (Kelley, 1973)

Attribution Theory is based on the idea that in order to make causal attributions about other people's behaviour, a human judge considers three relevant causal factors: persons (P), stimuli (S), and times (T). The attribution of a given person's response to a stimulus on a particular occasion is judged by (i) the *consensus* with other people's response to the same stimulus, (ii) the *consistency* with this person's response to the same stimulus on other occasions, and (iii) the *distinctiveness* from this person's response to other stimuli (Kelley & Michela, 1980). In other words, causal attribution is based on perceived covariance in the environment, between explanatory factors and the effect being judged, and the effect is judged as being caused by the factor with which it co-varies, or is perceived to co-vary. Where repeated observations are not available in order to judge covariance, judgments may be based on memories of similar judgments in the past or on counterfactual reasoning about non-common effects had the person acted in a different way, i.e. estimates of covariance based on past experience. This form of Attribution Theory is therefore sometimes described as the 'ANOVA model' because "the assumption is that the man in the street, the naive psychologist, uses a naive version of the method used in science" (Kelley, 1973).

Why is Attribution Theory of relevance to the kinds of intuitive statistical judgments considered in this thesis? Firstly, it raises the possibility that judgments of store prices may also be made on the basis of the perceived covariance of the price level of a store with other factors: "Store X is cheap because it shares Attribute Y with other stores which are also cheap". Some of those factors, which would have to be controlled for or measured in experimental research, are reviewed later in this chapter. Secondly, it raises the wider question of whether, and to what degree, humans have the ability to make accurate statistical judgments concerning the world around them, in this case to judge the covariance between a factor and an effect. Discrimination between complex stimuli, such as distributions of prices and distributions of inter-store differences in item prices, may also require intuitive judgments of statistical properties such as means, variances and covariances. In the next section I review the evidence related to intuitive statistical judgments, especially judgments of frequency information, and briefly discuss the implications of some of the apparent limitations and biases that have been found.

1.3.2 Intuitive Statistics

1.3.2.1 Intuitive Judgments of Statistical Properties

With the popularization of inferential statistics and Bayesian probability theory in psychology, the statistical methods of psychologists became increasingly viewed as a normative standard for 'correct' reasoning by those same researchers (Gigerenzer & Murray, 1987). Research was then directed at attempting to determine whether or not the human mind is able to calculate probabilities, means, variances and correlations, and whether intuitive reasoning follows the 'rational' laws of probability theory and inferential statistics. The main findings of this body of work were summarized by Peterson and Beach (1967), including both intuitive descriptive statistics (the process of describing samples of data) and intuitive inferential statistics (the process of inferring population statistics from samples). Of particular relevance is research related to people's ability to accurately judge the mean of a distribution. The first finding is that ‘average’ is variously interpreted by participants as one of several measures of central tendency, including mean, mode median and mid-range. However, if ‘mean’ is specified in the task instructions, then judgments are relatively accurate with no apparent biases. The variance between estimates of a mean tends to increase with the variance of the sample, with the sample size and with the speed of presentation, which suggests that participants are actually making inferences about the population mean (Peterson & Beach, 1967). Judgments of variance show two interesting effects. Firstly, they appear to be negatively correlated with the mean, suggesting that it is actually the coefficient of variation (the ratio of standard deviation to mean, related to the Weber fraction $\Delta I/I$) that is being estimated. Secondly, even once this is accounted for, judgments of variance appear susceptible to both underweighting and overweighting of large and small deviations from the mean, depending upon the task instructions and the distribution of values being judged. Instructions that emphasize small deviations or distributions with many small deviations (e.g. a normal distribution) tend to lead to underestimation of variance, whilst instructions that emphasize large deviations or distributions with prominent large deviations (e.g. a saddle-shaped distribution) tend to lead to overestimation of variance (Peterson & Beach, 1967).

Inferences of population averages (mean, median and mode) also appear to be relatively accurate, although when shown a J-shaped distribution – with many low values and a few high values - estimates of the mean are biased towards the median, i.e. the mean is underestimated (Peterson & Miller, 1964). This is equivalent to underweighting large deviations, leading Peterson and Beach to conjecture that “subjects may have regarded them as unrepresentative and thus not more important than the most frequently occurring events” (Peterson & Beach, 1967). Most of the inferential tasks involved showing numerical values on cards drawn from a pack and asking participants for inferences about the mean or variance of the pack. For example, Irwin, Smith and Mayfield (1956) drew cards one at a time from a pack of 500 cards and asked participants to judge whether the mean value of the pack was greater or less than zero using a confidence scale ranging from -100 (absolutely certain the mean is negative) to +100 (absolutely certain the mean is positive). The judgment was given after both 10 and 20 cards, although the second 10 cards were fixed to be identical to the first 10 cards, shuffled into a different order. The actual mean and variance of the packs were manipulated between different packs. Confidence ratings were directionally accurate, and participants were more confident when the actual mean of the pack was further from zero, when the variance of the pack was lower and when they had seen more cards.

However, a closer parallel to the inter-store price comparison task of Alba et al can be found in a second experiment involving two packs of cards (Irwin, et al., 1956). This time a card was drawn from each pack simultaneously and participants had to judge which pack had the higher mean value using a similar confidence scale to the previous experiment. Confidence ratings were obtained after each pair of cards was drawn, with 20 pairs of cards drawn in total. As before, the second 10 cards in each pack were the same as the first 10 cards, shuffled into a different order. Again confidence ratings were directionally accurate and participants were more confident when the actual difference in means between the two packs was greater, when the variance in each pack was smaller and when more cards had been seen. In every test the values were normally distributed and one pack always had a higher mean than the other, so there were no conditions which directly paralleled the price distributions used by Alba et al with identical means. Irwin, Smith and Mayfield also manipulated the mean and variance of each pack of cards separately, so did not measure or control the frequency of occasions on which the card drawn from each pack had a larger or smaller value than the other pack.

These kind of card-drawing experiments have subsequently been criticized because the judgment task was not truly ‘intuitive’, in the sense of knowing or learning something without conscious reasoning (Malmi & Samson, 1983). An alternative testing methodology was subsequently used that presented a large sample of numbers at high speed, e.g. 50 values for 0.5 seconds each. Furthermore, two different distributions of values, indicated with a category label, were randomly interleaved so that each participant would see 100 values in total, before being asked to estimate the average of each category. Even with extremely skewed distributions, estimates tended to reflect the mean rather than any other measure of central tendency. Estimate accuracy was unaffected by presentation time or by changing the differences of the mean and variance between the two distributions. Participants were also able to give accurate estimates for the variance and skew of each distribution, based upon estimates of the frequency with which different bands of values were presented. When asked, post hoc, to exclude values above or below a certain threshold and to re-estimate the average of the remaining values estimates were also reasonably accurate. This suggests that a sample of the observed values was represented in memory rather than the participants calculating a running mean as the items were presented and discarding the individual values from memory (Malmi & Samson, 1983).

1.3.2.2 Representations of Sets in the Visual System

Similar ‘intuitive’ experiments have been carried out using shapes and colours instead of numerical information, in order to understand how the visual system represents sets of items presented simultaneously or with a very short delay. In one example (Ariely, 2001), participants were shown sets of different sized circles for 500ms, followed by a test circle. In one task participants had to say whether the test circle had been a member of the set they had just viewed, while in a different task they had to say whether the test spot was larger or smaller than the mean of the set. The size of the sets and the similarity of the circle sizes were varied between trials. Participants were unable to perform the first task (set membership identification) any better than chance, but were highly accurate in discriminating the mean set size, with a 75% accuracy threshold of about 4-6% of circle size for similar sets and 6-12% of circle size for dissimilar sets. Performance did not differ between sets of different sizes. These findings suggest that the visual field represents statistical properties of sets, such as the mean and range, rather than the full descriptive details of each individual member of the set (Ariely, 2001). Subsequent studies employing mean size comparisons between two different sets of circles have found that judgments of mean size appear to use a value somewhere between the mean circle area and the mean circle diameter; that accuracy is unaffected by exposure duration or by simultaneous or sequential presentation of the sets; that accuracy is the same for different types of distribution and is only slightly worse when the shapes of the two distributions are different from each other (Chong & Treisman, 2003); that neither set size nor density affects accuracy; and that segregation of sets by colour gives mean discrimination thresholds that are no greater than when sets are segregated by location (Chong & Treisman, 2005).

Representation of mean set size by the visual system appears to be fast, accurate and automatic. Studies suggest that other properties of sets and sequences, including higher order temporal structure such as item frequencies and joint probabilities of consecutive appearances of item pairs, may also be represented automatically by the visual field (Fiser & Aslin, 2002).

1.3.2.3 Representation and Processing of Numerical Information

The experiments described in the two prior sections suggest distributions of numerical information are represented and processed differently from distributions of visual information: numerical information is retained in memory as individual values whilst visual information is represented by summary statistical descriptors. This modality difference is not particularly surprising: sensory information processing (visual, auditory, etc) is a primitive task, shared by our distant ancestors and most other species of animal, while numerical information processing is a relatively new and unique task in our evolutionary development. It is therefore plausible that visual information is represented and processed by specialized brain circuitry in a highly automatic way while numerical information is represented and processed by a more general cognitive architecture. Nonetheless, there is evidence that learning can result in automatic processing of basic numerical information, such as classifying digits as large or small (Tzelgov, Meyer, & Henik, 1992). Experimental findings suggests that the digit ‘5’ has a special status in numerical processing, being the exact halfway point in the number line from zero to ten. Digits smaller than five are automatically classified as ‘small’ whereas digits larger than five are classified as ‘large’. This automatic processing leads numerical size to interfere with judgments of physical size in a Stroop-like *size congruity* effect: when physical size and numerical size are inconsistent then reaction times for physical size judgments are increased (Tzelgov, et al., 1992). However, it also appears that the magnitude of these interference effects is malleable, for example by changing the discriminability of each dimension, suggesting that numerical processing is not strongly automatic because activation of numerical size judgments is not immune to changes in task demands and attention (Pansky & Algom, 1999). In a similar fashion, judgments of numerosity also interfere with judgments of numerical size, but the effects can be moderated by changes in dimension discriminability, motivation and practice (Pansky & Algom, 2002).

Research into numerical cognition has found a number of other effects which shed further light on how numbers are represented and processed in the human brain. One of the earliest and most robust findings is the *distance effect* in discrimination between two single-digit numbers (McCloskey & Macaruso, 1995). Discriminating between two digits is faster and easier when the two numbers are distant (e.g. 2 and 8) than when the two numbers are close (e.g. 7 and 8). A similar *problem size effect* is found in algebraic tasks, in which tasks involving larger digits (e.g. 7 + 8) take longer to perform than tasks involving smaller digits (e.g. 2 + 3) (Ashcraft, 1992). It is not yet clear whether numbers in different formats, such as Arabic numerals (1, 2, 3) or verbal numerals (one, two, three) written or spoken, are first transcoded into an internal semantic representation or whether the brain works simultaneously (asemantically) with representations in different formats. Proponents of the single-format position point out that children who learn an arithmetic fact in one format (e.g. they are told by a teacher that “eight times four equals thirty two”) are able to transfer that knowledge to arithmetic tasks in other formats (e.g. they can solve the written Arabic problem “4 × 8”). Proponents of the multi-format hypothesis point to evidence of differences in arithmetic fact retrieval performance depending upon the format of the task and also argue that cross-format learning of arithmetic facts is not complete (McCloskey & Macaruso, 1995). Other studies suggest that numerical information is represented spatially in the form of a number line, with small digits associated with the left hand end and large digits with the right hand end (Longo & Lourenco, 2007). Furthermore, it appears that larger numbers are less discriminable than smaller numbers (as if they obey Weber's Law) which has led to suggestions that the number line may be logarithmically compressive or that the variability of internal representations (discriminal dispersion in the language of Thurstone's LCJ) increases with the size of numbers. Evidence for the spatial nature of the mental number line include a correlation between individual biases in bisecting spatial and numerical lines, and the fact that number line bisection tasks are much harder when the larger of the two numbers is presented on the left (Longo & Lourenco, 2007).

It is not immediately clear how these numerical cognition findings could explain the frequency effect observed in inter-store price comparisons. On the one hand, a frequency store would contain a larger number of small prices with a few large prices, which could bias estimates of the mean price downwards if the number line is non-linear. On the other hand, the distance effect would suggest that the large price advantages of a magnitude store would be more discriminable (and perhaps therefore more salient) than the small price advantages of a frequency store. In any case, the research into intuitive statistical judgments described earlier suggests that real-world inter-store price comparisons are a cognitively demanding task and we should not expect non-intentional price judgments to be carried out by specialized brain circuitry in a fast and accurate manner. Rather, the discrimination task is likely to be biased, subject to influences such as changes in attention or motivation, and could also be influenced by potentially irrelevant cues that are easily and automatically processed. As suggested by Alba et al, one such candidate cue is the *frequency* of price advantages. In the next section I review the results of research into intuitive judgments of frequency, and evidence for the automatic processing of frequency information in the environment.

1.3.3 Frequency Judgments

1.3.3.1 Probability Learning

From the 1950s onwards, a large amount of research was dedicated to understanding human probability learning, with the growing belief that humans are information processors and decision makers, using informative feedback from their past actions rather than being shaped by the effect of rewards and punishments in the form of mechanical reinforcement learning (Estes, 1976). However, a large body of evidence suggests that in many circumstances decisions do not correspond to rational choice theory (Shanks, Tunney, & McCarthy, 2002). In a series of repeated decisions between two alternatives with different payoff probabilities, rational choice theory would predict that the higher probability alternative should be selected 100% of the time, after sufficient trials have been conducted in which the payoff probabilities can be learned. However, a large number of experiments found asymptotic choice percentages that were lower than predicted, often matching the payoff probabilities exactly. For example, if option A had a payoff probability of 70% and option B had a payoff probability of 30% then participants would choose option A on about 70% of trials: they appeared to be *probability matching*. More recent experiments have shown that this tendency can be reduced or removed through the use of large financial incentives, meaningful and regular feedback, and extensive training (Shanks, et al., 2002). Nonetheless, in many situations the tendency to probability match seems extremely common. More recent research has utilized multiple cue probability learning (MCPL) tasks, in which a number of cues (e.g. different medical symptoms) are presented and the participant has to judge which outcome will occur (e.g. disease A or B is present). Again, suboptimal responding and probability matching was often found, especially when no feedback was given (Estes, Campbell, Hatsopoulos, & Hurwitz, 1989; Shanks, 1990).

Probability learning research also suggests that probability estimates are based upon observed frequencies of events, which can lead to biases in those probability estimates. For example, Estes (1976) carried out an experiment in which participants were shown pairs of products as repeated observations. They were told that each observation represented a response from a consumer survey, and on each trial one of the products was indicated as being preferred. By manipulating the relative frequency of item appearance and item preference, Estes was able to directly pit the frequency cue against the observed probability. For example, a pair of items AB was presented 100 times and another pair of items CD was presented 200 times. Within the pairs, A had a 75% chance of being preferred over B whilst C and D each had a 50% chance of being preferred over the other. Thus C has been presented as the winning product on 100 observations while A has only been presented as the winning product on 75 occasions. If asked to judge the likely winner from a comparison between A and C, a judgment based on observed probability would favour A whilst a judgment based on observed frequency would favour C. In a series of experiments Estes showed that it was in fact the frequency cue that tended to dominate, leading to the conclusion that “probability estimates, relative frequency judgments, and predictive behaviour all share a common basis in associative memory” (Estes, 1976, p. 62).

1.3.3.2 Frequency Learning

A large body of research supports the idea that frequency information is accurately encoded and available for recall in judgment and probability estimation tasks. Reviews of frequency learning experiments have consistently reached the same conclusion. Peterson and Beach (1967) state that "the most striking aspect of the results is that the relation between mean estimates and sample proportions is described well by an identity function" while Howell (1973) concludes that "the main point of agreement in the experiments reviewed here is that subjects show a remarkable facility for synthesizing and storing the repetitive attribute of event occurrences". Furthermore, Hasher and Zacks (1984) state that "the major conclusion of this area of research stands on a firm empirical base: the encoding of frequency information is uninfluenced by most task and individual difference variables". There is some evidence of a mean reversion bias in some circumstances, in which low frequencies are overestimated and high frequencies are underestimated (Lichtenstein, Slovic, Fishhoff, Layman, & Combs, 1978; Peterson & Beach, 1967), although it has subsequently been argued that this does not necessarily reflect a systematic bias but can be explained in terms of unsystematic variance (Hertwig, Pachur, & Kurzenhäuser, 2005).

It has further been argued that the encoding of frequency information is automatic, i.e. that it is low effort, does not draw on limited attentional resources, and that it does not interfere with other on-going cognitive processes. In this respect it is a fundamental aspect of the flow of information encoded in memory, along with spatial and temporal location (Hasher & Zacks, 1979). One piece of evidence supporting this conclusion is that, unlike most other cognitive processes, it does not appear to show any developmental trends: children and young adults show the same ability as adults to accurately judge frequency of occurrence (Hasher & Chromiak, 1977). Other factors which usually influence cognitive performance, such as practice or feedback, similarly appear to have no impact on frequency encoding and recall (Hasher & Chromiak, 1977). Frequency information may also be implicated in a wide range of tasks and skills including memory for events, organizing existing knowledge and acquiring new knowledge, decision-making, and cognitive and social development (Hasher & Zacks, 1984).

1.3.3.3 Animal Foraging

Humans are not the only animals that are sensitive to frequencies of occurrence in the environment. Studies of animal foraging behaviour consider the degree to which animals are sensitive to probabilities (frequencies) in their environment. For example, bumblebees might be faced with two colours of flowers, blue and yellow. The blue flowers might yield a consistent small amount of nectar while the yellow flowers yield a much larger amount of nectar, but only a proportion of the flowers contain nectar. Experiments typically examine the degree to which animals are sensitive to the trade-off between mean and variance in different foraging options, and whether they are consistently risk-averse or risk-seeking. Sensitivity to mean and variance has been found in a wide range of species, including bumblebees, sparrows, bananaquits (a small nectar-eating bird), shrews, wasps, warblers, rats and pigeons (Real, 1991; Real & Caraco, 1986). The experimental evidence suggests that – just like humans – animals such as bumblebees estimate probabilities from experienced frequencies of events. However, this leads to a bias in the opposite direction from the mean reversion bias described earlier: bumblebees overweight high frequencies and underweight low frequencies. This may be because bumblebee’s probability judgments are subject to memory or perceptual constraints, but it has also been argued that this bias may be adaptive in certain (spatially autocorrelated) environments (Real, 1991). In any case, sensitivity to frequency information appears to be a cognitive ability shared by many species.

1.3.3.4 The Frequentist Hypothesis

The universality, automaticity and accuracy of frequency judgments have led some researchers to conclude that human statistical cognitive architecture is ‘tuned’ to information in a frequency format, because this is how it is naturally encountered. Probabilities and contingencies are learned through sequential encoding and updating of event frequencies. Statistical judgments are most accurate when performed on information presented in this natural format, or formats designed to invoke a frequency representation, rather than standard percentage descriptions of probabilities (Gigerenzer & Hoffrage, 1995). The ‘frequentist hypothesis’ is that “some of our inductive reasoning mechanisms do embody aspects of a calculus of probability, but they are designed to take frequency information as input and produce frequencies as output” (Cosmides & Tooby, 1996). When problems are presented in a frequency format, many previously observed biases in statistical judgment such as overconfidence, the conjunction fallacy and base-rate neglect disappear (Cosmides & Tooby, 1996). Furthermore, the frequency computation system appears to work best with representations of whole objects, events and locations, i.e. ‘natural units’ (Brase, Cosmides, & Tooby, 1998).

The frequentist hypothesis and the body of research concerning frequency judgments are important to the topic of this thesis for two reasons. Firstly, they highlight the importance of *ecological validity*: in order to understand how human decision makers perform a specific task, the experimental methodology must accurately recreate the format in which information is received in the real world and allow participants to respond in a naturalistic format. This point will be discussed in more detail later in this chapter. Secondly, if frequency information is readily available in the environment then we should not be surprised if it dominates intuitive statistical judgments. On the other hand, given that price information is presented in an 'unnatural' Arabic numeral format, subtle changes in task and information format which reduce the availability of frequency information might have a dramatic impact on the way price judgment tasks are performed. For example, in the Alba et al experiments participants were encouraged to pay attention to every single price whereas in real-world shopping tasks consumer are likely to sample information selectively from a much larger number of prices. Research suggesting that processes such as sampling and information search have strong implications for intuitive statistical judgments will be reviewed in the next section.

1.3.4 Naïve Intuitive Statistics

1.3.4.1 Naïve Information Sampling

For various reasons – time constraints, financial costs, cognitive or memory constraints – people do not usually sample the full population of information about which they want to make a judgment, but rely on samples of information. For example, when asked to estimate the average height of an adult male it would be impractical to try and recall the height of every adult male ever encountered. Instead a small sample of examples might be recalled and used to form an estimate. This is a case in which a sample statistic is used to estimate a population statistic. Samples might be drawn from memory or actively searched for. Sample means and sample proportions are unbiased estimates of their population values, but sample variance for a sample size of n is smaller than the population variance by a factor of $n/(n+1)$. Experimental evidence suggests that intuitive judgments of variability are not corrected for this bias in randomly-sampled information and that therefore humans perceive the world as less variable than it really is (Kareev, Arnon, & Horwitz-Zeliger, 2002). Because sample variability is a biased estimator but sample proportion is unbiased, changing the format of estimation tasks from one that encourages the use of variability estimates to one that encourages the use of proportions can remove systematic biases such as overconfidence in interval estimates (Winman, Hansson, & Juslin, 2004). As well as failing to correct for biases in sample data, people appear to treat sampled data as if it were randomly sampled and representative of the population. For example, the reported frequency of violent deaths reported in the media influences the judged risk of violent death, which is one explanation for why some low frequency events might be overweighted (Lichtenstein, et al., 1978). Similarly, sampling can also lead to *underweighting* of rare event probabilities in choice behaviour for two reasons. Firstly, rare events are often not encountered in small samples of information. Secondly, if recent information is weighted more strongly in estimation due to the updating process then common events will be weighted more strongly than rare events, because they are more likely to have occurred recently. This explains why underweighting of rare events is found in both human and animal studies, such as the bumblebee experiments cited earlier, which involve sampling of information from the environment (Hertwig, Barron, Weber, & Erev, 2004). Such findings have led to the suggestion that humans are ‘naïve intuitive statisticians’, in that cognitive processes accurately represent and process the available information but that people are naïve with respect to the origin and estimator properties of sampled information (Juslin, Winman, & Hansson, 2007).

The idea that cognitive processes accurately represent the available information has been supported by recent research concerning people's knowledge of the distribution of real-world phenomena such as life-spans, movie runtimes and gross takings, poem lengths, cake baking times and the length of reigns of pharaohs (Griffiths & Tenenbaum, 2006). A novel methodology for obtaining estimates of the shape of each distribution involved asking participants to make judgments of the most likely outcome given a piece of information. For example, participants were told than a man was currently aged t and were asked to estimate his most likely total lifespan t_{total}. A Bayesian statistician would apply Bayes' Rule to compute a probability distribution over t_{total} given t:

$$p(\, t_{\text{total}} \mid t \,) \propto p(\, t \mid t_{\text{total}} \,) \cdot p(\, t_{\text{total}} \,)$$

Assuming uniform random sampling, the first term is simply $1 / t_{\text{total}}$ for all values between 0 and t_{total}, while the second term reflects general expectations about the probability distribution of t_{total}. The relationship between t and predicted values of t_{total} are quite different for different forms of prior distribution (normal, power-law, Erlang etc) so participants' responses can be used to determine the most likely distribution of prior beliefs. Compared to real-world data on objective probability distributions, participants' median responses were extremely close to the optimal Bayesian response for a wide range of phenomena with quite different distributions. Griffiths and Tenenbaum conclude that their results "demonstrate that, at least for a range of everyday prediction tasks, people effectively adopt prior distributions that are accurately calibrated to the statistics of relevant events in the world" (Griffiths &

Tenenbaum, 2006, p. 772). Although the results were subsequently criticized because they used median responses aggregated across a large number of subjects – and hence may not have demonstrated that an individual participant held detailed information about the full prior distribution of an event – a follow-up study has shown that explanations relying on ‘the wisdom of crowds’ cannot explain all the observed data, and individuals appear to hold relatively sophisticated knowledge about everyday events (Lendanowsky, Griffiths, & Kalish, in press).

1.3.4.2 Information Integration Theory

If humans act as naïve intuitive statisticians, sampling information from their environment to make informed judgments, how do they integrate multiple pieces of information in order to make an overall judgment? One possible answer to that question is given by Information Integration Theory (N. H. Anderson, 1965, 1970, 1971). Information Integration Theory (IIT) has been used with varying success to explain findings from a range of fields including judgment and decision making, linguistics and social psychology. IIT assumes that a judgment involves three stages: a valuation stage in which each stimulus is mapped onto an internal interval scale; an integration stage where each of the subjective values is combined to form an overall impression; and a response stage in which the internal impression is translated into an overt response. Example valuation functions were given earlier in the description of Buyyukkurt’s (1986) experiment: individual item prices were valued by comparing them to an internal reference price. The integration function is usually assumed to be an additive or averaging process, with weighting often applied to account for serial order effects (primacy or recency) or other attentional differences between stimuli. The response function is usually assumed to be a linear mapping. Because IIT is not prescriptive about the exact format of the valuation and integration functions it offers a useful framework for modelling the integration of serially sampled information but it does not readily generate falsifiable predictions. Rather, it can only be used to compare different functional forms for the valuation and integration functions, assuming that IIT is an accurate description of how people integrate information from a range of stimuli.

1.3.4.3 Heuristics and Prior Beliefs

Time or resource constraints, low motivation or expertise, and task complexity may not only encourage naïve statistical judgments to be based on limited samples of information, but might also encourage the use of naïve decision-making heuristics. For example, judgments of frequency and probability may be based on recall of information from memory. Rather than recalling lots of events from memory, people sometimes rely on the relative ease with which different instances come to mind, the so-called *availability heuristic* (Tversky & Kahneman, 1973). Whilst availability is usually correlated with environmental frequency, it is also influenced by other factors such as familiarity, salience, and imaginability, which can cause systematic biases in judgments based on availability. For example, when people judge the frequency of different lethal events they not only overweight rare events and underweight common events, as described previously, but also overweight certain specific events such as motor vehicle accidents, homicides, floods and tornadoes. These are all events commonly reported by the media, so may be both salient and over-represented in memory (Lichtenstein, et al., 1978).

Another judgment heuristic is the *representativeness heuristic*: judging probabilities based on the degree to which an item resembles the class it belongs to (Tversky & Kahneman, 1974). For example, a sequence of coin tosses HTHTTH is judged more probable than HHHTTT (which appears to be non-random) and HHHHTH (which does not appear to represent a fair coin). Yet another is the *anchor-and-adjust heuristic* which describes people's tendency to make estimates by starting from an initial value – often suggested by the problem or the result of a partial calculation – and adjust (usually insufficiently) to reach their final answer (Tversky & Kahneman, 1974). Thus different 'anchors' yield different estimates, which are biased towards the initial value. For example, people estimating the product $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ tend to give a higher estimate than people who estimate the product $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$, because their initial partial calculation from the first few digits is higher. Finally, people may simply fall back on their prior beliefs when a judgment task is too cognitively demanding, especially in a complex multivariate environment or when they are required to make a prediction (Broniarczyk & Alba, 1994).

Of particular relevance to this thesis is a decision-making heuristic first identified by Dawes (1979) and sometimes referred to as Dawes' Rule. This heuristic applies to the choice between two or more options that differ on a number of attributes. The rational way to approach such a choice is described by Multi-Attribute Utility Theory and involves making a linear combination of the impact of each attribute, weighted by the attributes' relative importance or utility. Dawes described a simplifying heuristic in which each attribute is equally-weighted and the magnitude of attribute differences is ignored, i.e. the decision-maker chooses the option with the highest frequency of attribute advantages over the alternatives. This heuristic directly parallels the explanation given by Alba et al (1994) for the frequency effect they observed, although Dawes' Rule utilizes the frequency of advantages across multiple attributes rather than multiple instances of a single attribute, such as item prices.

1.3.4.4 System I and System II

The bodies of research results on frequency judgments, intuitive statistics, and heuristics and biases appear somewhat contradictory. On the one hand, frequency estimates and other statistical judgments can be fast and accurate, especially if information is presented to the visual field in naturalistic units or is serially experienced. Biases in judgments appear to be due to informational biases rather than processing errors: frequency estimates for rare events can be biased downwards by limited search or biased upwards by availability in memory. On the other hand, numerical processing appears to be deliberate and difficult, and often inaccurate. Heuristics and prior beliefs can systematically bias statistical judgments, even in experts. One proposed solution to such conundrums is that people have two different cognitive systems with which they make judgments, corresponding to intuition and reasoning (Kahneman, 2003). The two systems, also sometimes referred to as System 1 and System 2 (Stanovich & West, 2000), have quite different properties. Intuition (System 1) is - like perception - fast, effortless, and implicit while reasoning (System 2) is slow, controlled and effortful. Unlike perception, which acts only on current stimuli, both intuition and reasoning can act on conceptual representations including ideas evoked by language. The principal differences between the different cognitive systems are summarized in Figure 1.5, reproduced from Kahneman (2003).

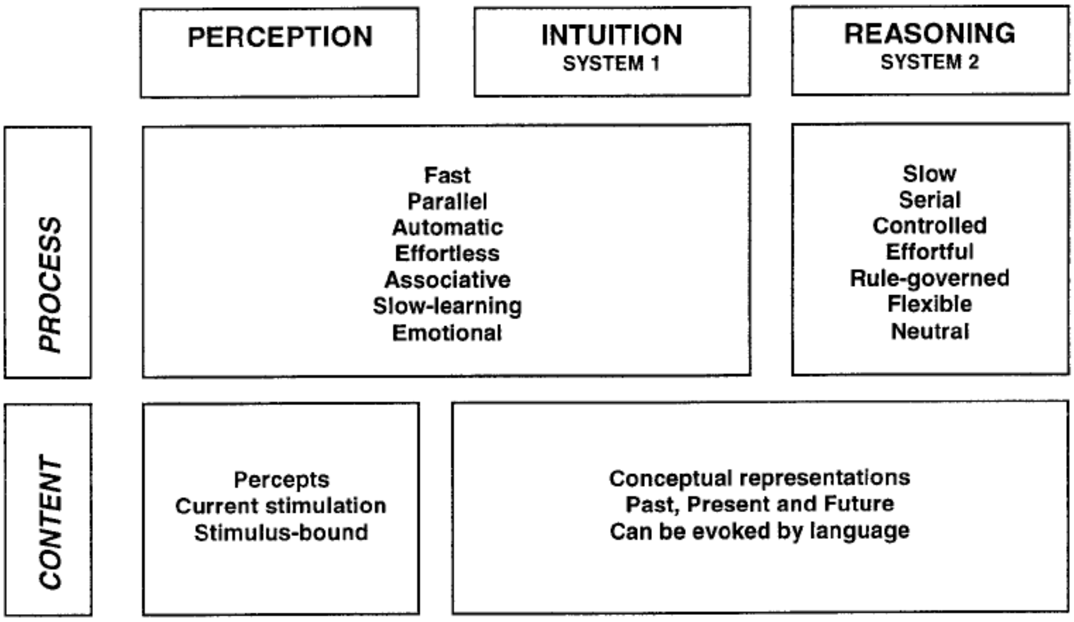

Figure 1.5: The three cognitive systems involved in judgments and decision making (Kahneman, 2003, p. 698, Figure 1).

There is considerable evidence supporting the idea of two cognitive systems for intuition and reasoning. For example, using familiarity as a basis for recognition judgments is believed to be an automatic use of memory (System 1) but recollection is believed to be intentional (System 2). Recollection is hampered when attention is divided, whereas recognition performance is invariant across full or divided attention (Jacoby, 1991). Similarly, dividing attention during a task hampers later recollection whilst leaving automatic memory processes unchanged, but switching task modality between study and test removes automatic memory influences whilst leaving deliberate recollection unaffected (Jacoby, 1996). Everyday life also abounds with examples of two modes of thinking: so-called ‘conflicts between head and heart’; the appeal of narratives and pictures over informative, dry texts; the persistence of superstitions and irrational fears; and possibly even the ubiquity of religion (Epstein, 1994). The key concept defining whether a judgment is made intuitively or deliberately is the degree to which the relevant concepts are *accessible* in memory, a term which includes notions of salience, attention, training, activation, and priming (Kahneman, 2003). While there is yet no general theoretical account of the determinants of accessibility, further consideration of the role of memory in judgments is merited in order to understand the part memory might play in the comparative judgments considered in this thesis.

1.4 Memory

1.4.1 Memory Architecture

While there is considerable disagreement on the exact nature of the architecture of memory and cognitive systems which access memory, certain key features are common to most memory models, in particular the need for a number of different memory ‘stores’ (Cowan, 1988). The earliest models proposed three stores: a sensory store for un-analysed information, a limited-capacity short-term store in which selected information is held for further coding, and a long-term store composing knowledge. In addition, various control processes or a *central executive* may be required to manage the voluntary transfer of information and to switch attention. Within this architecture, further divisions have been proposed, for example the division of long-term memory into episodic memory (memory for particular events) and semantic memory (memory for abstract knowledge and concepts). It has also been suggested that the sensory memory store consists of a visual sensory store lasting a few hundred milliseconds and an auditory sensory store lasting several seconds. The details of such divisions are not relevant to understanding the role of memory in intuitive statistical judgments so will not be considered further here.

The short-term memory store is often referred to as ‘working memory’ because it provides an interface between perception, long-term memory, and action (Baddeley, 2003). Working memory too appears to have subdivisions, in particular a *phonological loop* for processing sound and language and a *visuospatial sketchpad*. The phonological loop can hold memory traces for a few seconds before they fade, although rehearsal can refresh those traces. Similarly, the visuospatial sketchpad also appears to be limited in capacity, typically to about three or four objects. Finally, the central executive processes in working memory can also be subdivided into two forms of control: habit patterns and schemas, implicitly guided by environmental cues, and an attentionally-limited supervisory activating system (SAS) that intervenes when routine control is insufficient. More recently, a fourth component of working memory, the *episodic buffer*, has been proposed. This buffer is a limited-capacity store that binds together information into integrated episodes. This differs from previous conceptions of working memory merely activating existing episodic memories in the long-term memory store because it emphasizes the ability of working memory to manipulate and create new representations.

1.4.2 Memory Storage and Retrieval

Within the memory architecture described in the prior section, it is also necessary to consider what form a memory takes and how it might be stored in or retrieved from a memory store. Some of the earliest models focused on working memory because of the relative ease with which memory limitations could be tested in an experimental setting. For example, experimenters observed a serial order effect in memory for short lists of stimuli. If a participant is presented with a list of items, e.g. numerical digits, followed by a test stimulus then recognition performance differs according to the serial position of the stimulus within the list. In general, participants are more confident in their recognition judgments for later items (a recency effect) but participants also have a good memory for the first item presented (a primacy effect). Wicklegren and Norman fitted different models in which each number presented created a memory 'trace', an activation corresponding to the appropriate point on the number line (Wickelgren & Norman, 1966). Over time, as additional numbers were presented, the earlier traces gradually decay, leaving a weaker trace. When the test item is presented it also generates an activation, which is compared to each of the memory traces using a Signal Detection Theory decision criterion. If no match is observed then the participant responds that the test stimulus was not in the list. The best fitting model indicated (i) that memory traces decay exponentially, (ii) that participants only report a match if memory strength exceeds the decision criterion, and (iii) that the primacy effect is due to the initial memory trace being stronger than subsequent traces, rather than decaying more slowly or leaving an additional trace in long-term memory. Thus, memories are modelled as activations at a specific location with normally distributed noise, analogous to the normally-distributed activation of a stimulus in the LCJ or SDT.

Subsequent research into episodic long-term memory persisted with the idea that a memory is an activation trace in a multi-dimensional space. Recognition occurs when the activation caused by a cue stimulus is matched with an existing memory trace, along the lines of SDT. Recall follows a similar process, but the cue is not the stimulus itself but other information associated with the memory, such as contextual cues. One important finding was that the context in which a memory is encoded influences the subsequent ability for a particular cue to trigger recall, known as the *encoding specificity principle* (Tulving & Thomson, 1973). For instance, the word 'table' can aid in the recall of the previously presented word 'chair' when the word was encoded on its own, but not when the word was encoded in a list with other words. When ‘chair’ is encoded on its own then the semantic association with related words may also be encoded, whereas when it is encoded with other words in a list then these new associations are stored instead. Some models treat memory as a large array of traces, in which retrieval involves a probabilistic search of the memory space using cues (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981). These models are able to explain counter-intuitive results in free recall experiments, such as the ‘part-list cuing effect’. When presented with a list of words and later asked to recall as many words as possible, participants who are presented with a selection of words from the list as prompts perform worse at recalling the remaining words than other participants without prompts. The search models explain this as a result of the search cues used by each set of participants and the additional associations between the prompt words created by re-presenting them. The participants with prompt words utilize these as cues, and as a result frequently recall other words from the prompt list. The other participants are free to use other cues such as contextual information, so are more likely to recall the other words from the original list. Other models treat long-term memory as a distributed system in which all items and events are random vectors, so search does not require the identification of the appropriate localized portion of memory space (Murdock, 1982).

More recent models, again based on a memory store consisting of a vast number of memory traces, assume that retrieval is based on the summed response of all memory traces to a cue or ‘probe’, each of which depends upon the similarity between the trace and the probe (Hintzman, 1988). If a secondary *intertrace resonance* feature is added, in which traces activated by a probe can pass activation to other related traces, then such models can account for experimentally observed effects in frequency judgment, such as the tendency for presentation frequency judgments to be higher when the test word is presented in the same context as it was originally presented, relative to the case when the test word is presented in a different context. Similarity relations among multiple memory traces are also found in exemplar categorization models, which attempt to explain how items are perceptually grouped and the perceived similarity between items in different contexts. In contrast to prototype models which assume that category information is stored in an abstract summary category representation, exemplar models are able to explain context-dependency in norm construction (Kahneman & Miller, 1986), face categorization (Lamberts, 1994), and a wide range of other perceptual classification phenomena (Nosofsky & Johansen, 2000).

1.4.3 Rational Analysis of Memory

The view of long-term memory as a vast store of memory traces which has to be searched through in the retrieval process does not account for search costs such as retrieval time or metabolic expenditure. However, with a non-zero cost to memory search, an adaptive or 'rationally-designed' memory system would stop searching when the probable gain from a successful search is outweighed by the costs of that search (J. R. Anderson & Milson, 1989). This assumes that knowledge structures are ordered in order of plausibility and that memory search ignores structures below a certain threshold plausibility, although search will be more extensive when the gain is high (an important task) or the costs are low. Rational analysis of memory is based on the idea that "a system that is faced with the same statistics of information usage as a library or a file system and that is optimized [...] will produce the basic human memory functions" (J. R. Anderson & Milson, 1989, p. 705). From this perspective, memory is viewed as being optimized to maximize the probability of recall at the lowest cost, given a typical pattern of information retrieval demands. For example, as described earlier, memory recall is better for items that have been seen more frequently or more recently. Similar patterns are observed in real environments – words which appeared in the New York Times more often in a 100 day period were more likely to appear again on the 101st day, and words which appeared less recently within the last 100 days are less likely to appear on the 101st day (J. R. Anderson & Schooler, 1991). Rational analysis argues that human memory mirrors the information structure in the environment, and that memory is therefore adaptive (Schooler & Anderson, 1997).

In summary, human memory consists of a number of stores, and judgments require information to be retrieved from long-term memory in order to be used in working memory. Long-term memory appears to consist of a large number of memory traces, which include contextual information dependent on how they are encoded. Retrieval processes, such as recognition and recall, involve the comparison of a probe to each memory trace. The similarity between the probe and each trace, perhaps based on a SDT-type decision criterion, determines retrieval performance. Comparative price judgments that rely on recalled price information will therefore be influenced both by attentional biases and contextual factors when price information is encoded, and also by the specific retrieval cue used and the retrieval context. Because memory storage and retrieval is an imperfect process, price judgments from memory will be inherently different from price judgments made only using information currently available. Although Rational Analysis argues that memory retrieval is adapted to the environment, the degree to which memory is adapted to a particular task depends upon the specific task environment. Hence, to properly assess performance on a price judgment task in an experimental setting, the experimental methodology must faithfully recreate the basic information structure and memory demands of the real world. Certain theoretical perspectives emphasize the importance of task and environment in psychological research for other reasons. In the next section I briefly review some of the other arguments in favour of ecologically-valid research designs.

1.5 Task and Environment

1.5.1 The Importance of Context

One of the earliest proponents of the importance of the environment in psychological research was Brunswik, who spent much of his early career considering phenomena of *perceptual constancy* (Brunswik, 1937). One example is body-size constancy, the empirical observation that “the unconstrained observer will find it easy and natural to perceive and to compare bodies satisfactorily with respect to their own measurable physical sizes, regardless of all changes in distance or spatial orientation” (Brunswik, 1937, p. 228). He argued that because the perceptual system has no information except that which it receives from the environment (e.g. the light received on the retina) then the translation from received stimuli to perception must utilize information contained in the environment. His explanation was that the perceptual system integrates individual elements of a stimulus into a functional whole and combines that information with additional indirect stimulus effects, such as distance cues. He used the metaphor of a lens in which marginal rays are collected and converge, allowing the location of the radiating point to be determined. His ‘objective science’ was concerned with understanding the causal links between environmental cues and perception, and the focus of ‘constancy research’ was on measuring and correlating reactions with environmental traits.

Brunswik later moved onto studying behaviour, but his emphasis on the interaction between organism and environment persisted. He rejected the idea of universal laws of behaviour, arguing that they only apply within the context of the relevant *ecology*, which he defined as "the natural-cultural habitat of an individual or group" (Brunswik, 1955, p. 198). Because the ecology is semi-erratic (i.e. random but consisting of constant causal relationships) then psychology is inherently probabilistic. Furthermore, Brunswik argued that psychological research should adopt a *representative design*, in which the emphasis is on sampling from an ecology rather than controlled factorial experiments.

The view that perception is intrinsically linked to the environment was taken to its most extreme conclusion by the so-called *realists*, such as Gibson (Gigerenzer & Murray, 1987). Gibson addressed the question of whether we learn to perceive and to what degree, or whether perception is innate. Whilst Brunswik claimed that perception was inherently probabilistic, requiring the organism to make intuitive judgments from the available cues, Gibson argued that there was no need to assume a difference between sensory input and perception, but instead that "the stimulus input contains within it everything that the percept has" (Gibson & Gibson, 1955, p. 34). Perceptual learning was explained not as improved inference due to building up a store of relevant memories but rather as improved discrimination between stimuli: learning to differentiate between more subtle differences in the energy received by sensory receptors. Thus, just as Brunswik concluded, Gibson also rejected the traditional experimental approach to studying perception, claiming instead that "the laboratory *must* be like life!" (Gigerenzer & Murray, 1987, p. 84). More recent judgment research provides support for the Brunswik-ian notion of probabilistic inference from environmental cues, even when those cues are non-diagnostic in the experimental setting (Josephs, Giesler, & Silvera, 1994).

1.5.2 Information Format

As mentioned earlier in the discussion of the frequentist hypothesis, changing the format in which information is presented and the required output format of a task can have a significant impact on participants' judgments. Presenting information in a frequency format appears to eliminate biases that are consistently observed when judgments are based on information presented in a mathematically-identical percentage format (Gigerenzer & Hoffrage, 1995). In addition to the information format, the way in which that information is experienced also influences subsequent judgments and behaviour. For example, Prospect Theory, a theory of decision-making under risk, assumes that outcomes are weighed relative to a reference point and that probabilities are transformed into decision weights (Kahneman & Tversky, 1979). Based upon a previously-observed *certainty effect*, in which people underweight outcomes that are merely probable in favour of certain outcomes, the decision weights assume that people overweight low probabilities and underweight high probabilities. A later version of the theory, Cumulative Prospect Theory, added cumulative decision weights and allowed for different weights for gains and losses (Tversky & Kahneman, 1992). The model shows a 'four-fold pattern of risk attitudes': risk-aversion for gains and risk-seeking for losses of high probability; risk-seeking for gains and risk-aversion for losses of low probability. Critically, the models were based on experimental findings from tasks where risky or uncertain options ('prospects') were presented descriptively, e.g. "25% chance to win $150 and 75% chance to win $50". In contrast, in experimental paradigms where participants have to learn the underlying probabilities and payoffs from feedback based on their previous actions ('small feedback-based decisions'), participants are risk-seeking for gains and risk-averse for losses, and they underweight low probabilities (Barron & Erev, 2003). As explained earlier, this is probably due to reliance on small samples and over-weighting recent information (Hertwig, et al., 2004). Furthermore, feedback in experiential learning biases people towards sampling apparently favourable options and away from sampling apparently unfavourable options, which can also lead to risk-seeking or risk-aversion without requiring any assumptions about the marginal utility of losses and gains (March, 1996). Recent evidence suggests that risk aversion is more closely related to the Coefficient of Variation (CV) than to the variance of different options when information is sampled sequentially, which may be ecologically adaptive given the prevalence of Zipf (J-shaped) distributions in real-world phenomena (Weber, Shafir, & Blais, 2004).

1.5.3 Task Design

In order to study intuitive statistical judgments experimentally, it is desirable to place participants in a situation in which they actually make intuitive judgments rather than attempting to calculate the answer, guess, or follow some other cognitive process than they would use outside of the laboratory. This is especially true of intuitive statistical judgments where untrained or mathematically-naïve participants may not understand terms such as 'mean' or 'variance'. This is also highly important for certain types of task, particularly tasks involving memory for everyday events or tasks given to young children who would be unable to understand more abstract instructions (Neisser, 1991). An excellent example of a naturalistic statistical judgment task was used by Hutchinson and Alba (1997) to determine the effects of context on intuitive judgments of correlation. One task involved judging the partial correlations between four columns of input variables and an outcome variable. The input variables were described as advertising expenditures for four different products and the outcome variable was described as total store revenue, with each row labelled as a different year. Participants were asked to use the information in the table to decide how best to allocate an advertising budget across the four products. The optimal allocation would involve assigning the budget depending upon the relative correlation between each input variable and the outcome (as the input variables were uncorrelated) so the budget allocations could be used as a proxy for correlation estimates. In fact, it turned out that various task manipulations that left the correlations unchanged had a significant impact on intuitive assessments of correlation. Rising or falling trends over time made correlation more salient than saw-tooth patterns when the rows were labelled as years, but not when the rows were labelled as different stores. Adding a constant to one of the input variables or multiplying it by a constant increased the perceived strength of correlation. The experimental results suggest that a number of different decision heuristics were used to make the judgments, rather than a single automatic process, which is perhaps not surprising given the presentation of information in numerical tabular format. Hence the findings could be reasonably extrapolated to other correlation assessment tasks presented in a similar format, but not to tasks where the information was experienced serially or selectively sampled. Following the same logic, inter-store price comparison tasks such as Alba et al (1994) represent a naturalistic task for assessing participants' ability to discriminate between the means of two paired distributions, but only in contexts where all the information is presented simultaneously in a written numerical format. While such naturalistic tasks are not common in many areas of experimental psychology, they are the norm in applied consumer research. In the next section I briefly review some of the key findings from prior research into price judgments from the marketing and consumer research literature.

1.6 Consumer Price Research

1.6.1 Subjective Perceptions of Price

Price is central to classical economic theory, acting as a signal of the relative levels of supply and demand for a good and the costs faced by each supplier. If supply is greater than demand than suppliers must lower prices in order to clear their stock, while if demand is greater than supply then suppliers can raise prices to reduce demand. The market price represents the market-clearing balance between supply and demand. Under perfect competition suppliers continually undercut each other's prices to capture market share, until the point where marginal revenue is equal to the marginal cost of production. Within this framework, consumers are assumed to have perfect knowledge of prices and to make consistent utility-maximizing choices, such as always choosing the lowest price provider of a good. In reality, price perceptions – sometimes referred to as *price image* - differ significantly from objective prices and this can lead to consumers making sub-optimal choices between goods and stores (Brown, 1969). In an early study of price perceptions, Brown used ordinal store price judgments to place stores on an interval scale (using Thurstone's LCJ) and then correlated price perceptions in five different cities with objective prices based on a market basket of 80 commonly purchased products. Although shoppers were generally in agreement on the ranking of stores by price level (rank correlation of 0.92) the correlation between price image and reality varied between 0.98 in the best case and 0.00 in the worst case, indicating significant discrepancy between perception and reality. Psychological theories have commonly been employed to explain such phenomena. For example, some early price studies indicated that shoppers were insensitive to item price differences that fall within a threshold, which theorists attributed to a Weber-Fechner just-noticeable difference in prices (Monroe, 1971a, 1971b, 1973). Others argued that prices are judged relative to a reference price, much as in Prospect Theory, although the exact nature of the reference price was unclear, viewed as a blend of aspiration prices (e.g. "the price I would like to pay"), market prices (e.g. "the average retail price"), and historical prices (e.g. "the last price I paid") (Klein & Oglethorpe, 1987). In the following sections I survey some of the main findings from price perception research, beginning with price knowledge and search, followed by perceptions of item prices and ending with perceptions of store prices, the main topic of this thesis.

1.6.2 Price Knowledge and Search

1.6.2.1 Price Knowledge

The first studies into the accuracy of consumers' price knowledge showed mixed results, with higher accuracy often associated with methodological shortcomings such as bias in item selection, although they typically indicated relatively poor price knowledge that deteriorated rapidly with time from the moment of purchase (Dickson & Sawyer, 1986). Price awareness, measured by the ability to accurately recall prices, differs across product categories, stores and shoppers, with awareness of prices for store-brand products being better than for branded items (McGoldrick & Marks, 1987). Even when shoppers are intercepted moments after choosing a product from the shelf, price recall is poor with 21% of shoppers unable to even estimate the item price and less than half of shoppers able to give the exact price (Dickson & Sawyer, 1990). Socio-demographic differences in price knowledge vary from study to study, but typically lower-income customers have better price knowledge (Wakefield & Inman, 1993). Price awareness is generally better for discount stores with smaller ranges but even shoppers in these stores are relatively unaware of specific item prices (McGoldrick, Betts, & Wilson, 1999). However, store price knowledge was found to be accurate when rival retailers actively competed on price and heavily advertised discounts (Seiders & Costley, 1994). Over the longer term, consumers appear to underestimate the effects of price inflation, with recent prices being under-estimated and past prices being over-estimated (Kemp & Willetts, 1996). When a new currency is introduced, the new prices are learnt fastest for goods and services where prices are directly proportional to quantity (Juliusson, Gamble, & Garling, 2005).

Although consumers' ability to accurately recall item prices is poor, other measures of price knowledge show greater accuracy. Relative price knowledge, measured by the ability to correctly rank order brands from lowest to highest price, is much more accurate and only weakly correlated with the ability to recall specific prices (Connover, 1986). Meta-analysis of a large number of studies of consumer price knowledge indicates that a significant proportion of variation in price knowledge accuracy is driven by differences in research design characteristics such as the presence of financial incentives, task size, and the price elicitation method (Estelami & Lehmann, 2001). Some researchers, drawing on findings from research into memory and learning, have argued that poor price recall does not necessarily indicate poor price knowledge (Monroe & Lee, 1999) and that recognition tasks may be more appropriate than recall tasks (Monroe, Powell, & Choudhury, 1986). Recent studies employing recognition measures indicate that recognition of regular prices is poor, with less than 15% of respondents able to recognize the correct price from a range of prices within 5% of the actual price. However, deal recognition is very accurate, with over 85% of respondents able to discriminate between regular selling prices and prices reduced by 20% (Vanhuele & Dreze, 2002).

1.6.2.2 Price Search

One potential explanation for poor price knowledge is a low level of price search. Within the grocery supermarket context, consumers report very low levels of price checking, with only 58% of consumers claiming to have checked the price of an item before placing it in their basket and only 22% checking the price of an alternative brand (Dickson & Sawyer, 1990). Shoppers who deliberately set out to learn prices are better at price recall than relative price ranking, while those who acquire price information incidentally are better at relative price ranking and poor at recalling specific prices (Mazumdar & Monroe, 1990). Inter-store price comparisons improve price recall accuracy, but increase price recall confidence judgments to a greater extent, suggesting that consumers are over-confident in the amount that price search aids them in recalling prices (Alba & Hutchinson, 2000; Magi & Julander, 2005; Mazumdar & Monroe, 1992). The duration of price search is longer for products with a higher base price, but exhibits significant variation between individual shoppers (Oliveira-Castro, 2003). In conclusion, price knowledge appears to be poor for everything except significant discounts and most consumers do not spontaneously invest a lot of time or resources in searching out price information. Thus it is essential that experimental price perception research does not artificially encourage participants to pay attention to price information to a greater extent than they do in the real world.

1.6.3 *Item Price Perceptions*

1.6.3.1 Reference Price Research

Research into reference prices and reference price effects represents a growing body of literature in consumer behaviour, which has been thoroughly reviewed by Biswas, Wilson and Licata (1993) and more recently by Mazumdar, Raj and Sina (2005). A reference price is defined as a standard against which the purchase price of a product is judged, which can be external or internal to memory (Biswas & Blair, 1991; Monroe, 1973). An external reference price (ERP) is one which exists in the environment, either provided by a retailer in advertising or at the point of sale, or in the price of a comparable product being sold in the same location. An internal reference price (IRP) is a predictive price expectation against which the consumer judges the actual purchase price of an item in a particular context. Various conceptualizations of IRP have been suggested, but it is likely that IRP is not the same for all consumers, and it may not even be the same for a single consumer over multiple purchases and over time. In the following sections I shall briefly outline some of the main findings from consumer research into contextual ERPs, advertised ERPs, IRPs, and other influences on item price perceptions.

1.6.3.2 Contextual External Reference Prices

One of the earliest experimental studies into the effect of contextual ERPs provided participants with a range of prices for trousers, which had to be categorized into different categories of acceptability, with the participants free to choose the number and size of each category (Monroe, Dellabitta, & Downey, 1977). Half the participants were provided with a set of prices which ranged from $1 to $25.50 (short price series) and the other half were given prices ranging from $1 to $50 (long price series). There was no difference between the short and long price series groups in the number of categories formed or the lower limit of acceptability, but the long price series group had a significantly higher upper limit for acceptable prices. A later experimental study presented participants with a target mobile phone priced at \$159.65, in either a high price context (two other phones at \$197.85 and \$184.75) or a low price context (two other phones at \$127.65 and \$134.75) (Adaval & Monroe, 2002). Participants were later asked to compare it with a new mobile phone priced at \$157.89. The results showed that participants judged both the target and the new phone to be less expensive in the high price context, even though their subsequent recollection of the price of the target phone was also higher in this condition. Interestingly this priming effect was found even when a contextual price (low or high) was instead provided subliminally as part of a visual perception experiment. Fitting a model of reference price effects to empirical data on 42,000 purchases of saltines (crackers) over a two year period showed effects of both previously observed prices and current contextual prices in consumers' brand choices (Rajendran & Tellis, 1994). Of all the possible contextual ERPs, the lowest price seen appeared to be the most important cue for estimating a reference price, and there was some evidence that contextual prices had more influence on customers who shopped infrequently for that product.

An explanation for such contextual ERP effects is provided by range-frequency theory, an exemplar-based model of categorization in which an item is classified based on (i) its value relative to the minimum and maximum of all exemplars (range) and (ii) its rank position in all exemplars (frequency) (Parducci, 1965). In an experimental test of range-frequency effects in price judgments, a price of \$1.25 was judged in the context of a low range of prices (\$0.75 to \$1.50), a moderate range of prices ($0.75 to $1.75), or a high range of prices ($1.00 to $1.75) holding the rank position of the judged price constant (Janiszewski & Lichtenstein, 1999). The price was judged as most attractive in the high price context and least attractive in the low price context. The range of acceptable prices in the low price context ($0.93 to $1.51) was lower than in the high price context ($1.02 to $1.66). A later follow-up study used a similar methodology, but holding the range of prices constant while manipulating the rank position of the test price within the set of contextual prices (Niedrich, Sharma, & Wedell, 2001). Effects of both range and frequency on price attractiveness ratings were obtained across three experiments, supporting both range-frequency theory and an exemplar representation of contextual ERPs.

1.6.3.3 Advertised External Reference Prices

Purchase prices are often presented in conjunction with an advertised ERP, such as a previous selling price, a competitor's price or a recommended selling price (Urbany, Bearden, & Weilbaker, 1988). This may be done in comparative advertising or at the point of sale. It has long been recognized that such advertised ERPs have the potential to mislead consumers, with the Federal Trade Commission in the US publishing guidelines about deceptive comparative pricing as long ago as 1958 (Dellabitta, Monroe, & McGinnis, 1981), while legal action has been taken against major retailers over comparative price advertising practices (Grewal & Compeau, 1992). Including an advertised ERP is consistently found to improve perceptions of a purchase price (Alford & Engelland, 2000; Biswas & Blair, 1991; Compeau & Grewal, 1998; Grewal, Monroe, & Krishnan, 1998; Urbany, et al., 1988). Increasing the gap between the purchase price and the advertised ERP generally improves perceptions of the offer (Biswas & Blair, 1991; Biswas, Pullig, Krishnan, & Burton, 1999; Compeau & Grewal, 1998; Compeau, Grewal, & Chandrashekaran, 2002; Urbany, et al., 1988). It has been argued that as the advertised ERP increases, the plausibility of the ERP decreases and consumers will discount high ERPs, but the experimental evidence is mixed with some researchers finding a decreased impact of implausible ERPs (Alford & Engelland, 2000; Biswas & Blair, 1991; Compeau, et al., 2002; Kopalle & Lindsey-Mullikin, 2003) while other researchers found no difference between plausible and implausible ERPs (Compeau & Grewal, 1998; Urbany, et al., 1988). Other factors found to increase the impact of an advertised ERP include the type of store (Biswas & Blair, 1991); an absence of other external price information such as competitor advertising (Biswas, et al., 1999); naming specific competitors (Pechmann, 1996); and specifying the savings amount (Dellabitta, et al., 1981; Pechmann, 1996).

1.6.3.4 Internal Reference Prices

Purchase prices for items can vary significantly between purchase occasions, due to inflation, seasonal fluctuations, price differences between retailers, and promotional offers. Because of uncertainty about the true price of an item, it is assumed that consumers form a price forecast or internal reference price (IRP) which they compare against an observed price. Models fit to consumer purchase data that incorporate an IRP show a better fit than those which only use contextual prices (Kalyanaram & Winer, 1995; Urbany & Dickson, 1991; Winer, 1986). Typically the IRP is extrapolated from previously observed prices (Winer, 1986), and the range of acceptable prices are less variable than market prices, indicating that IRPs can be used as a purchase decision criterion (Urbany & Dickson, 1991). Consistent with other research on loss aversion, consumers appear to be more sensitive to prices above their IRP than those below (Kalyanaram & Winer, 1995). Consumers also appear to have brand-level rather than category-level IRPs, and use a weighted average of past prices for a brand to estimate an IRP, placing more weight on more recent observations (Briesch, Krishnamurthi, Mazumdar, & Raj, 1997). The particular exemplar price memories recalled when estimating an IRP can be influenced by priming different features of the product, causing it to be categorized in different ways, particularly by novice consumers (Herr, 1989). Experienced shoppers are more likely to rely on an IRP, especially when price claims are higher than expected, whereas novice shoppers are more influenced by advertised price claims (Yadav & Seiders, 1998). Importantly for this thesis, the existence and influence of IRPs indicates that consumers can store and recall item-level price information and subsequently use that information to make price judgments and purchase decisions.

1.6.3.5 Other Influences on Item Price Perceptions

Beyond the influence of reference prices, item price perceptions have been found to be influenced by a wide range of other factors. For example, price and quality are strongly linked in consumers' minds, with high price often treated as an indicator of high quality (Monroe, 1973). The perceived *value* of an item is a function of both price and perceived quality. If consumers place more weight on the price then low price items are judged to be better value, while if consumers place more weight on the quality then high price items are judged to be better value. Consumers are more likely to utilize the price-quality heuristic when they have low motivation or insufficient time to process product information systematically (Suri & Monroe, 2003). The tendency to use the price cue to judge quality is highest amongst novice shoppers and decreases with product experience, although expert shoppers also utilize a price-quality heuristic when price is a reliable environmental predictor of quality (Rao & Monroe, 1988). Item prices are often judged in the context of other prices and the general circumstances in which a price is presented. Consumers judge a price increase that exploits an external shift in demand (e.g. doubling the price of shovels when it snows) as unfair, but price increases to prevent a loss or protect profits (e.g. passing on an increase in wholesale costs) as fair (Kahneman, Knetsch, & Thaler, 1986). Because of diminishing marginal utility for both losses and gains, consumers prefer to integrate two price increases (losses) but to separate two discounts (gains), although this is dependent upon the way in which the price increases and losses are described (Heath, Chatterjee, & France, 1995). Small price changes which reduce the leftmost digit (e.g. reducing $3.00 to $2.99) have a disproportionate effect on the perceived size of a price, especially when that price is being compared to a competing product's price which is very close in magnitude (Thomas & Morwitz, 2005). Finally, item price perceptions exhibit significant individual differences. Accurate price perceptions are only weakly related to demographic and shopping variables (Brown, 1971), but consumers differ markedly in 'shopping personality' factors such as their tendency to focus exclusively on paying the lowest possible price ('price consciousness') and their propensity to respond to prices framed as promotional offers ('sale proneness') (Alford & Biswas, 2002). A more general personality factor that appears to be related to the degree of price search is 'Need for Closure', defined as "the desire for clear, definite, or unambiguous knowledge that will guide perception and action" (Vermeir & Van Kenhove, 2005, p. 73). Most of these additional factors are irrelevant to the topic of this thesis and so are not discussed further here.

1.6.4 Store Price Perceptions

1.6.4.1 Item and Store Selection

Retailers and consumer researchers have been greatly concerned with understanding item price perceptions as pricing and promotional activity has been consistently shown to influence consumer choice behaviour (Lattin & Bucklin, 1989). However, item and brand choice is not the only area in which price and price perceptions play a role in consumer decision making: price and price image are also known to influence store choice, especially for grocery shoppers. Using a multi-attribute utility approach, Schuler (1979) found that the price of the merchandise was second only to the quality of the merchandise in determining consumers' choice of store. An international study, modelling consumer store choice behaviour in six countries and over seven years, found that although store choice determinants varied between countries and over time, price was consistently the second most important factor after locational convenience (Arnold, Oum, & Tigert, 1983). Across a body of research into store choice behaviour, inclusion of a store image parameter generally improves fit relative to models which only consider spatial factors (Craig, Ghosh, & McLafferty, 1984). Consumers themselves also rate price as one of the most important factors in their choice of store, especially when switching their patronage to a different store (Seiders & Costley, 1994).

1.6.4.2 Determinants of Store Price Perceptions

In contrast to research into item price perceptions, research into the determinants of store price perceptions is relatively scarce. Buyukkurt and Buyukkurt (1986) argue that consumers find it difficult to judge a store's prices purely from price samples, so instead base their perception on store attributes that they perceive to correlate with a store's prices, as predicted by Attribution Theory. Their survey research found that consumers believe that stores with extra operating costs and investments (e.g. short queues, long opening hours, attractive displays) are more expensive, as are smaller businesses such as family-owned or independent stores. Stores with an extended range, including deli and bakery, gourmet food and non-food products are also perceived as more expensive. Finally, stores that advertise more heavily were also believed to be more expensive. Consumers were more likely to infer store price perceptions from store attributes when they perceived a wide variation in price between different retailers. The design of a store's advertising also influences perceptions of price, with advertised prices presented as a reduction from a previous price (i.e. providing an advertised ERP) leading to the most favourable store-level price perceptions (Cox & Cox, 1990). However, repeated use of this strategy can have the opposite effect on store-level price perceptions if consumers associate the high advertised ERP with the store's regular selling prices (Blair, Harris, & Monroe, 2002). Heavy price advertising is related to intense price competition between stores, and consumers respond by shopping around more and updating their price beliefs about each retailer (Seiders & Costley, 1994). The objective accuracy of consumers' store-level price perceptions is improved by inter-store price search, number of stores shopped, and length of residence in the market, but only the first factor is subjectively perceived by consumers to improve store price knowledge (Magi & Julander, 2005). It is possible that other cues that are in reality irrelevant to store prices may also influence price judgments, as has been observed in related judgments of the value of different loyalty programs (Van Osselaer, Alba, & Manchanda, 2004).

In one of the few experimental studies assessing how judgments from memory influence store price perceptions, knowledgeable consumers were found to use a different judgment heuristic from less knowledgeable consumers (Ofir, Raghubir, Brosh, Monroe, & Heiman, 2008). Participants were intercepted while approaching a supermarket and were asked to recall either two or five low-priced products sold by the store. They were then asked to rate the prices in the store on two different seven-point scales. The authors suggested two alternative heuristics that consumers could use to make the store price judgment: the availability heuristic, where price image is judged by the ease of recalling low prices, or the 'numerosity' heuristic, where price image is judged by the total number of low prices recalled i.e. a frequency judgment. They found that knowledgeable customers had more favourable perceptions of the store's prices when they had to recall more products, indicating the use of a frequency heuristic. On the contrary, less knowledgeable customers had more favourable perceptions of the store's prices when they had to recall fewer products, which is an easier task, indicating the use of an availability heuristic. Discussing these results in light of the findings of Alba et al (1994), the authors suggested that the previously observed frequency effect was due to testing a stimulus-based judgment task, while their study tested a memory-based judgment task. If this is the case, then one would not expect to see a frequency effect in inter-store price judgments, as the large salient price advantages of a magnitude store would be more available for memory recall than the small price advantages of a frequency store. However, the artificial priming effect of the earlier price recall task is unlikely to reflect the way intuitive store price judgments are usually made. Hence, use of the availability heuristic may not be observed in a more naturalistic judgment task. To the best of my knowledge, with the exception of this study, no other consumer research into memory-based judgments of store prices has been conducted to date.

1.6.4.3 Store Price Format

Grocery retailers have to make strategic decisions concerning both the *price level* of their store and also the *price format* of the store. The price level is the average price charged for a comparable basket of items, whereas the price format is the distribution of item prices around that mean level. The marketing literature traditionally distinguishes between two polar opposites of price format: Every Day Low Pricing (EDLP) and Promotional pricing (PROMO, sometimes referred to as HiLO). EDLP is a price strategy in which all prices are kept as low as possible, hence they tend to cluster around the mean item price. PROMO pricing involves higher item prices across most of the range, but with a few deep discounts. The two formats correspond crudely to Alba et al's frequency and magnitude stores. In reality, PROMO stores are also usually associated with a higher price level, but also offer a higher service level as demanded by time-constrained shoppers (Lal & Rao, 1997). Although it has been argued that the two store formats co-exist because they serve different market segments, empirical data suggests that stores tend to adopt a similar pricing strategy to local competitors, leading to clusters of similar price format stores (Ellickson & Misra, 2008).

Recent research findings suggest that consumer price perceptions of the two price formats depend upon the type of consumer, specifically whether the household shops frequently and buys a small number of items on each occasion (small basket shoppers) or shops infrequently and buys a large number of items on each occasion (large basket shoppers) (Bell & Lattin, 1998). Large basket shoppers have a relatively high probability of purchase from every product category so are less responsive to prices within individual product categories and more responsive to the expected basket price. As a result, large basket shoppers are more likely to choose EDLP stores while small basket shoppers are more likely to choose PROMO stores, even if the average item price is slightly higher, because they can cherry-pick the best offers on each occasion. Empirical testing using scanner data with the purchase behaviour of 1,042 households over two years supports this hypothesised pattern in store choice behaviour (although does not measure price perceptions directly). In an experimental study of the determinants of store price perceptions, Desai and Talukdar (2003) explored whether discounts applied to different types of item have a differential impact on store price perceptions. Based on survey research with a student population, they found that price reductions on items with a short consumption span and a large unit price had the greatest impact on store price perceptions. A store with a PROMO price format will have a better or worse price image depending upon which particular items are chosen for promotion.

1.7 Motivation for Thesis

The goal of this first chapter was to describe a collection of empirical findings concerning intuitive statistical discrimination judgments of grocery store prices from the consumer research literature, notably Alba et al (1994), which are suggestive of an important role for frequency information in a cognitively complex judgment task. The theories and empirical findings related to stimulus discrimination and intuitive statistical judgments were outlined, including a detailed discussion of the role of frequency information in different task formats. An overview of human memory was given, focusing on results relevant to the storage and retrieval of exemplars in episodic memory for use in intuitive judgments. Some motivations for designing naturalistic or representative tasks and contexts for experimental research were also sketched out. Finally, the consumer research literature concerning price perceptions and price judgments was summarized, highlighting the lack of empirical data concerning the interaction between store price format and store price perceptions. Furthermore, it was shown that no analogous intuitive statistical judgments have been tested in the psychology literature. No experiments were found in which participants compared the mean size of two *paired item* distributions, either in a descriptive or serially sampled format.

As a consequence, my own research aims to address both of these gaps, and to contribute to a growing body of empirical results on human intuitive statistical judgments in naturalistic tasks. While the problem and context of grocery store price comparisons is adopted from the consumer research literature, and the findings will no doubt be of interest to practitioners in that field, the primary focus is on exploring the cognitive processes involved in the task. This review has laid out why the task under consideration (inter-store price comparisons) might differ from the related empirical findings described at the start of this chapter: switching from forced attention to incidental sampling of price information; and switching from simultaneous presentation (stimulus-based judgment) to serial presentation (memory-based judgment). By the same token, these factors clearly delineate how far the findings from the experiments in this research might be extrapolated to analogous judgment tasks. Specifically, this thesis aims to address the following research questions:

1. How sensitive are participants at discriminating between the means of two paired item distributions?

2. Do the frequency or magnitude of paired item differences bias intuitive judgments when the two distributions have equal means?
3. Are judgments made differently in stimulus-based and memory-based tasks?
4. Are judgments made differently in forced-attention and incidental-sampling experimental paradigms?
5. What cognitive processes are involved in making these intuitive statistical judgments?

I will provide evidence that the biasing effect of frequency dominates mean discrimination judgments, even in serial presentation of incidentally-sampled information, although the effect is strongest in simultaneous presentation of information with forced attention. Furthermore, I will show that sensitivity even to large differences in mean between two paired item distributions is weak when serial sampling and incidental acquisition of information are involved. I will demonstrate that basket cost estimates are an inappropriate measure of store-level price perceptions, which can even be negatively correlated with price judgments under certain circumstances, and test hypotheses from the consumer research literature concerning the interaction between basket size and price perceptions. Finally, I will fit different cognitive process models to individual-level data to determine which theories offer the best explanation of the observed judgment bias, and provide evidence for a judgment process involving pair-wise item price comparisons that are additively integrated and weighted by attention.

A secondary contribution of this thesis to the literature on judgments and decision making is a methodological one. Experiments 3 and 4 were conducted online and required participants to carry out a relatively lengthy, complex and immersive simulation task. The use of the web as a research tool and the simultaneous collection of data from hundreds of participants in controlled experiments outside of the traditional laboratory setting are unusual but not unique in the experimental psychology literature. Equally, the use of complex naturalistic tasks carried out in a representative ecology of information is not new, although such tasks differ from the classical stripped-down laboratory methods. However, the combination of the two is novel, as prior web-based research has tended to utilize much simpler and less-involving tasks. Thus the research presented in this thesis, although not designed as a comparison between online and offline research, demonstrates the methodological possibilities opened up by the latest developments in software and web technology, and the increasing ubiquity of web access. Because of the relative novelty of web-based psychology experiments, I conclude this literature review with a brief survey of prior research and best practice in conducting experiments via the internet.

1.8 Web-Based Experiments

1.8.1 Development of the Internet as a Research Tool

The internet, particularly the World Wide Web, has developed into a platform for mainstream communication, leading to increased interest from psychologists into the potential of internet-based experimental research. In fact, psychology experiments have been running on the web since at least 1995 (Reips, 2001) and web-based research studies are now commonly found in peer-reviewed APA journals (Skitka & Sargis, 2006). In the field of judgment and decision making research, web-based studies from a diverse range of topics have been published, including probability learning (Birnbaum & Wakcher, 2002), medical decision making (Waters, Weinstein, Colditz, & Emmons, 2006), choice between risky gambles (Birnbaum, 1999), reaction times in binary choice (Reimers & Stewart, 2007), and task-switching (Reimers & Maylor, 2005), as well as many other areas of psychology such as psychometric personality assessment (Buchanan & Smith, 1999). The internet is also increasingly being used as a medium for publishing and disseminating psychology research studies (Brezsnyak, 1999).

The correspondence between web and lab experiments is generally very high, with comparison studies obtaining broadly similar findings from lab and web samples (Birnbaum, 1999, 2004; Buchanan & Smith, 1999; Smith & Leigh, 1997). The main differences between web and lab, particularly in early web-based studies, were demographic differences between samples. Compared to traditional undergraduate participant samples, internet users tend to be older, more highly educated and are more likely to be male (Birnbaum, 1999; Smith & Leigh, 1997). However, web samples also tend to be more demographically diverse than student populations (Birnbaum, 1999; Smith & Leigh, 1997) and the web potentially offers access to more specialized populations (Birnbaum, 2004; Skitka & Sargis, 2006). Increased demographic variance potentially both reduces sample bias but also adds noise through an additional uncontrolled source of variability (Buchanan & Smith, 1999; Schmidt, 1997). As with traditional lab studies, random sampling helps to reduce sample bias but researchers have to be aware of a potential self-selection bias in the type of internet user who completes online surveys and experiments (Birnbaum, 2004; Buchanan & Smith, 1999; Smith & Leigh, 1997).

1.8.2 Best Practice in Web-Based Research

In addition to access to a large and demographically diverse pool of research participants, numerous other advantages have been cited for conducting research on the web, including increased speed, lower cost of materials, participant anonymity, increased flexibility over time and place of interaction, reduced experimenter bias, increased experimental power, the ability to create interactive and tailored experiments, and automation of data acquisition and analysis (Birnbaum, 2004; Hewson, Laurent, & Vogel, 1996; Schmidt, 1997; Skitka & Sargis, 2006; Smith & Leigh, 1997). Many of these advantages are not specific to web-based experiments but apply to offline computer-based testing as well. Some criticisms and concerns have also been raised, particularly concerning ethical issues, participant fraud, high dropout and non-response rates, limitations in the type of stimuli that can be delivered, technical constraints and lack of control over the PC and screen being used, greater variability in the context in which an experiment takes place, lack of participant accountability, multiple submissions, and inaccurate control and measurement of temporal intervals (Birnbaum, 2004; Schmidt, 1997; Skitka & Sargis, 2006; Smith & Leigh, 1997).

As a result of these concerns a number of researchers have laid out best practice guidelines for web-based research. For example, just as in lab experiments, participants must be able to give informed consent, have the right to withdraw, and participants' anonymity and data security must be protected (Schmidt, 1997; Smith & Leigh, 1997). Participants must also be debriefed as to the purpose of a study after completion and should not be subject to deception (Reips, 2001). The risk of multiple submissions can be reduced through giving clear participation instructions, by removing incentives for those who participate more than once, by tracking IP addresses or personal identifiers such as e-mail addresses, and by filtering data for identical or nearly-identical records (Birnbaum, 2004; Smith & Leigh, 1997). Participants can be recruited either passively (e.g. through advertising) or actively (e.g. through targeted e-mails) but contact must never be unsolicited (Birnbaum, 2004; Schmidt, 1997). Experiments should be piloted before launching them on the web in order to check that instructions are clear, that data is being recorded correctly, and to check that participants complete the experiment in the required manner (Birnbaum, 2004). Web experiments should be as standardized as possible, not requiring any special software or hardware, and should be designed to run on as many different web browsers as possible (Hewson, et al., 1996; Schmidt, 1997). All these guidelines were carefully followed in the web experiments presented in this thesis, in order to maximize data quality and accuracy.

1.8.3 Web-Based Interactive Simulations

Computer-based simulations have previously been shown to be reasonable predictors of consumers' real-world shopping decisions (Burke, Harlam, Kahn, & Lodish, 1992). Simulation tasks predict consumer choices best when the appropriate product cues can be reproduced, so computer-based simulations are less useful for product choices involving non-visual cues such as texture or smell. Increasing the realism of store design and shopping task features also increases the validity of observed choices. Participants are generally more sensitive to price and promotions in simple simulations and tend to 'buy' a greater quantity of the discounted good, as they face no budget or space constraints. Finally, time-compression means that repeated choices in simulations tend to show a greater degree of repetition and routine than in the real world. However, the overall correspondence between simulation and real-world is good, especially for relative rather than absolute measures of choice behaviour.

Advances in web-based software mean that relatively rich, complex and involving simulations can now be delivered via the web, as well as automating the allocation of participants to experimental conditions and the collection of data. A variety of options are available for delivering web-based experiments, both client-side (running on the participant's local machine) and server-side (running on a central server) although data collection obviously has to occur on a central server (Birnbaum, 2004). Javascript has been commonly used for online experimental studies (Birnbaum & Wakcher, 2002), but the browser extension Adobe Flash (formerly known as Macromedia Flash) is gaining in acceptance (Reimers & Stewart, 2007). The advantages of Flash are that it allows attractive animated content to be created quickly and easily, it is ubiquitous (Adobe's website claims over 99% of internet users have the plugin) and freely available, and pre-prepared components like menus, text boxes and buttons make the creation of an ergonomic interface straightforward (Reimers & Stewart, 2007; Schmidt, 2001), thus avoiding problems caused by unclear design (Birnbaum, 2004). For these reasons, Flash is particularly suited to creating simulated environments such as the shopping task in Experiments 3 and 4, and was therefore chosen to implement these experiments.

1.9 Thesis Structure

The rest of this thesis is organized into five further chapters. Chapter 2 presents the results of two experiments in which the findings of Alba et al are replicated (Experiment 1) and then extended into a serial presentation format (Experiment 2). Chapter 3 describes a novel online shopping simulation task used to determine participants' sensitivity to inter-store mean price differences (Experiment 3). Chapter 4 consists of a further experiment using the same experimental paradigm, testing for a biasing effect of frequency or magnitude cues in inter-store price discrimination (Experiment 4). Finally, Chapter 5 compares the ability of different cognitive process models to explain the data observed in Experiments 3 and 4, before Chapter 6 concludes with a summary of the findings, conclusions that can be drawn from the research presented, and concrete suggestions for future avenues of exploration that would further develop understanding of naturalistic intuitive judgments.

CHAPTER 2

BASKET COST ESTIMATES IN PAIRED VS. POOLED PRICE INFORMATION PROCESSING (EXPERIMENTS 1 AND 2)

2.1 Introduction

In the original study by Alba et al (1994), as described in Chapter 1, various experiments tested the relative impact of the frequency and magnitude cues in comparative store price judgments. Although a number of attempts were made to sensitize participants to the trade-off between frequency and magnitude (Experiment 5) or to emphasize the magnitude cue (Experiment 6), in all cases the prices in the two stores were presented side-by-side, enhancing the salience and accessibility of the frequency cue. I shall refer to this mode of presentation as *Paired Presentation*. The central purpose of the two experiments described in this chapter was to compare and contrast Paired Presentation of prices with a format in which all prices from one store are sampled before all the prices in the second store are seen. This kind of *Pooled Presentation* format both reduces the salience and availability of the frequency cue (by requiring each item price from the first store to be recalled from memory) and is also a closer replication of the sequence of price information observed when shoppers engage in comparison shopping in two different stores.

The purpose of Experiment 1 was to replicate a Paired Presentation experiment in a UK context, in order to confirm the relative importance of the frequency cue, the magnitude cue and prior beliefs upon basket cost judgments in paired presentation with forced attention, as found by Alba et al (1994). Experiment 2 adapted the methodology of Experiment 1 to present the same information in a

Pooled Presentation format. A meta-analysis across the two experiments was conducted in order to examine the impact of presentation format upon participants' basket cost estimates.

2.2 Paired Price Presentation (Experiment 1)

2.2.1 Method

2.2.1.1 Participants

64 participants, consisting of 38 men and 26 women aged between 19 and 69 years with a mean age of 32 years, were recruited on the campus of the University of Warwick via face-to-face recruitment. In order to ensure sufficient prior experience of and accurate prior beliefs about UK supermarket prices, older participants including staff, postgraduate students and visitors were targeted. All participants received £5 for their participation.

2.2.1.2 Stimuli

30 pairs of prices were selected from the two largest UK grocery retailers, Tesco and Asda. The selected items are all commonly purchased items, sold in both stores, and available in local stores at the time price data was collected. The total cost of the items in the two retailers (at the time of conducting the experiment, excluding promotional offers) was £29.23 in Tesco and £29.85 in Asda. Tesco had the lower price for 10 of the items, by an average of 27p. Asda had the lower price for the remaining 20 items, by an average of 10p.

2.2.1.3 Design and Procedure

The experiment was set up as a between-subjects design with two conditions: a Magnitude condition, using the actual item prices, and a Frequency condition in which the Tesco prices were altered in order to double the number of items for which Tesco had the lower price, whilst maintaining the same total basket cost and approximate magnitudes of price advantages and disadvantages. In the Frequency condition, Tesco's 20 price advantages were lower by an average of 15p and Asda's 10 price advantages were lower by an average of 23p. The complete list of items and prices in each condition is shown in Table 2.1. Assignment to the two conditions was random. 32 participants were assigned to the Magnitude condition and the remaining 32 participants were assigned to the Frequency condition.

The experiment was implemented on a laptop computer using the E-Prime software package. After giving their age and gender, and reading some brief instructions, participants were asked about their prior price beliefs about the two stores using a 5-point response scale:

"Compared to other supermarkets, on the products I buy regularly, [Tesco/Asda] has:"

1. Cheaper prices on all
2. Cheaper prices on some
3. Average prices on all
4. More expensive prices on some
5. More expensive prices on all
6. Don't know

Participants were next told that they would be shown the prices of a basket of thirty items sold in Tesco and Asda. They were asked to read the information carefully and to memorize as much as possible about the prices, in order to answer some subsequent questions. They were also told that there was no time limit and they were free to spend as long as necessary on each item.

TABLE 2.1
Items and prices used in Experiment 1

	Asda	Tesco	
Item Description		Magnitude	Frequency
Baby Potatoes (1 Kg)	88p	58p	58p
Baking Potatoes (1 Kg)	78p	85p	99p
Broccoli	68p	99p	99p
Bunch of Spring Onions	47p	48p	45p
Celery	47p	48p	45p
Chicken Tonight (500g Jar)	74p	75p	73p
Conference Pears (1 Kg)	£1.28	£1.19	£1.19
Crusty White Split Tin Loaf (800g)	58p	63p	56p
Frozen Peas (1 Kg)	43p	78p	84p
Golden Delicious Apples (1 Kg)	£1.08	£1.18	£1.24
Iceberg Lettuce	58p	57p	57p
Large Eggs (6 Pack)	68p	82p	82p
Lemons	17p	22p	16p
Low Fat Fruit Yoghurt (4 Pack)	98p	83p	83p
Medium Eggs (6 Pack)	58p	72p	75p
Medium Sized Tomatoes (1 Kg)	£1.28	£1.29	£1.25
Medium Tomatoes (6 Per Pack)	68p	49p	49p
Muller Fruit Corner (175g)	34p	38p	30p
PG Tips Pyramid Tea Bags (80)	£1.38	£1.44	£1.59
Plum Tomatoes (400g)	24p	19p	19p
Radishes (125g)	37p	44p	45p
Royal Gala Apples (1 Kg)	£1.28	£1.49	£1.49
Sliced Danish Loaf (400g)	28p	30p	25p
Sparkling Mineral Water (2 litres)	43p	18p	18p
Strawberries (1 Kg)	£7.84	£6.23	£6.23
Swede (1 Kg)	59p	99p	99p
Toilet Tissue (4 Pack)	44p	42p	42p
White Finger Rolls (6 Pack)	36p	37p	35p
Whole Fresh Chicken (Large - Per Kg)	£1.97	£1.95	£1.95
Whole Fresh Chicken (Medium - Per Kg)	£1.99	£2.00	£1.95
	£29.85	£29.23	£29.23

Each item description was displayed in the centre of the screen with the two store prices below, one on the left and one on the right. Each price was labelled with the store name. The store order was counter-balanced between participants but did not vary within the experiment[1]. Each price was displayed for a minimum of two seconds, after which the participant could press the spacebar to progress to the next item. All thirty items were shown in a random order without replacement.

Participants were then asked to judge which store they thought was cheapest for the items shown, using a 7-point response scale:

"The total cost of these items was different in the two stores. In which store do you think the total cost was lowest?"

1. Definitely Asda
2. Almost certainly Asda
3. Probably Asda
4. They seemed the same to me
5. Probably Tesco
6. Almost certainly Tesco
7. Definitely Tesco

In addition, participants estimated the total cost of the items in each store: "What was the total cost of these items in [Asda/Tesco]?" A manipulation check was also carried out to determine whether participants were aware of the frequency cue: "For how many of the thirty items was Tesco cheaper than Asda?" Finally, participants gave their posterior beliefs about the prices in each store, using the same 5-point

[1] All ANOVA and ANCOVA analyses were repeated with the store order as an additional between-subjects factor, but no significant effects were found ($\alpha<0.05$). The effect of store order is therefore not reported in these results for the purpose of clarity.

response scale as at the start of the experiment. At the end of the experiment participants were debriefed as to the purpose of the study and informed that the prices they had been shown may not have reflected actual prices.

2.2.2 Results

2.2.2.1 Cost Estimates

There was a wide spread of cost estimates for both Asda (M = £32.53, SD = £17.13) and Tesco (M = £33.67, SD = £16.56). The mean cost estimates for Asda did not vary significantly between the Magnitude and Frequency conditions (£32.17 vs. £32.89, $t(62) = -0.165$, $p = 0.87$, two-tailed) and neither did the mean cost estimates for Tesco vary significantly between the two conditions (£35.66 vs. £31.68, $t(62) = 0.960$, $p = 0.34$, two-tailed). The cost estimates in each store were highly correlated, $r(62) = 0.96$, $p<0.001$. A one-way ANCOVA model was used to partial out the variance in Tesco cost estimates explained by the cost estimates in Asda, the effect of which was highly significant ($R^2 = 0.94$; $F(1,61) = 881.49$, $p<0.001$). The adjusted mean cost estimates for Tesco given by participants in the Magnitude condition (M_{adj} = £35.99) were higher than the adjusted mean cost estimates given by participants in the Frequency condition (M_{adj} = £31.35) and the difference was highly significant ($F(1,61) = 19.04$, $p<0.001$, $\eta^2=0.24$).

2.2.2.2 Prior Beliefs

Participant's beliefs about the prices in each store were collected at the start and end of the experiment. Prior beliefs about Asda did not vary significantly between the Magnitude condition ($M = 2.0$) and the Frequency condition ($M = 2.3$)

($t(47) = -1.156$, $p = 0.25$, two-tailed)[2]. Similarly, prior beliefs about Tesco did not vary significantly between the Magnitude condition ($M = 2.1$) and the Frequency condition ($M = 2.3$) ($t(60) = -0.991$, $p = 0.25$, two-tailed)[3]. Hence, the random allocation of participants between the two experimental conditions cannot explain the previous difference in Tesco cost estimates.

The residual of the previous ANCOVA model was significantly correlated with the participants' prior beliefs about prices in Tesco, $r(60) = 0.33$, $p<0.01$. Participants with more favourable prior beliefs about Tesco tended to give lower cost estimates. Including prior belief as an additional covariate in the ANCOVA model showed significant effects of both prior beliefs and cost estimates for Asda (Table 2.2). The adjusted mean cost estimate for Tesco given by participants in the Magnitude condition (M_{adj} = £35.92) was higher than the adjusted mean cost estimate given by participants in the Frequency condition (M_{adj} = £30.88).

TABLE 2.2
ANCOVA model of cost estimates in Tesco including cost estimate for Asda and prior beliefs about Tesco prices as covariates (Experiment 1)

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Cost Estimate Asda	15333.82	1	15333.82	917.10	<0.001	0.94
Prior Beliefs Tesco	124.68	1	124.68	7.46	<0.01	0.11
Price Condition	387.17	1	387.17	23.16	<0.001	0.29
Error	969.76	58	16.72			

2.2.2.3 Confidence Judgments

Participants' confidence judgments of which store was cheaper were uncorrelated with prior beliefs about Asda, $r(47) = 0.184$, $p = 0.21$. Similarly,

[2] Fifteen participants, nine in the Magnitude condition and six in the Frequency condition, answered 'Don't know' to the question concerning their prior beliefs about prices in Asda.
[3] Two participants in the Magnitude condition answered 'Don't know' to the question concerning their prior beliefs about prices in Tesco so were excluded from the subsequent ANCOVA model.

confidence judgments were uncorrelated with prior beliefs about Tesco, $r(60) = -0.05$, $p = 0.69$. Confidence judgments do not appear to have been significantly influenced by participants' prior beliefs.

Participants' confidence judgments of which store was cheaper differed significantly between the Magnitude and Frequency conditions ($t(62) = -5.236$, $p<0.001$, two-tailed). In the Magnitude condition participants were confident that Asda was the cheaper of the two stores ($M = 2.6$, $SD = 1.5$), and the mean rating differed significantly from the neutral rating of 4 ($t(31) = -5.230$, $p<0.001$, two-tailed). In the Frequency condition participants were weakly confident that Tesco was the cheaper of the two stores ($M = 4.7$, $SD = 1.7$) and the mean rating differed significantly from the neutral rating ($t(31) = 2.323$, $p<0.05$, two-tailed).

2.2.2.4 Manipulation Check

Participants' estimates of how many of the items were cheaper in Tesco were uncorrelated with prior beliefs about Asda, $r(47) = 0.166$, $p = 0.25$. Similarly, estimates of the frequency of Tesco's price advantages were uncorrelated with prior beliefs about Tesco, $r(60) = -0.151$, $p = 0.24$. Frequency estimates do not appear to have been significantly influenced by participants' prior beliefs.

Participants' mean estimate of how many of the items were cheaper in Tesco were significantly lower in the Magnitude condition ($M = 12.1$, $SD = 5.9$) than in the Frequency condition ($M = 17.1$, $SD = 5.8$) ($t(62) = -3.417$, $p<0.001$, two-tailed). Participants were both sensitive to differences in the frequency cue and able to make reasonably accurate estimates of the true frequency.

2.2.3 Discussion

Consistent with the study by Alba et al (1994), Experiment 1 found that total basket cost estimates are strongly influenced by the frequency cue when comparative price information is presented in a Paired format and that the frequency cue dominates the magnitude cue and prior price beliefs. Experiment 1 replicated these findings in a UK context, with representative items and prices. In addition, participants' qualitative confidence judgments concerning which store was cheaper mirrored the basket cost finding.

Unlike the US study, where all price information was presented in a booklet and was therefore readily available to participants, the method of Paired Presentation deliberately emphasized the frequency cue. This format of presentation makes basket cost estimates more difficult. In order to calculate the basket cost in each store a participant would have to maintain two running totals simultaneously. In order to calculate the cheapest store a participant would have to calculate the magnitude of the price difference in each case and maintain a running total. It is therefore not surprising that the frequency cue dominated the magnitude cue.

Given the difficulty of the task, it is perhaps surprising that participants did not generalize from their prior beliefs about the prices in each store when estimating the total costs. A simple heuristic would be to estimate a total cost in the first store and then to adjust the second estimate up or down from the first, congruent with prior beliefs. Although there is a weak effect of prior beliefs, the relative effect size as indicated by the partial eta-squared values in the ANCOVA model was only about one third as large as the effect of the frequency cue. Participants were also sensitive to the frequency manipulation as shown in their estimated values of the frequency cue. In summary, Experiment 1 demonstrated that the frequency cue dominates participants' cost estimates in a Paired Presentation format which enhances the salience and availability of frequency information.

As described earlier, the purpose of Experiment 2 was to contrast the Paired Presentation format of Experiment 1 with a Pooled Presentation format, in which all the item prices for one store are viewed before moving onto the item prices for the second store. This has the effect of making the frequency cue less salient (by no longer presenting matched item prices in pairs) and less readily available. In a Paired Presentation format, participants can encode the information as to which store was cheaper for each item directly into memory. To utilize the frequency cue each participant then had only to maintain a running count of frequency for each store or to retrieve the encoded binary information (cheaper / more expensive) from memory. In contrast, in a Pooled Presentation format participants would have to encode both the item description and actual price in the first store, in order to have the information available (assuming accurate recall) to produce and encode binary information as to which store was cheaper for each item. Hence, ceteris paribus, one would expect the frequency cue to have less or no impact in the case of Pooled Presentation (Hypothesis 1a).

A Pooled Presentation format could, on the other hand, make the magnitude cue more salient. If encoding or retrieval of item price information from the first store is noisy or inaccurate then small price differences between items could lead to uncertainty over which store is cheaper for that item, or even an incorrect judgment. In contrast, large price differences are less prone to such errors and uncertainty. Therefore, even if participants are attempting to use a frequency heuristic, one would expect to see an increased impact of the magnitude cue. In the current experimental procedure, this would be indistinguishable from the effect described in the previous paragraph as the frequency and magnitude cue are negatively correlated between the two price conditions. Hence, evidence for decreased impact of the frequency cue (Hypothesis 1a) could also be interpreted as evidence for a greater impact of the magnitude cue (Hypothesis 1b).

Finally, given the above discussion concerning the increased task complexity, one might expect to see an increased reliance on prior beliefs when price information is given in a Pooled Presentation format (Hypothesis 2). Similarly, one would expect participants' confidence judgments as to which store was the cheaper to be less extreme, reflecting greater uncertainty due to increased task complexity and the reduced availability of the frequency cue (Hypothesis 3).

In the following section the method and results for Experiment 2 are described. A meta-analysis across the two experiments is then presented, which explicitly tests the prior hypotheses outlined above concerning the impact of presentation format on cost estimation in a store comparison task.

2.3 Pooled Price Presentation (Experiment 2)

2.3.1 Method

2.3.1.1 Participants

64 participants, consisting of 25 men and 39 women aged between 18 and 57 years with a mean age of 27 years, were recruited on the campus of the University of Warwick via face-to-face recruitment. In order to ensure sufficient prior experience of and accurate prior beliefs about UK supermarket prices, older participants including staff, postgraduate students and visitors were targeted. Participants from

Experiment 1 were excluded from Experiment 2. All participants received £5 for their participation.

2.3.1.2 Stimuli

The 30 pairs of prices for items used in Experiment 1 were repeated in Experiment 2. As before, the total cost of the items in the two retailers was £29.23 in Tesco and £29.85 in Asda. Tesco had the lower price for 10 of the items, by an average of 27p. Asda had the lower price for the remaining 20 items, by an average of 10p.

2.3.1.3 Design and Procedure

Experiment 2 was an exact replication of Experiment 1, being a between-subjects design with two conditions: a Magnitude condition, using the actual item prices, and a Frequency condition in which the Tesco prices were altered in order to double the number of items for which Tesco had the lower price, whilst maintaining the same total basket cost and approximate magnitude of price advantages and disadvantages. In the Frequency condition, Tesco's 20 price advantages were lower by an average of 15p and Asda's 10 price advantages were lower by an average of 23p. The complete list of items and prices in each condition is shown in Table 2.1. Assignment to the two conditions was random. 32 participants were assigned to the Magnitude condition and the remaining 32 participants were assigned to the Frequency condition.

The experimental procedure was identical to that used in Experiment 1, except for the format in which item prices were presented to participants. Each item description was displayed in the centre of the screen with the price in the first store below. Each price was labelled with the store name and was displayed for a minimum of two seconds, after which the participant could press the spacebar to progress to the next item. All thirty items were shown in a random order without replacement. After all the prices for the first store had been viewed the 30 items were repeated with the prices in the second store, labelled with the store name. The store order was counter-balanced between participants but did not vary within the experiment[4].

The instructions, questions and question ordering were unchanged from Experiment 1. As before, at the end of the experiment participants were debriefed as to the purpose of the study and informed that the prices they had been shown may not have reflected actual prices.

2.3.2 Results

2.3.2.1 Cost Estimates

There was a wide spread of cost estimates for both Asda (M = £29.12, SD = £11.17) and Tesco (M = £30.32, SD = £11.43). The mean cost estimates for Asda did not vary significantly between the Magnitude and Frequency conditions (£27.84 vs. £30.39, $t(62) = -0.912$, $p = 0.37$, two-tailed) and neither did the mean cost estimates for Tesco vary significantly between the two conditions (£30.10 vs. £30.53, $t(62) = -0.148$, $p = 0.88$, two-tailed). The cost estimates in each store were highly correlated, $r(62) = 0.94$, $p<0.001$. A one-way ANCOVA model was used to partial out the variance in Tesco cost estimates explained by the cost estimates in Asda, the effect of which was highly significant ($R^2 = 0.89$; $F(1,61) = 482.18$,

[4] All ANOVA and ANCOVA analyses were repeated with the store order as an additional between-subjects factor, but no significant effects were found ($\alpha<0.05$). The effect of store order is therefore not reported in these results for the purpose of clarity.

$p<0.001$). The adjusted mean cost estimate for Tesco given by participants in the Magnitude condition (M_{adj} = £31.34) was higher than the adjusted mean cost estimate given by participants in the Frequency condition (M_{adj} = £29.29) and the difference was small but significant ($F(1,61) = 4.385$, $p<0.05$, $\eta^2=0.07$).

2.3.2.2 Prior Beliefs

Participant's beliefs about the prices in each store were collected at the start and end of the experiment. Prior beliefs about Asda did not vary significantly between the Magnitude condition ($M = 1.9$) and the Frequency condition ($M = 1.8$) ($t(50) = 0.181$, $p = 0.86$, two-tailed) [5]. Similarly, prior beliefs about Tesco did not vary significantly between the Magnitude condition ($M = 2.2$) and the Frequency condition ($M = 2.1$) ($t(61) = 0.542$, $p = 0.59$, two-tailed) [6]. Hence, the random allocation of participants between the two experimental conditions cannot explain the previous difference in Tesco cost estimates.

The residual of the previous ANCOVA model was not correlated with the participants' prior beliefs about prices in Tesco, $r(61) = 0.12$, $p = 0.34$. Consequently, including prior belief as an additional covariate in the ANCOVA model showed no significant effect of prior beliefs ($F(1,59) = 0.916$, $p = 0.34$).

2.3.2.3 Confidence Judgments

Participants' confidence judgments of which store was cheaper were uncorrelated with prior beliefs about Asda, $r(49) = 0.088$, $p = 0.54$. Similarly, confidence judgments were uncorrelated with prior beliefs about Tesco, $r(61)$ =

[5] Twelve participants, six in the Magnitude condition and six in the Frequency condition, answered 'Don't know' to the question concerning their prior beliefs about prices in Asda.
[6] One participant in the Magnitude condition answered 'Don't know' to the question concerning their prior belief about prices in Tesco so was excluded from the subsequent ANCOVA model.

0.057, $p = 0.66$. Confidence judgments do not appear to have been significantly influenced by participants' prior beliefs.

Participants' confidence judgments of which store was cheaper differed significantly between the Magnitude and Frequency conditions ($t(62) = -2.322$, $p<0.05$, two-tailed). In the Magnitude condition participants were weakly confident that Asda was the cheaper of the two stores ($M = 3.3$, $SD = 1.1$), and the mean rating differed significantly from the neutral rating of 4 ($t(31) = -3.650$, $p<0.001$, two-tailed). In the Frequency condition participants were unsure as to which store had the lowest total cost ($M = 4.1$, $SD = 1.5$) and the mean rating did not differ significantly from the neutral rating ($t(31) = 0.229$, $p = 0.82$, two-tailed).

2.3.2.4 Manipulation Check

Participants' estimates of how many of the items were cheaper in Tesco were uncorrelated with prior beliefs about Asda, $r(50) = -0.011$, $p = 0.94$. Similarly, estimates of the frequency of Tesco's price advantages were uncorrelated with prior beliefs about Tesco, $r(61) = 0.085$, $p = 0.51$. Frequency estimates do not appear to have been significantly influenced by participants' prior beliefs.

Participants' mean estimate of how many of the items were cheaper in Tesco did not differ significantly between the Magnitude condition ($M = 9.8$, $SD = 4.5$) and the Frequency condition ($M = 11.8$, $SD = 5.1$) ($t(62) = -1.716$, $p = 0.09$, two-tailed). Participants were insensitive to differences in the frequency cue and were unable to make reasonably accurate estimates of the true frequency, especially in the Frequency condition.

2.4 Meta-Analysis: Paired vs. Pooled Presentation

2.4.1 Cost Estimates

There was a wide spread of cost estimates for both Asda (*M* = £30.83, *SD* = £14.51) and Tesco (*M* = £31.99, *SD* = £14.27) across the two experiments. Two-way ANOVA models were used to test for differences in mean cost estimates for each store between price conditions (Magnitude or Frequency) and presentation conditions (Paired or Pooled). The central tendency and spread of cost estimates are summarized in Table 2.3.

TABLE 2.3
Summary of cost estimates for Asda and Tesco across Experiments 1 and 2

	Asda				Tesco			
	Paired		Pooled		Paired		Pooled	
Price Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Magnitude	£32.17	£13.88	£27.84	£10.17	£35.66	£15.45	£30.10	£11.36
Frequency	£32.89	£20.09	£30.39	£12.12	£31.68	£17.63	£30.53	£11.68

Visual inspection of Table 2.3 suggests that the assumption of equal variances is not met: the variances in the Pooled Presentation cells are lower than in the Paired Presentation cells for both stores. However, Levene's test of Equality of Variances[7] indicated that the differences in variances are insignificant in the full-factorial two-way design for cost estimates in both Asda ($F(3,124) = 2.329$, $p = 0.08$) and Tesco ($F(3,124) = 1.488$, $p = 0.22$). Homogeneity of variances is therefore assumed in both cases. No significant effect of price condition or presentation condition was found for either store. A summary of the two ANOVA models is shown in Tables 2.4 (Asda) and 2.5 (Tesco).

[7] Levene's test is preferred to Bartlett's test as it is less sensitive to departures from normality.

TABLE 2.4
Two-way ANOVA model of cost estimates in Asda

Source	*SS*	*df*	*MS*	*F*	*P*	η2
Price	85.25	1	85.25	0.403	0.53	0.003
Presentation	372.58	1	372.58	1.760	0.19	0.014
Price*Presentation	27.07	1	27.07	0.128	0.72	0.001
Error	26246.19	124	211.66			

TABLE 2.5
Two-way ANOVA model of cost estimates in Tesco

Source	*SS*	*df*	*MS*	*F*	*P*	η2
Price	100.91	1	100.91	0.495	0.48	0.004
Presentation	359.62	1	359.62	1.765	0.19	0.014
Price*Presentation	154.99	1	154.99	0.761	0.39	0.006
Error	25261.18	124	203.72			

The cost estimates in each store were highly correlated, $r(126) = 0.95$, $p<0.001$. ANCOVA models assume that the slopes of the regression lines are the same for each group formed by the categorical variables and measured on the dependent, i.e. homogeneity of correlations. The (Fisher-transformed) correlation coefficients for each cell were compared using a two-tailed t-test for independent samples[8]. The results, summarized in Table 2.6, support the null hypothesis of equality of correlations.

[8] The difference between correlations are tested using t-tests, where $t = \frac{Z(r_1)-Z(r_2)}{\sqrt{\frac{1}{(n_1-3)}+\frac{1}{(n_2-3)}}}$ and $Z(r_i)$ is Fisher's transformation where $Z(r) = 0.5 \ln\left|\frac{1+r}{1-r}\right|$. The degrees of freedom (*df*) are equal to $n_1 - 2 + n_2 - 2$.

TABLE 2.6
Correlation statistics for cost estimates in each store and t-tests for differences between correlation coefficients

	Presentation Condition			
	Paired		Pooled	
Price Condition	r	p	r	p
Magnitude	0.97 (r_1)	<0.001	0.93 (r_2)	<0.001
Frequency	0.98 (r_3)	<0.001	0.95 (r_4)	<0.001
t-test	t	df	p	
r_1 vs r_2	1.473	60	0.146	
r_1 vs r_3	-0.494	60	0.623	
r_1 vs r_4	0.808	60	0.422	
r_2 vs r_3	-1.967	60	0.054	
r_2 vs r_4	-0.665	60	0.509	
r_3 vs r_4	1.302	60	0.198	

A two-way ANCOVA model was used to partial out the variance in Tesco cost estimates explained by the cost estimates in Asda, the effect of which was highly significant (R^2 = 0.92; $F(1,123)$ = 1402.96, p<0.001). As before, Levene's test for Equality of Variances was used to check that the assumption of homogeneity of variances was satisfied ($F(3,124)$ = 0.393, p = 0.76). The adjusted mean cost estimate for Tesco given by participants in the Magnitude condition (M_{adj} = £33.65) was higher than the adjusted mean cost estimate given by participants in the Frequency condition (M_{adj} = £30.34) and the difference was highly significant ($F(1,123)$ = 21.125, p<0.001, η^2=0.15). The adjusted mean cost estimates for Tesco given in the Paired Presentation condition (M_{adj} = £32.06) and the Pooled Presentation condition (M_{adj} = £31.92) were not significantly different ($F(1,123)$ = 0.039, p = 0.84). The interaction between price condition and presentation condition was in the hypothesized direction and approached, but did not reach, statistical significance ($F(1,123) = 3.445$, $p = 0.07$). The results of the ANCOVA model are summarized in Table 2.7.

TABLE 2.7
ANCOVA model of cost estimates in Tesco including cost estimate for Asda as a covariate (Experiments 1 and 2).

Source	*SS*	*df*	*MS*	*F*	*P*	η2
Cost Estimate Asda	23224.99	1	23224.99	1402.96	<0.001	0.92
Price	349.71	1	349.71	21.13	<0.001	0.15
Presentation	0.64	1	0.64	0.04	0.844	0.00
Price*Presentation	57.03	1	57.03	3.45	0.066	0.03
Error	2036.18	123	16.55			

2.4.2 Prior Beliefs

Participant's beliefs about the prices in each store were collected at the start and end of the experiment. Prior beliefs about Asda did not vary significantly between the Magnitude condition ($M = 1.9$) and the Frequency condition ($M = 2.0$) ($t(99) = -0.753$, $p = 0.45$, two-tailed) [9]. Similarly, prior beliefs about Tesco did not vary significantly between the Magnitude condition ($M = 2.2$) and the Frequency condition ($M = 2.2$) ($t(123) = -0.508$, $p = 0.61$, two-tailed) [10]. Hence, the random allocation of participants between the two experimental conditions cannot explain the previous difference in Tesco cost estimates.

The residual of the previous ANCOVA model was significantly correlated with the participants' prior beliefs about prices in Tesco, $r(123) = 0.24$, $p<0.01$. Participants with more favourable prior beliefs about Tesco tended to give lower cost estimates. Including prior belief as an additional covariate in the ANCOVA model

[9] 27 participants, 15 in the Magnitude condition and 12 in the Frequency condition, answered 'Don't know' to the question concerning their prior beliefs about prices in Asda.
[10] 3 participants in the Magnitude condition answered 'Don't know' to the question concerning their prior belief about prices in Tesco so were excluded from the subsequent ANCOVA model.

showed significant effects of both prior beliefs and cost estimates for Asda (R^2 = 0.92; Table 2.8).

TABLE 2.8
ANCOVA model of cost estimates in Tesco including cost estimate for Asda and prior beliefs about Tesco prices as covariates (Experiments 1 and 2).

Source	*SS*	*df*	*MS*	*F*	*P*	η2
Cost Estimate Asda	22690.90	1	22690.90	1417.46	<0.001	0.92
Prior Beliefs Tesco	121.58	1	121.58	7.60	<0.01	0.06
Price	368.30	1	368.30	23.01	<0.001	0.16
Presentation	0.00	1	0.00	0.00	0.99	0.00
Price*Presentation	74.38	1	74.38	4.65	<0.05	0.04
Error	1904.97	119	16.01			

The adjusted mean cost estimate for Tesco given by participants in the Magnitude condition (M_{adj} = £33.68) was higher than the adjusted mean cost estimate given by participants in the Frequency condition (M_{adj} = £30.24). When a Paired Presentation format is used to display price information, the difference between the adjusted mean cost estimates in the Magnitude condition and the Frequency condition (£34.46 vs. £29.47) is greater than the difference when a Pooled Presentation format is used (£32.90 vs. £31.01) and the interaction is significant (F(1,119) = 74.384, p<0.05). The interaction between price condition and presentation condition is illustrated in Fig 2.1.

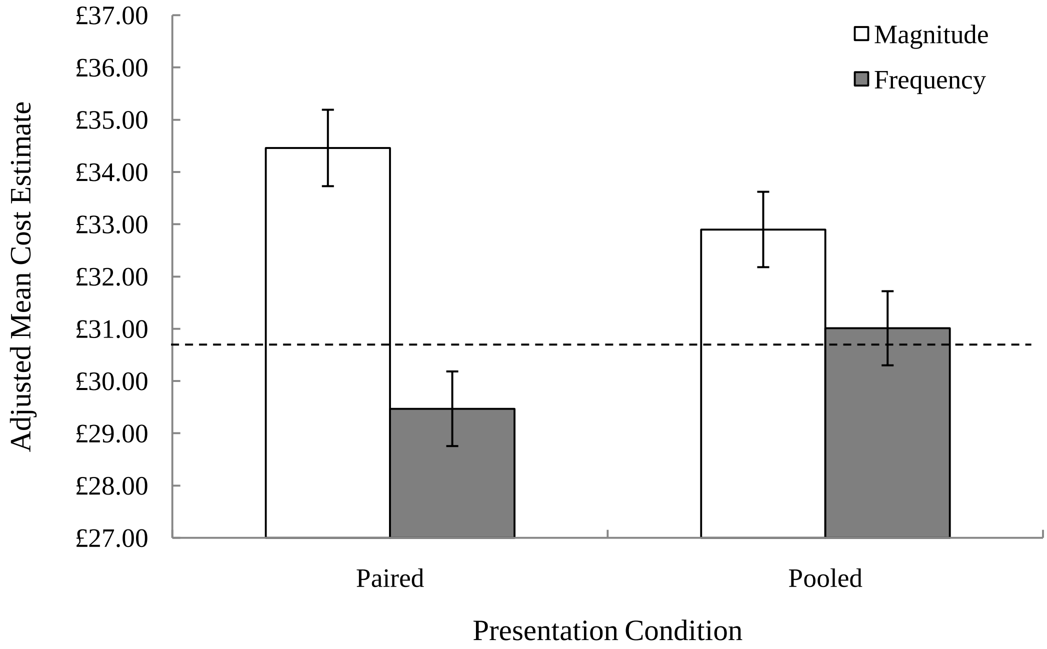

Figure 2.1: Interaction plot of the adjusted mean cost estimates for Tesco in Experiments 1 and 2 (Evaluated at mean cost estimate for Asda = £30.76).

2.4.3 Confidence Judgments

Participants' confidence judgments of which store was cheaper were uncorrelated with prior beliefs about Asda, $r(99) = 0.140$, $p = 0.16$. Similarly, confidence judgments were uncorrelated with prior beliefs about Tesco, r(123) = -0.016, $p = 0.86$. Confidence judgments do not appear to have been significantly influenced by participants' prior beliefs.

A two-way between-subjects ANOVA model was used to test for differences in the mean confidence judgment between price conditions (Magnitude or Frequency) and presentation conditions (Paired or Pooled). The central tendency and spread of confidence judgments are summarized in Table 2.9.

TABLE 2.9
Summary of confidence judgment ratings of the cheapest store across Experiments 1 and 2

	Paired		Pooled	
Price Condition	*M*	*SD*	*M*	*SD*
Magnitude	2.6	1.5	3.3	1.1
Frequency	4.7	1.7	4.1	1.5

Participants' confidence judgments of which store was cheaper differed significantly between the Magnitude and Frequency conditions ($F(1,124) = 30.263$, $p<0.001$, $\eta^2=0.20$). In the Magnitude condition participants were confident that Asda was the cheaper of the two stores ($M = 2.9$, $SEM = 0.185$) while in the Frequency condition participants were weakly confident that Tesco was the cheaper store ($M = 4.4$, $SEM = 0.185$). The 95% confidence intervals indicate that the mean ratings differed significantly from the neutral rating of 4 in both the Magnitude condition ($2.572 < M < 3.303$) and the Frequency condition ($4.009 < M < 4.741$). The ANOVA model also indicated a small but significant interaction between the price and presentation conditions ($F(1,124) = 6.307$, $p<0.05$, $\eta^2=0.05$). When a Paired Presentation format is used to display price information, the difference between the mean confidence judgments in the Magnitude condition and the Frequency condition (2.6 vs. 4.7) is greater than the difference when a Pooled Presentation format is used (3.3 vs. 4.1). The interaction between price condition and presentation condition is illustrated in Fig 2.2 and the ANOVA model ($R^2 = 0.23$) is summarized in Table 2.10.

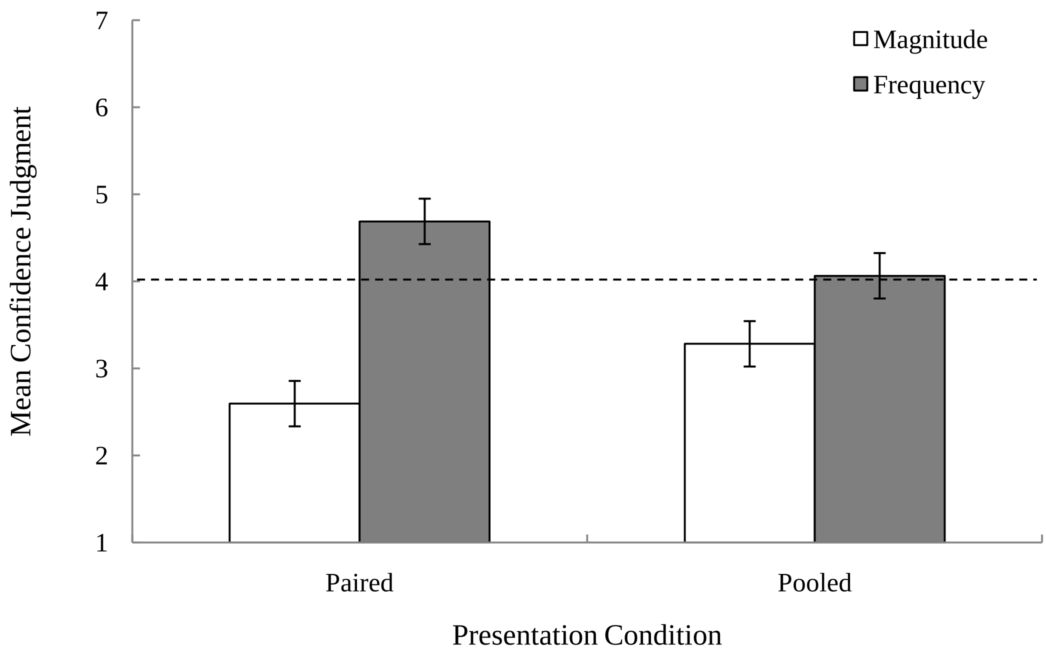

Figure 2.2: Interaction plot of the mean confidence judgments of the cheapest store in Experiments 1 and 2 (Neutral rating = 4).

TABLE 2.10
ANOVA model of confidence judgment ratings of the cheapest store across Experiments 1 and 2.

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Price	66.13	1	66.13	30.26	<0.001	0.20
Presentation	0.03	1	0.03	0.01	0.91	0.00
Price*Presentation	13.78	1	13.78	6.31	<0.05	0.05
Error	270.94	124	2.19			

2.4.4 Manipulation Check

Participants' estimates of how many of the items were cheaper in Tesco were uncorrelated with prior beliefs about Asda, $r(99) = 0.134$, $p = 0.18$. Similarly, estimates of the frequency of Tesco's price advantages were uncorrelated with prior beliefs about Tesco, $r(123) = -0.038$, $p = 0.67$. Frequency estimates do not appear to have been significantly influenced by participants' prior beliefs.

A two-way between-subjects ANOVA model was used to test for differences in the mean frequency estimates between price conditions (Magnitude or Frequency) and presentation conditions (Paired or Pooled). The central tendency and spread of frequency estimates are summarized in Table 2.11.

TABLE 2.11
Summary of the estimates of the frequency of price advantages in Tesco across Experiments 1 and 2

	Paired		Pooled	
Price Condition	*M*	*SD*	*M*	*SD*
Magnitude	11.9	6.0	9.7	4.6
Frequency	17.1	5.8	11.8	5.1

Participants' mean estimates of the number of items that were cheaper in Tesco differed significantly between the price conditions ($F(1,124) = 13.909$, $p<0.001$, $\eta^2=0.10$). In the Magnitude condition participants gave lower estimates ($M = 10.9$) than in the Frequency condition ($M = 14.5$). Participants' mean frequency estimates also differed significantly between the presentation conditions ($F(1,124) = 15.685$, $p<0.001$, $\eta^2=0.11$). In the Paired Presentation format participants gave higher estimates ($M = 14.6$) than in the Pooled Presentation form ($M = 10.8$). The interaction between price and presentation conditions was not significant. The ANOVA model ($R^2 = 0.21$) is summarized in Table 2.12.

TABLE 2.12
ANOVA model of the estimates of the frequency of price advantages in Tesco across Experiments 1 and 2.

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Price	399.03	1	399.03	13.91	<0.001	0.10
Presentation	450.00	1	450.00	15.69	<0.001	0.11
Price*Presentation	69.03	1	69.03	2.41	0.12	0.02
Error	3557.44	124	28.69			

2.5 Discussion

2.5.1 Experimental Manipulation

Participants' estimates of the frequency with which Tesco was cheaper across the thirty items, asked as a check of the experimental manipulation, indicate that the between-subjects experimental manipulations worked as intended. The meta-analysis across the two experiments showed that participants were aware of the frequency cue as they gave higher frequency estimates in the Frequency condition relative to the Magnitude condition. Hence, the price manipulation was successful. Although the meta-analysis did not show a significant interaction between the price and presentation conditions, when the data from the two experiments were analysed separately the mean frequency estimates differed significantly between price conditions only for Experiment 1 (Paired Presentation) and not for Experiment 2 (Pooled Presentation). This suggests that the presentation manipulation reduced the salience and availability of the frequency cue in Experiment 2, although a firm conclusion cannot be drawn from these results.

2.5.2 Support for Hypotheses

The meta-analysis of Experiments 1 and 2 strongly supports Hypothesis 1: that Pooled Presentation of price information would reduce or remove the impact of the frequency cue (or strengthen the impact of the magnitude cue), relative to Paired Presentation. A between-subjects ANCOVA showed that mean total cost estimates in the manipulated store Tesco - when the correlation with total cost estimates for the control store Asda and prior beliefs about Tesco were controlled for – were lower in the Frequency condition than in the Magnitude condition for both presentation formats, but the difference was significantly smaller in the case of Pooled

Presentation. This pattern of results could not be explained by differences in prior beliefs between the experimental conditions. Nonetheless, it is important to note that the difference in cost estimates between the price conditions persisted and did not entirely disappear, despite participants being unable to accurately estimate or use the frequency cue in the case of Pooled Presentation.

A similar pattern of results supports Hypothesis 3, concerning differences between participants' mean confidence judgments of which store was the cheaper of the two. When price information was formatted in a Paired Presentation, participants were confident that Asda was the cheapest store in the Magnitude condition and weakly confident that Tesco was the cheapest store in the Frequency condition. When price information was formatted in a Pooled Presentation, participants were again confident that Asda was the cheapest store in the Magnitude condition (although less confident than in the Paired Presentation) but were undecided as to which store was cheaper in the Frequency condition. Not only are differences between total cost estimates less extreme when prices are presented in a Pooled format, but participants are also less confident about which store is the cheaper. However, it is again important to note that the change of presentation weakened but did not remove the impact of the price manipulation.

The results of the meta-analysis show mixed evidence concerning Hypothesis 2 – that participants would rely more on prior beliefs in the case of Pooled Presentation of price information – and, on balance, weakly support the null hypothesis that prior beliefs play a similar role across presentation formats. Firstly, participants' confidence judgments concerning which was the cheaper store were uncorrelated with prior beliefs. Although the fit of the ANCOVA model of total cost estimates in Tesco was improved by the addition of prior beliefs about Tesco as an additional covariate, when the data from the two experiments were analysed separately, prior beliefs only improved the model fit for Paired Presentation and not for Pooled Presentation. Whilst bearing in mind that the power of statistical tests involving prior beliefs was weakened due to the exclusion of participants using the 'Don't Know' rating – meaning care must be taking in extrapolating conclusions - the results overall show little evidence of participants relying upon prior beliefs and even less evidence of increased reliance upon prior beliefs in the case of Pooled Presentation relative to Paired Presentation of price information.

2.5.3 Implications

The central result of Experiments 1 and 2 is that removing the availability and salience of the frequency cue from comparative price data, via a change in presentation format, weakens but does not remove the effect of the experimental price manipulation upon either total cost estimates or confidence judgments as to which store was the cheaper. Even when the frequency cue was not easily available, participants tended to judge the test store as cheaper when it had many small price advantages compared to the control store and more expensive when it had a few large price advantages. As the total cost in the test store was unchanged between the price conditions, participants' judgments were being systematically biased by the distribution of price differences between the two stores. Alba et al (1994) explained this 'frequency effect' as being due to the salience and availability of the frequency cue in comparative price data. The persistence of a (weakened) frequency effect in the case of pooled price information has both theoretical and practical implications that merit further investigation.

From a theoretical perspective, the result raises the question – which is the topic of this thesis - of what cognitive process or processes underlie participants' overall comparative judgments of two collections of paired items. If a simple frequency cue heuristic is an insufficient explanation of the experimental findings, then some or all of the participants were using an alternative judgment and estimation process. Specifically, given that pooled presentation of price data requires the prices in one or both stores to be encoded and retrieved from memory, what models of judgment and decision-making from memory best explain the observed results? As the literature review in Chapter 1 showed, a number of competing models of decision-making from memory have been proposed, some of which would predict a frequency effect. Fitting different models to the same experimental data can lend support to particular models (in the sense that a 'model' is a hypothesis of the reality of the underlying cognitive process) in two ways. Firstly, goodness-of-fit statistics can show whether one model explains the observed data better than another. Secondly, the fitted parameters can be compared to those found when other researchers have applied the same model to experimental data from a different domain or decision-making context. Consistency in model parameters across domains can be interpreted as supporting the existence of a single judgment process employed in a range of different decisions.

The importance of the theoretical implications is strengthened by the potential practical implications of the findings. Paired presentation of price data is only found in comparative price advertising or on comparison websites. Such presentations (i) are therefore mostly limited to stores that practise such comparative advertising, (ii) are restricted to stores which choose to compare themselves against each other, and (iii) are likely to be seen only infrequently by many consumers.

Pooled presentation of prices is analogous to the experience of sequential exposure to prices in two stores, either through browsing, deliberate comparative price information searching, or shopping in different stores on consecutive visits. It may also be analogous to repeat visits to the same store on different occasions given that item prices change over time. Hence, exposure to comparative price data in a pooled presentation format is likely to be a common and frequent experience for most consumers. The existence of a frequency effect in actual - as opposed to advertised – prices would potentially link the price format of a retailer (e.g. the EDLP and PROMO strategies discussed in Chapter 1) with the price image of that retailer. As outlined in Chapter 1, store price image has consistently been shown to exert a strong influence on consumers' grocery store selection decisions; hence this would have significant economic implications for retailers.

2.5.4 Limitations of Present Study

Whilst the previous discussion highlighted the potential implications of the results of Experiments 1 and 2 and supports the case for further investigation, a number of methodological limitations in the design of the present study mean that strong conclusions cannot be drawn from the data and extrapolated to other (non-experimental) settings. In particular, a number of features of the experimental design do not reflect the real-world experience of comparison shopping, arguably lessening the ecological validity of the findings.

Firstly, total (basket) cost estimates were employed as a measure of price image. This is appropriate when the basket of items is fixed and constant across the two stores. In real-world grocery shopping, the basket of items will differ between trips and between stores because of (i) differences in the range of items stocked; (ii) preference for variety (Ratner, Kahn, & Kahneman, 1999; Simonson, 1990); (iii) different purchase needs on different occasions; (iv) changes in purchase behaviour driven by price differences; and (v) random variation in purchasing behaviour (a "trembling hand"). Relative basket cost estimates may be decoupled from relative store price image judgments when the baskets of items differ between the two stores. This would be particularly important in the case of (iv) above, where basket differences are driven by price differences between the two stores. One can plausibly imagine a scenario in which a store might be perceived as relatively expensive and hence a consumer might choose to buy fewer items or to switch to cheaper alternatives when shopping in that store. In that case, the estimated basket cost in the store judged to be more expensive would be *lower* than the estimated basket cost in the store judged to be cheaper.

Secondly, whilst the Paired Presentation format (especially the fifteen-items-per-page format adopted by Alba et al rather than the item-by-item presentation of Experiment 1) is a realistic reflection of the way in which comparative price advertising might be experienced by consumers, the sequential presentation of Experiment 2 is quite different from the experience of browsing the prices in a real store. Furthermore, in both experiments, the amount of attention given to the price data by participants is likely to be higher than in the real world due to the nature of the task given to the participants. In the current experiments participants were explicitly instructed to pay attention to the prices and to memorize as much information as possible in order to answer subsequent questions. In a real store prices are encountered incidentally as a result of browsing or shopping, and some prices may receive more or less attention than others due to the nature of the specific shopping task.

Thirdly, the thirty item prices used in the current experiments is orders of magnitudes fewer than the tens of thousands of items stocked in a large grocery store. A real-world consumer is likely to be sampling from a much larger distribution of prices, rather than explicitly considering every item. As discussed in Chapter 1, decisions from experience (i.e. serial sampling of information) introduce further potential biases such as under-weighting rare events. Whilst the current experiments allowed many participants to accurately determine information such as the frequency cue, real-world price judgments are likely to involve estimation of such cues, even if those cues are then subsequently used in a simple decision heuristic.

In addition to the concerns regarding the lack of ecological validity, the design of the current studies yields little information about the judgment process followed by participants, or the way in which they processed the information they were presented with. With the possible exception of the manipulation check, all measures collected were *outcome* variables related to the final price estimation judgment. No behavioural *process* measures were collected to record participants' behaviour during the experiment, which might shed light on how they made the subsequent price judgment.

These limitations of Experiments 1 and 2 motivated the design of Experiments 3 and 4, described in the next two chapters. The experimental design chosen was intended to increase the ecological validity of the way price information was encountered and processed, the experimental task, and the outcome variables collected. Additional process variables related to participants' behaviour during the experiment were also collected.

CHAPTER 3

TESTING THE SENSITIVITY OF DISCRIMINATION JUDGMENTS TO CHANGES IN MEAN ITEM PRICE IN AN ONLINE COMPARATIVE SHOPPING TASK (EXPERIMENT 3)

3.1 Introduction

In the previous chapter I demonstrated the persistence of a ‘frequency effect’ in comparative price estimates and judgments, even after the frequency cue is made unavailable by switching from a paired to a pooled format for price information. With the average item price held constant in both stores, the store with frequent small price advantages over the other is perceived as cheaper, while the store with a few large price advantages is perceived as more expensive. A frequency effect in pooled presentation of price data suggests that the frequency of price advantages will exert a significant influence on comparative price judgments when prices in two stores are experienced sequentially, for example in browsing two stores. However, a number of limitations in the experimental design, in particular concerns about the ecological validity of the presentation and task, mean it would be unwise to extrapolate strongly from the experimental results to real-world behaviour. The motivations for Experiment 3 were to design an experimental procedure that overcomes these limitations and then to test the sensitivity of the outcome variables to changes in the input prices. The results from Experiment 3 were intended to be used to calibrate a robust test of the frequency effect (Experiment 4) and to select a sample size with appropriate statistical power.

Experiment 3 simulates the task of browsing and purchasing from two grocery stores, before asking participants to make various judgments about the prices in each store. The experiment was designed to be as ecologically valid as possible in an online experimental setting, but there are a number of reasons why participants' price judgments might be less sensitive than in the real world. Firstly, the experiment is a 'one-shot' experience, while consumers may visit real stores repeatedly over time. Therefore consumers are likely to be both more knowledgeable about prices in real stores and also to be more confident in the accuracy of that knowledge. Secondly, there is no strong extrinsic or economic motive to give the correct response in an experiment, in contrast to real-world shopping where incorrect price judgments may have significant financial consequences. Thirdly, despite the best efforts of the experimenter, an experimental shopping task is a less stimulating and involving experience than shopping in a real store. In a real store items are presented in a visually attractive manner and can be physically handled. Prices may be presented in different fonts, sizes and colours, with attention drawn to promotional prices through visual cues and placement within the store. Consumers may spend half an hour or more browsing and shopping a real store, but are unlikely to spend more than a few minutes browsing a fictional store for an experimental task.

Nonetheless, one would reasonably hypothesize that, ceteris paribus, as prices are lowered in an experimental test store relative to a control store then participants' judgments concerning the prices in the test store will become more favourable (Hypothesis 1). However, there are a variety of reasons why one might expect participants to be unable to distinguish between the two stores when price differences are small. From a normative perspective[11], when the mean prices in the two stores (or rather the mean of a sample of prices drawn from each store) are closer to each other, then confidence that the detected difference is statistically significant rather than a Type I error is lower, and so participants should be more likely to rate the stores as having identical prices. In addition, encoding and retrieval of prices from memory may be a noisy process, which would introduce additional variance and widen the confidence intervals around estimated mean prices. Use of a response scale that only allows integer responses rather than being a true scale measure might also introduce a region in which prices are perceived as different but not sufficiently so to select different response values for the two stores. For these reasons, a just-noticeable difference in item price means is expected to be observed, below which the prices in the two stores are indistinguishable (Hypothesis 2).

3.2 Method

3.2.1 Participants

320 participants (40 men and 280 women, aged between 18 and 65 years, with a mean age of 36 years) took part in the web-based experiment. Participation was restricted to UK citizens. The recruitment was conducted via the iVillage.co.uk website by posting adverts containing a link to the experiment. All participants were entered into a draw for £1,000.

3.2.2 Stimuli

150 items and prices were selected from a local branch of the UK's largest supermarket, Tesco. The selected items are all commonly purchased items available

[11] By "normative", I mean here that I assume participants follow a Signal Detection Theory decision process, in which the decision criterion is held constant across multiple judgments.

in local stores at the time price data was collected. The items were chosen to be representative of the range of items found in a typical large grocery store. Fifteen items from ten product categories were chosen. The item prices (at the time of the experiment, excluding promotional offers) were between 11p and £14.98 with a mean of £1.95. A summary of the price distribution in each product department is given in Table 3.1 and the full item descriptions and and prices are included in Appendices A1 and A2.

TABLE 3.1
Summary of the item price distribution used for the control store in Experiment 3.

Product Department	Items	*M*	*SD*	*Min*	*Max*
Fruit and Vegetables	15	£0.92	£0.49	£0.34	£2.18
Meat and Poultry	15	£5.34	£2.89	£1.93	£11.31
Grocery	15	£0.82	£0.24	£0.28	£1.17
Canned Goods	15	£0.53	£0.49	£0.11	£1.73
Beverages	15	£1.46	£0.86	£0.14	£2.73
Household and Pet Food	15	£1.07	£0.81	£0.37	£3.58
Bakery	15	£0.67	£0.32	£0.23	£1.35
Dairy	15	£0.95	£0.70	£0.29	£3.08
Frozen Foods	15	£1.47	£0.53	£0.78	£2.54
Off-Licence	15	£6.24	£4.19	£0.86	£14.98
	150	£1.95	£2.56	£0.11	£14.98

Each participant saw two sets of prices for two different stores. Both stores contained the same 150 items. All participants saw the same prices in the control store (“Smith’s Supermarket”) but the item prices in the test store (“Jones’ Supermarket”) varied between subjects. Fictional store names were chosen to avoid any influence of prior beliefs on participants’ price judgments. Item descriptions contained a mixture of unbranded items (e.g. “White Seedless Grapes (1 Kg)”), familiar brand names (e.g. “Cadbury’s Dairy Milk (200g)”) and store own-brand items (e.g. “Smith’s Fresh Pure Orange Juice (1 Litre)”).

3.2.3 *Design and Procedure*

The experiment was set up as a between-subjects design with eleven discount conditions: 0%, 1%, 2%, 3%, 4%, 5%, 7.5%, 10%, 15%, 20% and 30%. In each case, all 150 items in the test store (Jones') were discounted by the same percentage amount. Assignment to the discount conditions was random. The presentation order of the two stores (Smith's-Jones' or Jones'-Smith's) was also randomized. The effect of drop-outs and randomization was that data was not evenly spread across the cells, which weakened the power of t-tests for significant differences. The allocation of participants to conditions is shown in Table 3.2.

TABLE 3.2
Allocation of participants to discount conditions and store orders in Experiment 3.

Discount	Smith's first	Jones' first	Total
0%	16	15	31
1%	18	20	38
2%	13	13	26
3%	17	18	35
4%	14	18	32
5%	12	14	26
7.5%	12	12	24
10%	18	7	25
15%	12	14	26
20%	12	13	25
30%	22	10	32
	166	154	320

The experiment was implemented in Adobe Flash, embedded in a standard HTML page. Participants were told that they were taking part in a shopping simulation in two fictional stores. Participants were told to imagine that they had no provisions in their house and that they must buy everything they needed for a week. Participants were instructed to try to buy the items that they would usually purchase, or the closest matching item stocked, but only to buy an item if the price was such that they would buy it on a normal shop given their normal budget. Participants were also informed that they would be unable to progress to the checkout until they had chosen a reasonable number and selection of items. Finally, participants were given detailed instructions on how to use the online store: how to navigate the store, how to add and remove items from the shopping basket, and how to progress to the checkout once the shopping trip was completed. Screen shots were used to illustrate the instructions.

The store was laid out with the store name at the top of the screen and the ten product departments listed down the left-hand side in a fixed order. When a product department is selected then the fifteen items in that department are listed in the centre of the screen in a fixed order, with an item description, unit price, quantity selection buttons and a purchase button. A list format was chosen as prior research has shown that this is best suited to browsing tasks in online shopping (Hong, Thong, & Tam, 2004). Selected items are added to a shopping basket at the bottom of the screen. The item descriptions and quantities were shown in the basket but not the prices. There was also a button next to each item allowing it to be removed from the basket. At the very bottom of the screen was a button which took the participant to the checkout. This button was greyed out and inactive until the participant had selected items from at least five different product categories. A screen-shot of the store is shown in Figure 3.1.

Figure 3.1: Screenshot of the online store used in Experiment 3.

After clicking on the ‘Go To Checkout’ button, participants were asked for their pre-checkout judgment of the prices in the first store, using a 5-item scale similar to the price beliefs measure used in Experiments 1 and 2:

“Compared to other supermarkets that I have shopped at in the last 3 months, on the products that I buy regularly, this supermarket had:

1. Cheaper prices on all
2. Cheaper prices on most
3. Average prices on all
4. More expensive prices on most
5. More expensive prices on all

Participants were then shown a checkout receipt with the details of their purchases. The store name, date and time were shown, followed by the details of the items and quantities purchased, with their total prices. The receipt was scrollable if the list of items was longer than the screen height. The total number of items and total cost were given at the end of the receipt. The total cost was also given in a large red font below the receipt to ensure all participants paid attention to it. The 'Continue' button was greyed out and inactive for 15 seconds. A screenshot of the checkout receipt is shown in Figure 3.2.

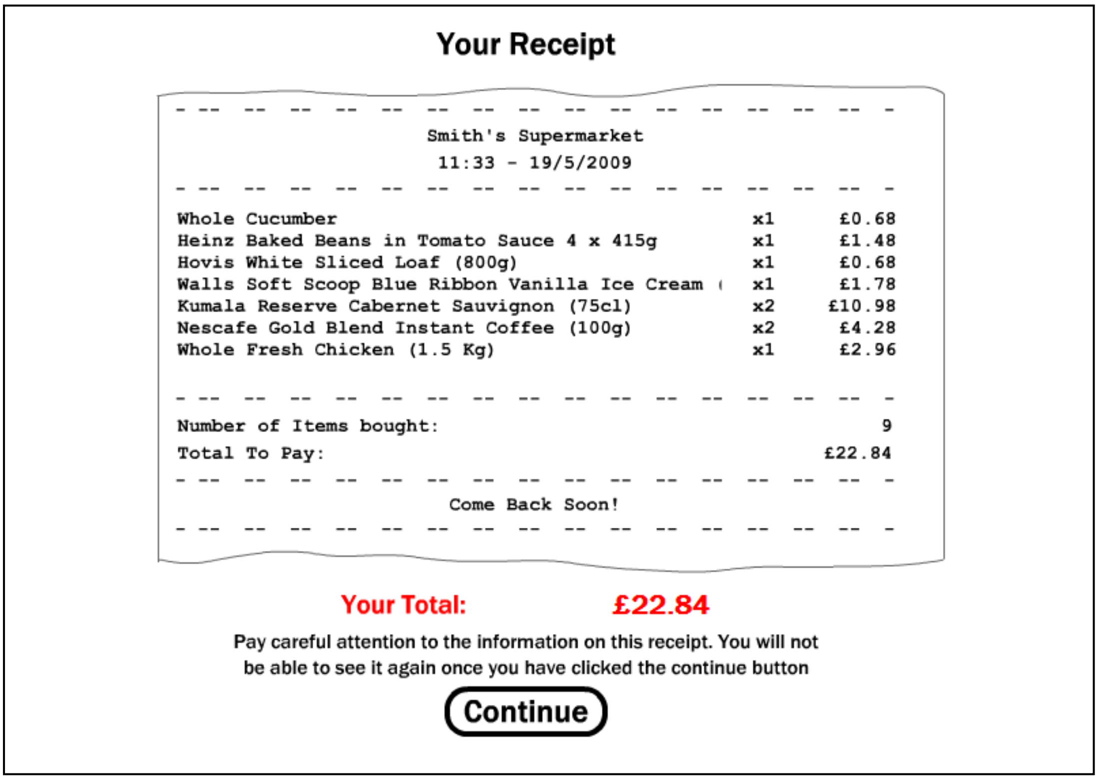

Figure 3.2: Screenshot of the checkout receipt used in Experiment 3.

After viewing the checkout receipt and total basket cost, participants were asked for their post-checkout judgment of the prices in the first store using the same 5-item response scale as before the checkout.

Participants were then given the same task description for the second store. In addition, participants were told not to try and buy exactly the same items as in the first store, as the two stores may stock different items and the prices may differ between the two stores. They then repeated the shopping task in the second store, which was laid out identically to the first store. As before, participants were unable to continue to the checkout until they had selected items from at least five different product categories. After clicking on the ‘Go To Checkout’ button, participants were asked for their pre-checkout judgment of the prices in the second store *relative to the first store*, using a 7-item scale:

“Compared to the first supermarket, I thought that the second store was:”

1. A lot cheaper
2. Cheaper
3. Slightly cheaper
4. About the same on price
5. Slightly more expensive
6. More expensive
7. A lot more expensive

In addition, participants were asked their pre-checkout judgment of the prices in the second store using the same 5-item response scale used for the first store. Participants were also asked to estimate what their basket of items in the second store would cost, in pounds and pence.[12] As before, participants were then shown their checkout receipt and total basket cost. The 7-item relative price judgment and 5-item price judgment were repeated after the checkout in the second store.

[12] A basket cost estimate was not collected before the checkout of the first store to discourage participants from keeping a running total in the second store.

After completing the two shopping tasks and making price judgments about both stores, participants were asked a manipulation check question concerning their estimate of the frequency cue for the second store:

“There were 150 products available in each store. Compared to the first supermarket, I think that the second store was cheaper on:”

1. 0 – 14 products
2. 15 - 29 products
3. 30 - 44 products
4. 45 - 59 products
5. 60 - 74 products
6. 75 - 89 products
7. 90 - 104 products
8. 105 - 119 products
9. 120 - 134 products
10. 135 – 150 products

Finally, participants supplied their gender, age, income, education level, employment status and location. The participant’s e-mail address was collected in order to administer the prize draw.

In addition to the price judgments and manipulation check already described, various behavioural process measures were collected. Participants’ shopping baskets in each store were recorded, as well as the time spent browsing each of the ten product categories in each store. The IP address of each participant was also recorded in order to check for multiple entries.

3.3 Results

3.3.1 Shopping Behaviour

3.3.1.1 Total Spend

There was a wide spread of total spends in both Smith's (*M* = £53.27, *SD* = £33.40) and Jones' (*M* = £49.48, *SD* = £25.31). Two-way ANOVA models were used to test for differences in mean spend for each store between discount conditions and presentation orders (Smith's-Jones' or Jones'-Smith's). The central tendency and spread of total spends are summarized in Table 3.3.

TABLE 3.3
Summary of total spends in Smith's and Jones' (Experiment 3)

Discount Condition	Smith's		Jones'	
	M	*SD*	*M*	*SD*
0%	£45.47	£21.87	£48.67	£25.47
1%	£50.50	£28.67	£49.74	£26.12
2%	£52.59	£35.48	£50.71	£32.63
3%	£53.92	£24.53	£53.48	£24.39
4%	£56.29	£26.45	£55.84	£23.72
5%	£51.55	£50.05	£48.73	£34.84
7.5%	£64.10	£42.19	£58.63	£29.46
10%	£45.45	£18.38	£47.49	£22.24
15%	£54.25	£31.38	£45.61	£18.05
20%	£60.80	£55.37	£44.68	£17.13
30%	£53.62	£24.88	£40.41	£18.33

No significant effect of discount condition or presentation order was found for either store. A summary of the two ANOVA models is shown in Tables 3.4 (Smith's) and 3.5 (Jones').

TABLE 3.4
Two-way ANOVA model of total spend in Smith's (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Discount	8015.36	10	801.54	0.71	0.71	0.02
Order	3258.51	1	3258.51	2.90	0.09	0.01
Discount*Order	9488.61	10	948.86	0.85	0.59	0.03
Error	334569.61	1	1122.72			

TABLE 3.5
Two-way ANOVA model of total spend in Jones' (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Discount	7768.16	10	776.82	1.16	0.28	0.04
Order	743.11	1	743.11	1.21	0.28	0.00
Discount*Order	5131.53	10	513.15	0.80	0.63	0.03
Error	190842.69	1	640.41			

The total spends in each store were highly correlated, $r(318) = 0.756$, $p<0.001$. A two-way ANCOVA model was used to partial out the variance in total spend in Jones' explained by the total spend in Smith's, the effect of which was large and highly significant ($R^2 = 0.61$; $F(1,297) = 416.31$, $p<0.001$, $\eta^2=0.58$). Levene's test for Equality of Variances was used to check that the assumption of homogeneity of variances was satisfied ($F(21,298) = 0.968$, $p = 0.50$). The adjusted mean total spend in Jones' differed significantly between discount conditions ($F(10,297) = 2.846$, $p<0.01$, $\eta^2=0.09$). Planned multiple comparisons between the discount levels show that the adjusted mean spend was significantly lower when prices were discounted by 20% (M = £40.36) or by 30% (M = £38.99) compared to the 0% discount condition (M = £53.20). The results of the ANCOVA model are summarized in Table 3.6, the adjusted mean spends are shown in Table 3.7 and the planned comparisons are shown in Table 3.8.

TABLE 3.6
ANCOVA model of total spend in Jones' including total spend in Smith's as a covariate (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Spend Smith's	111381.17	1	111381.17	416.31	<0.001	0.58
Discount	7613.20	10	761.32	2.85	<0.01	0.09
Order	31.91	1	31.91	0.12	0.73	0.00
Discount*Order	489.51	10	48.95	0.18	1.00	0.01
Error	79461.52	297	267.55			

TABLE 3.7
Summary of adjusted mean spends in Jones' (Experiment 3).

Discount Condition	Adjusted Mean	*SEM*	95% C.I. for Mean	
0%	£53.20	£2.95	£47.40	£59.00
1%	£51.32	£2.66	£46.09	£56.55
2%	£51.10	£3.21	£44.79	£57.41
3%	£53.13	£2.77	£47.68	£58.57
4%	£53.90	£2.92	£48.17	£59.64
5%	£49.78	£3.22	£43.45	£56.11
7.5%	£52.38	£3.35	£45.78	£58.98
10%	£51.77	£3.65	£44.58	£58.95
15%	£45.03	£3.22	£38.70	£51.36
20%	£40.36	£3.28	£33.91	£46.82
30%	£38.99	£3.12	£32.85	£45.13

TABLE 3.8
Pairwise adjusted mean spend comparison results (Experiment 3).

Comparison	Difference	*p*
0% vs. 1%	£1.88	0.635
0% vs. 2%	£2.11	0.629
0% vs. 3%	£0.08	0.985
0% vs. 4%	-£0.70	0.867
0% vs. 5%	£3.42	0.433
0% vs. 7.5%	£0.82	0.854
0% vs. 10%	£1.44	0.759
0% vs. 15%	£8.17	0.062
0% vs. 20%	£12.84	<0.01
0% vs. 30%	£14.21	<0.01

3.3.1.2 Basket Size

There was a wide spread of basket sizes in both Smith's (*M* = 39.1, *SD* = 17.4) and Jones' (*M* = 39.9, *SD* = 17.2). Two-way ANOVA models were used to test for differences in mean basket size for each store between discount conditions and presentation orders. The central tendency and spread of basket sizes are summarized in Table 3.9. No significant effect of discount condition or presentation order was found for either store. A summary of the two ANOVA models is shown in Tables 3.10 (Smith's) and 3.11 (Jones').

TABLE 3.9
Summary of basket sizes in Smith's and Jones' (Experiment 3)

	Smith's		Jones'	
Discount Condition	*M*	*SD*	*M*	*SD*
0%	35.3	14.8	35.9	15.5
1%	34.6	14.4	35.9	15.0
2%	40.2	15.9	40.3	20.6
3%	42.3	19.0	43.3	19.0
4%	41.9	18.1	41.6	16.7
5%	39.7	25.9	38.6	23.7
7.5%	43.8	21.9	43.9	14.8
10%	35.0	14.8	37.6	18.2
15%	38.9	14.7	40.0	13.9
20%	40.2	14.1	42.7	15.5
30%	39.1	16.0	40.8	15.1

TABLE 3.10
Two-way ANOVA model of basket size in Smith's (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Discount	2517.01	10	251.70	0.83	0.60	0.03
Order	268.32	1	268.32	0.89	0.35	0.00
Discount*Order	3719.79	10	371.98	1.23	0.27	0.04
Error	90064.36	298	302.23			

TABLE 3.11
Two-way ANOVA model of basket size in Jones' (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Discount	2226.62	10	222.66	0.74	0.69	0.02
Order	190.37	1	190.37	0.63	0.43	0.00
Discount*Order	1965.83	10	196.58	0.65	0.77	0.02
Error	89542.94	298	300.48			

The basket sizes in each store were highly correlated, $r(318) = 0.822$, $p<0.001$. A two-way ANCOVA model was used to partial out the variance in basket size in Jones' explained by the basket size in Smith's, the effect of which was large and highly significant ($R^2 = 0.70$; $F(1,297) = 645.45$, $p<0.001$, $\eta^2=0.69$). Levene's test for Equality of Variances was used to check that the assumption of homogeneity of variances was satisfied ($F(21,298) = 1.199$, $p = 0.25$). The adjusted mean basket size in Jones' did not differ significantly between discount conditions ($F(10,297) = 0.283$, $p = 0.99$). A small and marginally significant interaction between discount level and presentation order was indicated ($F(10,297) = 2.178$, $p<0.05$, $\eta^2=0.07$) but visual inspection of the profile plot showed no systematic relationship with discount level. The results of the ANCOVA model are summarized in Table 3.12.

TABLE 3.12
ANCOVA model of basket size in Jones' including basket size in Smith's as a covariate (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Basket Smith's	61324.81	1	61324.81	645.45	<0.001	0.69
Discount	268.80	10	26.88	0.28	0.99	0.01
Order	0.08	1	0.08	0.00	0.98	0.00
Discount*Order	2069.59	10	206.96	2.18	<0.05	0.07
Error	28218.14	297	95.01			

3.3.1.3 Total Shopping Time

Repeated-measures ANOVA, with trip (first or second) as a within-subjects factor and discount condition and presentation order as between-subjects factors, was used to test for differences in the mean time spent on each shopping trip. Outlier values >1000 seconds were removed from the analysis.[13] There was a large and significant difference between the mean times spent on each trip ($F(1,294) = 205.74$, $p<0.001$, $\eta^2=0.41$). The mean duration of the first trip ($M = 320$ seconds) was longer than the mean duration of the second trip ($M = 230$ seconds). There was no significant effect of the discount condition ($F(10,294) = 1.815$, $p = 0.06$). The results of the within-subjects tests are shown in Table 3.13 and the between-subjects tests are shown in Table 3.14.

TABLE 3.13
Within-subjects tests of repeated-measures ANOVA of total shopping time (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*P*	η2
Trip	1197711	1	1197711	205.74	<0.001	0.41
Trip*Discount	36630	10	3663	0.63	0.79	0.02
Trip*Order	3505	1	3505	0.60	0.44	0.00
Trip*Discount*Order	51795	10	5179	0.89	0.54	0.03
Error	1711483	294	5821			

TABLE 3.14
Between-subjects tests of repeated-measures ANOVA of total shopping time (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*P*	η2
Discount	454785	10	45479	1.81	0.06	0.06
Order	32691	1	32691	1.30	0.25	0.00
Discount*Order	256253	10	25625	1.02	0.43	0.03
Error	7371959	294	25075			

[13] The web-based experiment had no time limit, so participants could have been interrupted or taken a break during the shopping task and returned to it later.

Participants had lower basket costs in Jones' in the high discount conditions due to the lower item prices. There is no evidence that participants responded to the lower item prices by buying more items in the cheaper store. Participants took less time to shop on their second trip relative to the first, but the prices had no influence on the time they spent shopping in each store.

3.3.2 Price Judgments

3.3.2.1 Comparative Price Judgments

Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and discount condition and presentation order as between-subjects factors, was used to test for differences in the mean price judgment rating of the second store relative to the first store. There was a small and marginally significant difference in mean rating before and after the checkout ($F(1,298) = 5.035$, $p<0.05$, $\eta^2=0.02$). Mean ratings before the checkout are less favourable toward the second store ($M = 3.75$) than the mean rating after the checkout ($M = 3.55$). There was a significant effect of the presentation order ($F(1,298) = 147.784$, $p<0.001$, $\eta^2=0.33$) and the discount level interacted significantly with the effect of the presentation order ($F(10,298) = 12.785$, $p<0.001$, $\eta^2=0.30$). The results of the within-subjects tests are shown in Table 3.15 and the between-subjects tests are shown in Table 3.16. The means plot in Figure 3.3 shows that comparative price judgments concerning the second store tend to improve with discount level when the second store has a lower average price and tend to worsen with discount level when the second store has a higher average price.

TABLE 3.15
Within-subjects tests of repeated-measures ANOVA of comparative price judgments (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Checkout	6.39	1	6.39	5.04	<0.05	0.02
Checkout*Discount	4.53	10	0.45	0.36	0.96	0.01
Checkout*Order	0.05	1	0.05	0.04	0.85	0.00
Checkout*Discount*Order	22.08	10	2.21	1.74	0.07	0.06
Error	378.36	298	1.27			

TABLE 3.16
Between-subjects tests of repeated-measures ANOVA of comparative price judgments (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Discount	19.95	10	2.00	0.99	0.45	0.03
Order	296.82	1	296.82	147.78	<0.001	0.33
Discount*Order	256.78	10	25.68	12.79	<0.001	0.30
Error	598.52	298	2.01			

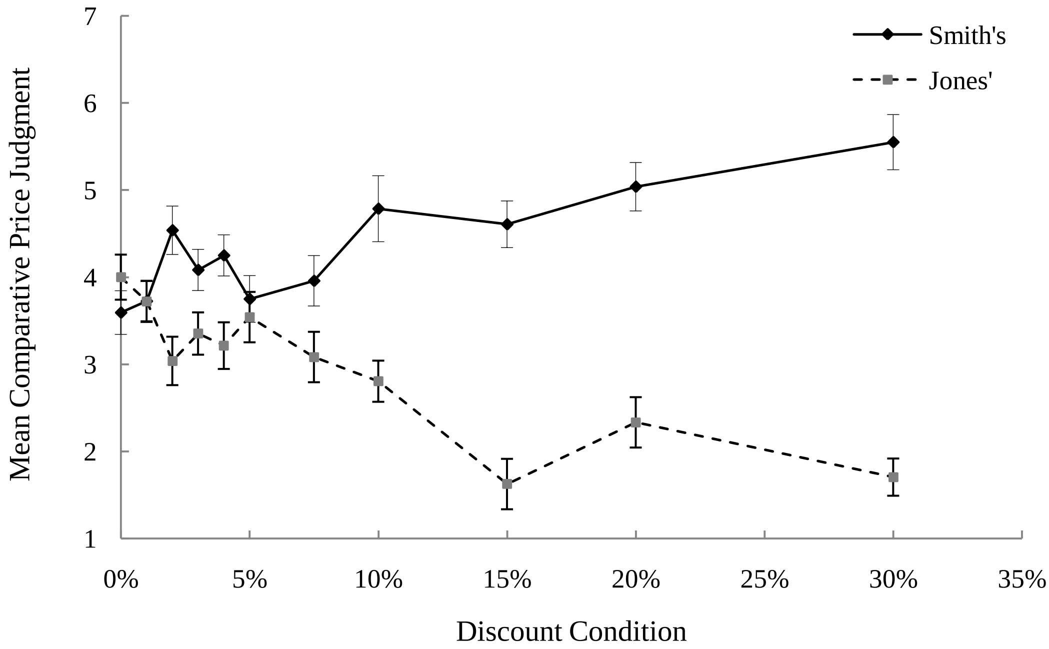

Figure 3.3: Interaction plot of the mean comparative price judgments of the second store relative to the first store across discount conditions and presentation orders (Experiment 3).

The data were re-coded to create a new percentage price decrease categorical variable with 21 levels, where each level represents the decrease in average item price in the second store as a percentage of the average item price in the first store. For example, a 30% discount in Jones' with presentation order Jones'-Smith's was re-coded as a -42.8% decrease (a 42.8% increase in average item price from the first store to the second store). Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and percentage price decrease as a between-subjects factor, was again used to test for differences in the mean price judgment rating of the second store relative to the first store. As before, there was a small and marginally significant difference in mean rating before and after the checkout ($F(1,299) = 4.170$, $p<0.05$, $\eta^2=0.01$). Mean ratings before the checkout are less favourable toward the second store ($M = 3.74$) than the mean rating after the checkout ($M = 3.55$). There was a significant effect of the percentage price decrease ($F(20,299) = 14.052$, $p<0.001$, $\eta^2=0.49$). Planned multiple comparisons between the percentage price decrease levels show that the mean comparative price judgment rating was significantly more favourable when prices decreased by 7.5% ($M = 3.08$), by 10% ($M = 2.81$), by 15% ($M = 1.63$), by 20% ($M = 2.33$) or by 30% ($M = 1.71$) compared to the 0% condition ($M = 3.79$). Similarly, the mean comparative price judgment rating was significantly less favourable when prices increased by 11.1% ($M = 4.79$), by 17.6% ($M = 4.61$), by 25.1% ($M = 5.04$) or by 42.8% ($M = 5.55$) compared to the 0% condition. The results of the within-subjects tests are shown in Table 3.17, the between-subjects tests are shown in Table 3.18, the mean price judgments are shown in Table 3.19 and the planned comparisons are shown in Table 3.20. Based on the pair-wise comparisons, the just-noticeable difference in average price between the two stores can be estimated as being between 7.5% and 10%. The relationship between the percentage price decrease and the mean comparative price judgment is plotted in Figure 3.4.

TABLE 3.17
Within-subjects tests of repeated-measures ANOVA of comparative price judgments using re-coded price conditions (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Checkout	5.282	1	5.282	4.17	<0.05	0.01
Checkout*Decrease	25.616	20	1.281	1.01	0.45	0.06
Error	378.683	299	1.266			

TABLE 3.18
Between-subjects tests of repeated-measures ANOVA of comparative price judgments using re-coded price conditions (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Decrease	282.486	20	14.124	14.05	<0.001	0.49
Error	300.539	299	1.005			

TABLE 3.19
Summary of mean price judgments in the second store (Experiment 3).

Percentage Price Decrease	*M*	*SEM*	95% C.I. for Mean	
-42.8%	5.55	0.32	4.93	6.17
-25.1%	5.04	0.28	4.49	5.59
-17.6%	4.61	0.27	4.08	5.13
-11.1%	4.79	0.38	4.04	5.53
-8.1%	3.96	0.29	3.39	4.53
-5.2%	3.75	0.27	3.22	4.28
-4.2%	4.25	0.24	3.79	4.72
-3.1%	4.08	0.24	3.62	4.55
-2%	4.54	0.28	3.99	5.09
-1%	3.73	0.22	3.28	4.17
0%	3.79	0.18	3.44	4.15
1%	3.72	0.24	3.26	4.19
2%	3.04	0.28	2.49	3.59
3%	3.35	0.24	2.87	3.83
4%	3.21	0.27	2.69	3.74
5%	3.54	0.29	2.97	4.11
7.5%	3.08	0.29	2.51	3.65
10%	2.81	0.24	2.34	3.27
15%	1.63	0.29	1.06	2.20
20%	2.33	0.29	1.76	2.90
30%	1.71	0.21	1.28	2.13

TABLE 3.20

Pairwise comparisons of mean comparative price judgments (Experiment 3).

Comparison	Difference	*p*
0% vs. -42.8%	-1.76	<0.001
0% vs. -25.1%	-1.25	<0.001
0% vs. -17.6%	-0.82	<0.05
0% vs. -11.1%	-1.00	<0.05
0% vs. -8.1%	-0.17	0.62
0% vs. -5.2%	0.04	0.90
0% vs. -4.2%	-0.46	0.12
0% vs. -3.1%	-0.75	0.33
0% vs. -2%	-.293	<0.05
0% vs. -1%	0.07	0.82
0% vs. 1%	0.07	0.82
0% vs. 2%	0.75	0.02
0% vs. 3%	0.44	0.15
0% vs. 4%	0.58	0.08
0% vs. 5%	0.71	0.47
0% vs. 7.5%	0.25	<0.05
0% vs. 10%	0.99	<0.001
0% vs. 15%	2.17	<0.001
0% vs. 20%	1.46	<0.001
0% vs. 30%	2.09	<0.001

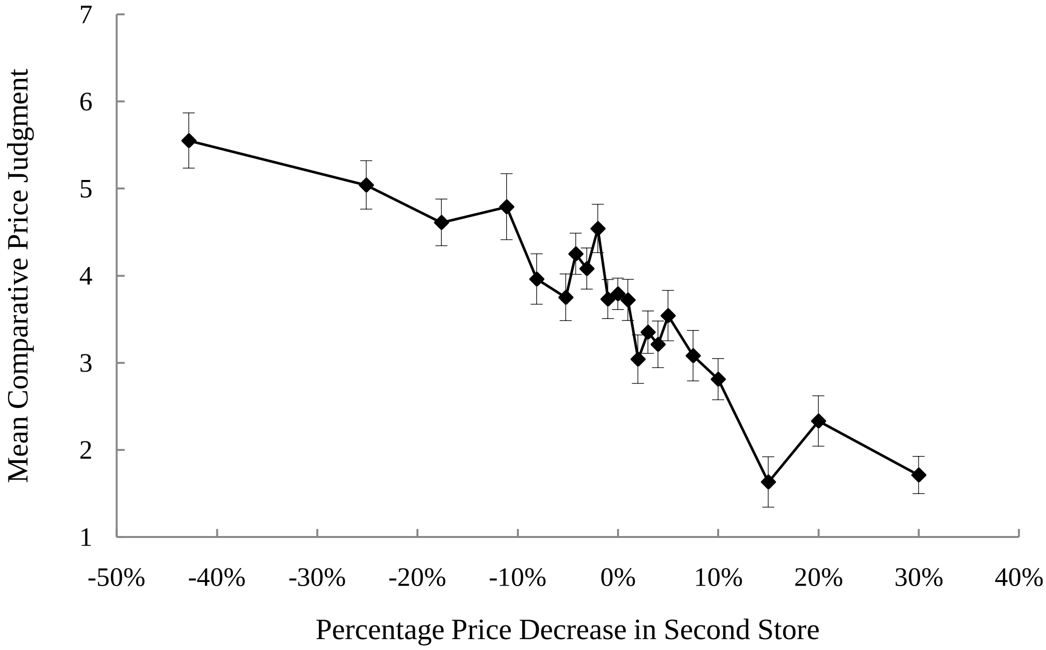

Figure 3.4: Plot of the mean comparative price judgments of the second store relative to the first store across percentage price decrease conditions (Experiment 3).

A more sensitive test of the just-noticeable difference in average price between the two stores was estimated by treating the comparative price judgment rating data as a pair-wise stimulus comparison task. For all cases except the 0% discount level, participants were presented with a pair of stimuli (stores) which differed on an attribute (average item price) by a pre-determined amount (the absolute value of the percentage price change between store 1 and store 2). The participant may or may not have correctly identified that the second store was cheaper or more expensive than the first store. A binary response variable indicating a correct response was coded as:

CORRECT = 1 if RESPONSE = [1,2,3] when PRICE CHANGE<0

CORRECT = 1 if RESPONSE = [5,6,7] when PRICE CHANGE>0

CORRECT = 0 otherwise

The coding was applied to both pre- and post-checkout relative price judgments. The proportion of correct responses for each magnitude of price change is shown in Figure 3.5.

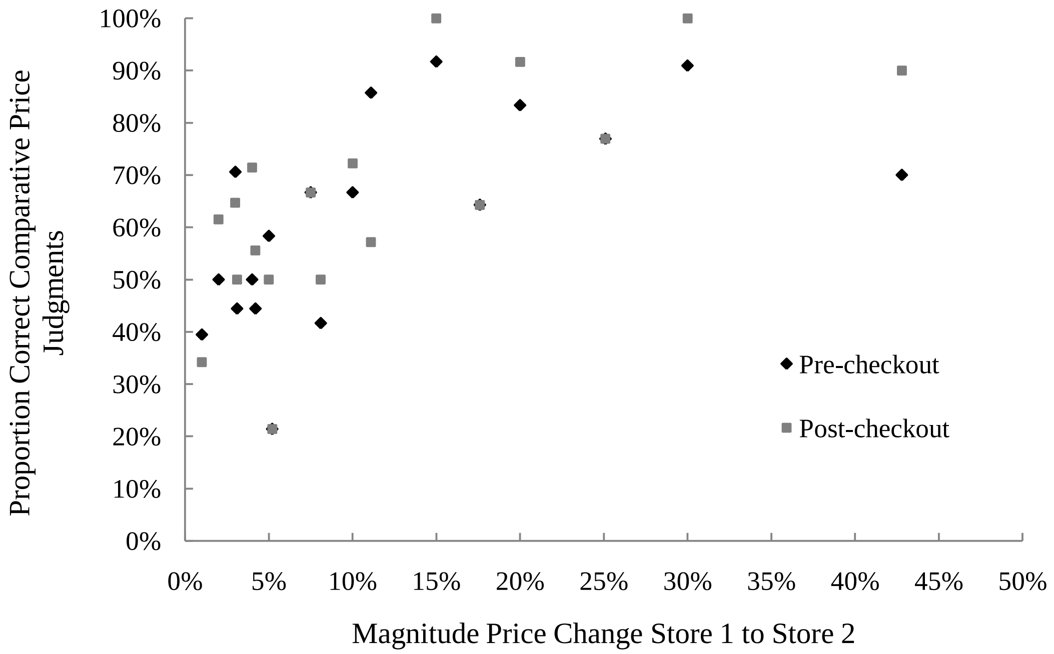

Figure 3.5: Plot of the magnitude of the percentage price change between store 1 and store 2 vs. the proportion of participants correctly identifying the cheaper store (Experiment 3).

A logistic regression model was estimated for the response data, with the binary correct response as the dependent variable and the magnitude of the price change as the independent variable, using an iterative maximum likelihood procedure. A test of the full model versus a model with intercept only was significant, $\chi^2(1, N = 578) = 59.363$, $p<0.001$. With a cut-off of $p = 0.5$, the model was correctly able to classify 83.9% of participants who made a correct relative price judgment and 31.6% of participants who made an incorrect judgment, for an overall success rate of 63.5%. Table 3.21 shows the logistic regression coefficient, Wald test and odds ratio for the model variables. The partial effect of the predictor variable was significant, $\chi^2(1, N = 578) = 44.071$, $p<0.001$. A one point increase in the magnitude of the percentage price change from the first store to the second store increases the probability of a correct relative price judgment by a factor of 1.08. The model predicts that an average item price difference of +/-3% or greater in the second store relative to the first store would be correctly identified by more than half of participants[14]. Hence, a better estimate of the just-noticeable difference in average item prices is 3%. The predicted relationship between the price change and the proportion of correct relative price judgments is shown in Figure 3.6 and the relationship between the observed and predicted proportions of correct judgments is shown in Figure 3.7.

TABLE 3.21
Logistic regression predicting correct comparative price judgment from magnitude of price change between stores (Experiment 3).

Predictor	*B*	*SE*	Wald χ^2	*df*	*p*	*Exp(B)*
Price Change	0.075	0.011	44.07	1	<0.001	1.078
Constant	-0.224	0.124	3.26	1	0.07	0.799

[14] As three of the seven possible responses are 'correct', the base rate for random guessing is 43%.

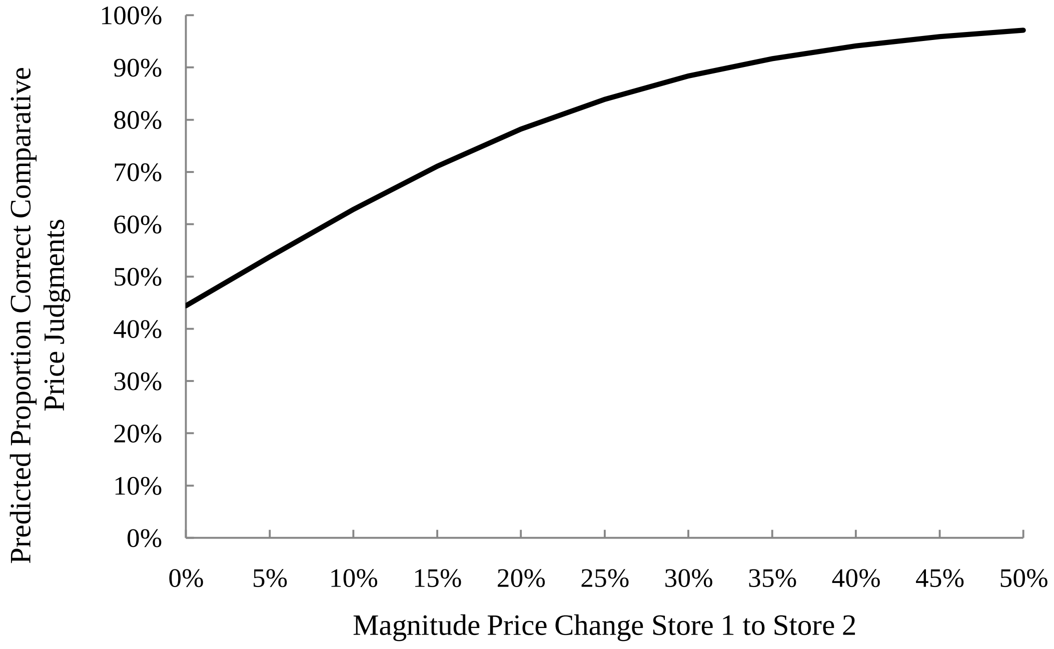

Figure 3.6: Plot of the predicted relationship between the percentage price change between store 1 and store 2 and the proportion of participants correctly identifying the cheaper store based on a binary logistic regression model (Experiment 3).

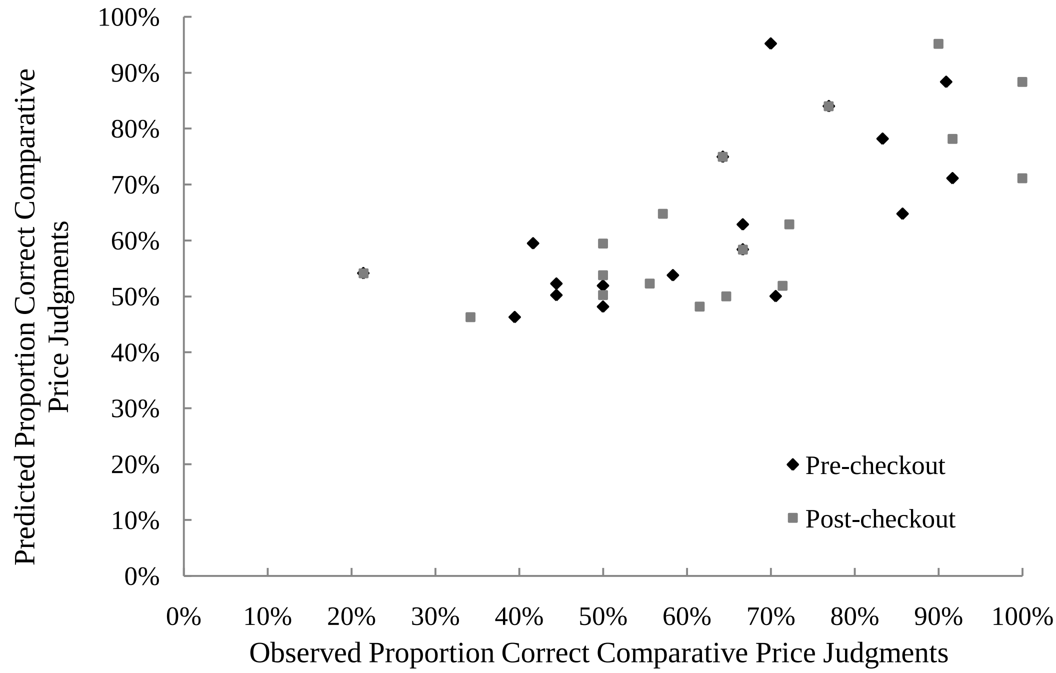

Figure 3.7: Plot of the relationship between the observed and the predicted proportions of participants correctly identifying the cheaper store based on a binary logistic regression model (Experiment 3).

3.3.2.2 Absolute Price Judgments

The absolute price ratings of each store showed a similar pattern of results to the comparative price judgments. Analysis was conducted on the change in ratings between the two stores (a negative difference score implies the second store's prices are perceived more favourably than the first store's prices) both before and after the checkout. The change in pre-checkout ratings is significantly correlated with the pre-checkout comparative price judgments, $r(318) = 0.662$, $p<0.001$. Similarly, the change in post-checkout ratings is significantly correlated with the post-checkout comparative price judgments, $r(318) = 0.664$, $p<0.001$.

Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and discount condition and presentation order as between-subjects factors, was used to test for differences in the mean change in price judgment ratings between the two stores. There was no significant difference in rating changes before and after the checkout ($F(1,298) = 0.514$, $p = 0.47$). There was a significant effect of the presentation order ($F(1,298) = 98.619$, $p<0.001$, $\eta^2=0.25$) and the discount level interacted significantly with the effect of the presentation order ($F(10,298) = 8.738$, $p<0.001$, $\eta^2=0.23$). The results of the within-subjects tests are shown in Table 3.22 and the between-subjects tests are shown in Table 3.23. The means plot in Figure 3.8 shows that changes in absolute price judgments between the two stores tend to become more favourable towards the second store with discount level when the second store has a lower average price and tend to worsen with discount level when the second store has a higher average price.

TABLE 3.22
Within-subjects tests of repeated-measures ANOVA of inter-store changes in absolute price judgments (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Checkout	0.29	1	0.29	0.51	0.47	0.00
Checkout*Discount	9.13	10	0.91	1.60	0.11	0.05
Checkout*Order	0.61	1	0.61	1.08	0.30	0.00
Checkout*Discount*Order	3.93	10	0.39	0.69	0.73	0.02
Error	169.60	298	0.57			

TABLE 3.23
Between-subjects tests of repeated-measures ANOVA of inter-store changes in absolute price judgments (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Discount	7.92	10	0.79	0.68	0.75	0.02
Order	115.47	1	115.47	98.62	<0.001	0.25
Discount*Order	102.32	10	10.23	8.74	<0.001	0.23
Error	348.92	298	1.17			

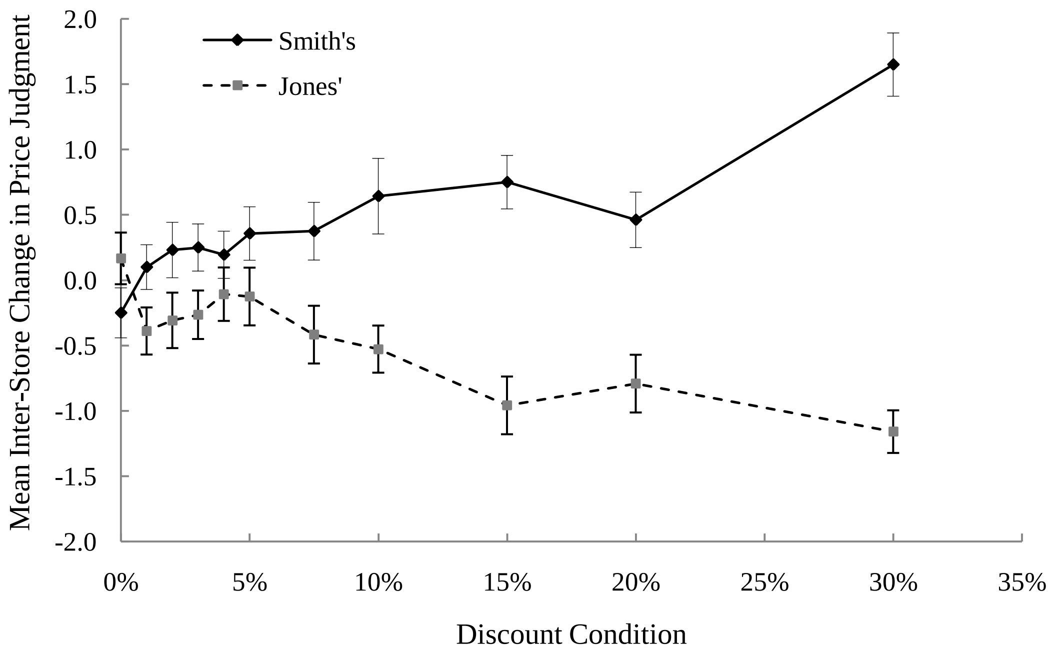

Figure 3.8: Interaction plot of the mean change in absolute price judgments between the two stores across discount conditions and presentation orders (Experiment 3).

As before, the data were re-coded to create a new percentage price decrease categorical variable with 21 levels, where each level represents the decrease in average item price in the second store as a percentage of the average item price in the first store. Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and percentage price decrease as a between-subjects factor, was again used to test for differences in the mean change in absolute price judgment rating between the two stores. Once again, there was no significant difference in mean change in rating before and after the checkout ($F(1,299) = 0.588$, $p = 0.44$). There was a significant effect of the percentage price decrease ($F(20,299) = 9.041$, $p<0.001$, $\eta^2=0.38$). Planned multiple comparisons between the percentage price decrease levels show that the mean change in absolute price judgment rating was significantly more favourable towards the second store when prices decreased by 10% ($M = -0.53$), by 15% ($M = -0.96$), by 20% ($M = -0.79$) or by 30% ($M = -1.16$) compared to the 0% condition ($M = -0.05$). Similarly, the mean change in absolute price judgment rating was significantly less favourable when prices increased by 11.1% ($M = 0.64$), by 17.6% ($M = 0.75$), by 25.1% ($M = 0.46$) or by 42.8% ($M = 1.65$) compared to the 0% condition. The results of the within-subjects tests are shown in Table 3.24, the between-subjects tests are shown in Table 3.25, the mean price judgments are shown in Table 3.26 and the planned comparisons are shown in Table 3.27.

TABLE 3.24
Within-subjects tests of repeated-measures ANOVA of inter-store changes in absolute price judgments using re-coded price conditions (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Checkout	0.334	1	0.334	0.588	0.44	0.00
Checkout*Decrease	12.877	20	0.644	1.135	0.31	0.07
Error	169.617	299	0.567			

TABLE 3.25
Between-subjects tests of repeated-measures ANOVA of inter-store changes in absolute price judgments using re-coded price conditions (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Decrease	212.63	20	10.63	9.04	<0.001	0.38
Error	351.61	299	1.18			

TABLE 3.26
Summary of mean change in price judgments between the two stores (Experiment 3).

Percentage Price Decrease	*M*	*SEM*	95% C.I. for Mean	
-42.8%	1.65	0.24	1.17	2.13
-25.1%	0.46	0.21	0.04	0.88
-17.6%	0.75	0.21	0.35	1.15
-11.1%	0.64	0.29	0.07	1.21
-8.1%	0.38	0.22	-0.06	0.81
-5.2%	0.36	0.21	-0.05	0.76
-4.2%	0.19	0.18	-0.16	0.55
-3.1%	0.25	0.18	-0.11	0.61
-2%	0.23	0.21	-0.19	0.65
-1%	0.10	0.17	-0.24	0.44
0%	-0.05	0.14	-0.32	0.22
1%	-0.39	0.18	-0.75	-0.03
2%	-0.31	0.21	-0.73	0.11
3%	-0.27	0.19	-0.63	0.10
4%	-0.11	0.21	-0.51	0.30
5%	-0.13	0.22	-0.56	0.31
7.5%	-0.42	0.22	-0.85	0.02
10%	-0.53	0.18	-0.88	-0.17
15%	-0.96	0.22	-1.39	-0.52
20%	-0.79	0.22	-1.23	-0.36
30%	-1.16	0.16	-1.48	-0.84

TABLE 3.27
Pairwise comparisons of mean inter-store changes in price judgments (Experiment 3).

Comparison	Difference	*p*
0% vs. -42.8%	1.65	<0.001
0% vs. -25.1%	0.46	<0.05
0% vs. -17.6%	0.75	<0.001
0% vs. -11.1%	0.64	<0.05
0% vs. -8.1%	0.38	0.11
0% vs. -5.2%	0.36	0.10
0% vs. -4.2%	0.19	0.29
0% vs. -3.1%	0.25	0.19
0% vs. -2%	0.23	0.27
0% vs. -1%	0.10	0.50
0% vs. 1%	-0.05	0.14
0% vs. 2%	-0.39	0.31
0% vs. 3%	-0.31	0.35
0% vs. 4%	-0.27	0.81
0% vs. 5%	-0.11	0.77
0% vs. 7.5%	-0.13	0.16
0% vs. 10%	-0.42	<0.05
0% vs. 15%	-0.53	<0.001
0% vs. 20%	-0.96	<0.01
0% vs. 30%	-0.79	<0.001

Based on the pair-wise comparisons, the just-noticeable difference in average price between the two stores can be estimated as being around 10%, indicating that the change in absolute price judgment ratings between stores is a slightly less sensitive measure than the comparative price judgments. The relationship between the percentage price decrease and the mean inter-store change in absolute price judgments is plotted in Figure 3.9.

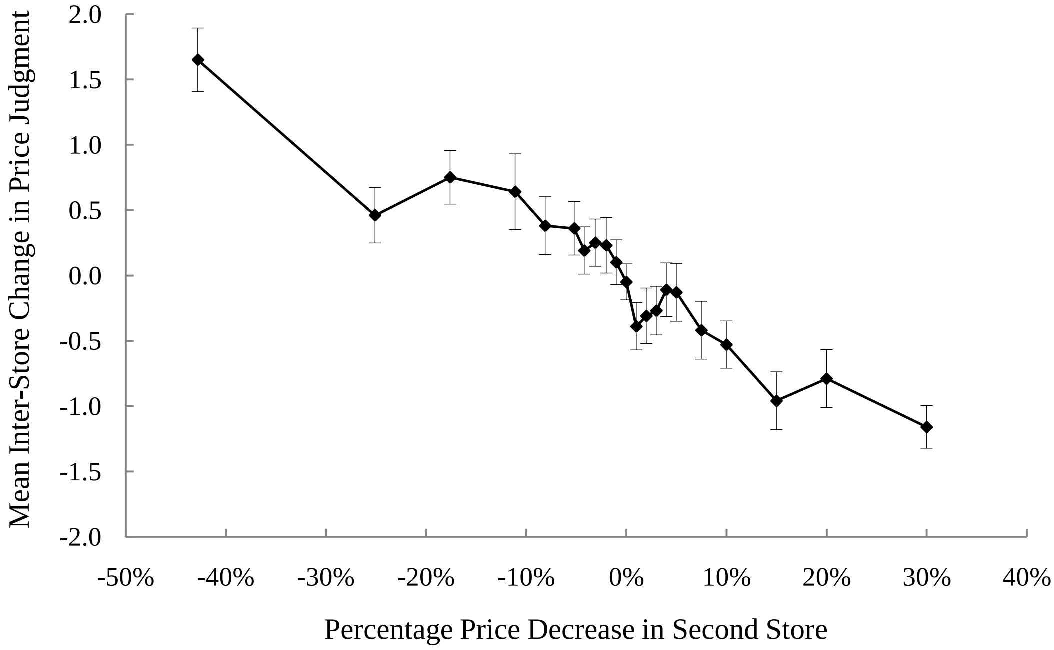

Figure 3.9: Plot of the mean change in absolute price judgments between the first store and the second store across percentage price decrease conditions (Experiment 3).

3.3.2.3 Basket Cost Estimates

There was a wide spread of basket cost estimates in the second store (*M* = £51.05, *SD* = £25.41). A two-way ANOVA model was used to test for differences in mean basket cost estimate between discount conditions and presentation orders. The central tendency and spread of basket cost estimates are summarized in Table 3.28. There was a small and marginally significant difference between the mean basket cost estimates for each store ($F(1,298) = 3.983$, $p<0.05$, $\eta^2=0.01$). The mean basket cost estimate in the discounted store Jones' (*M* = £48.03) was lower than in the control store Smith's (*M* = £54.25). A summary of the ANOVA model is shown in Table 3.29.

TABLE 3.28
Summary of basket cost estimates in the second store (Experiment 3)

	Smith's		Jones'	
Discount Condition	*M*	*SD*	*M*	*SD*
0%	£43.57	£18.38	£53.43	£25.72
1%	£54.77	£31.06	£43.60	£21.81
2%	£48.61	£27.09	£44.25	£19.57
3%	£52.24	£24.04	£53.51	£21.66
4%	£64.53	£23.08	£43.73	£20.27
5%	£52.31	£39.15	£40.48	£20.63
7.5%	£74.85	£47.88	£51.40	£18.88
10%	£48.81	£19.00	£46.40	£19.16
15%	£54.75	£29.63	£50.30	£15.58
20%	£50.50	£31.05	£56.05	£25.04
30%	£48.80	£18.16	£46.73	£16.62

TABLE 3.29
Two-way ANOVA model of basket cost estimates in the second store (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Discount	6080.01	10	608.00	0.96	0.48	0.03
Order	2531.00	1	2531.00	3.98	<0.05	0.01
Discount*Order	7465.65	10	746.57	1.18	0.31	0.04
Error	189345.50	298	635.39			

The basket cost estimates in the second store were strongly correlated with the actual basket cost in the first store, $r(318) = 0.883$, $p<0.001$. A two-way ANCOVA model was used to partial out the variance in basket cost estimates in the second store explained by the actual basket cost in the first store, the effect of which was large and highly significant ($R^2 = 0.82$; $F(1,297) = 1180.93$, $p<0.001$, $\eta^2=0.80$). The adjusted mean basket cost estimates differed significantly between presentation orders ($F(1,297) = 15.25$, $p<0.001$, $\eta^2=0.05$) and the discount level interacted significantly with the presentation order ($F(10,297) = 2.74$, $p<0.05$, $\eta^2=0.09$). The results of the ANCOVA model are summarized in Table 3.30.

TABLE 3.30
ANCOVA model of basket cost estimate in second store including basket cost in first store as a covariate (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Cost First Store	151295.14	1	151295.14	1180.93	<0.001	0.80
Discount	1867.35	10	186.74	1.46	0.16	0.05
Order	1954.12	1	1954.12	15.25	<0.001	0.05
Discount*Order	3513.72	10	351.37	2.74	<0.05	0.09
Error	38050.36	297	128.12			

The means plot in Figure 3.10 shows that adjusted basket cost estimates for the second store tend to decrease with discount level when the second store has a lower average price and tend to increase with discount level when the second store has a higher average price.

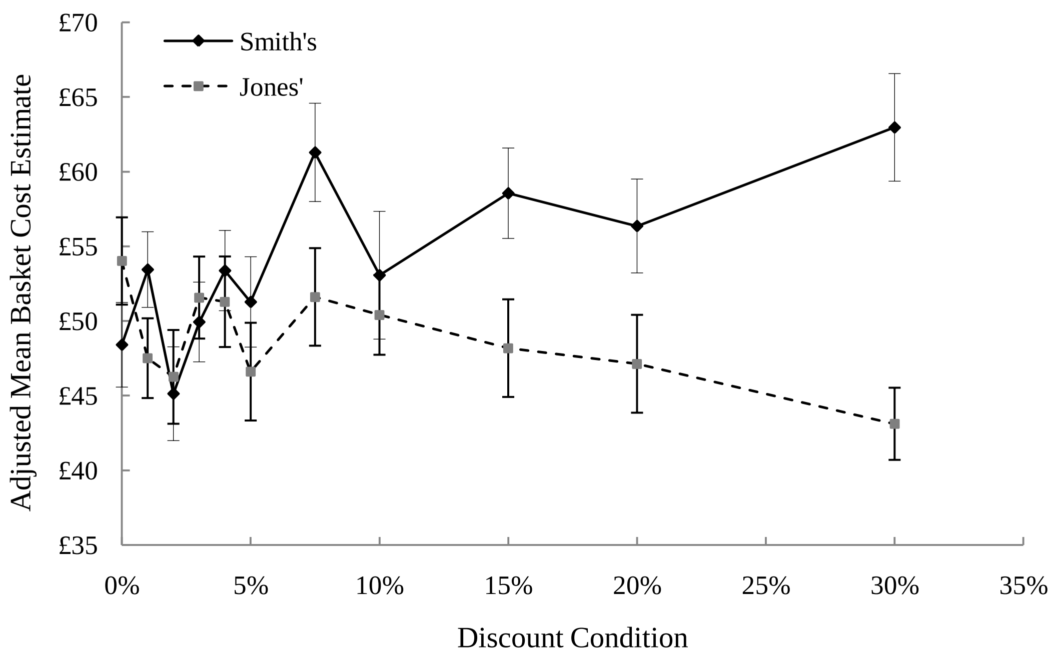

Figure 3.10: Interaction plot of the adjusted mean basket cost estimates for the second store across discount conditions and presentation orders (Experiment 3).

The basket cost estimates in the second store were also strongly correlated with the actual basket cost in the second store, $r(318) = 0.863$, $p<0.001$. A two-way

ANCOVA model was used to partial out the variance in basket cost estimates in the second store explained by the actual basket cost in the second store, the effect of which was large and highly significant (R^2 = 0.77; F(1,297) = 873.85, p<0.001, η^2=0.75). However, the adjusted mean basket cost estimates did not differ significantly between presentation orders (F(1,297) = 0.004, p = 0.95) and the discount level did not interact significantly with the presentation order (F(10,297) = 1.745, p = 0.07). The results of the ANCOVA model are summarized in Table 3.31.

TABLE 3.31
ANCOVA model of basket cost estimates in the second store including the actual basket cost in the second store as a covariate (Experiment 3).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Cost Second Store	141315.65	1	141315.65	873.85	<0.001	0.75
Discount	1727.63	10	172.76	1.07	0.39	0.04
Order	0.63	1	0.63	0.00	0.95	0.00
Discount*Order	2821.99	10	282.20	1.75	0.07	0.06
Error	48029.85	297	161.72			

The results of the two ANCOVA models suggest that participants tended to form their basket cost estimates for the second store by anchoring on the basket cost in the first store and adjusting up or down in response to observed price differences between the two stores. Hence, basket cost estimates tended to be positively related to the discount in the second store and the accuracy of participants' estimates was not systematically worse in high discount conditions.

3.3.3 Manipulation Check

A two-way ANOVA model was used to test for differences in participants' mean estimate of the number of items that were cheaper in the second store between discount conditions and presentation orders. The mean estimate differed significantly between the two stores (F(1,298) = 25.167, p<0.001, η^2=0.08). The estimated proportion of items cheaper was higher in the discounted store Jones' (*M* = 34%) than in the control store Smith's (*M* = 21%). The discount level interacted significantly with the presentation order ($F(10,298) = 2.766$, $p<0.01$, $\eta^2=0.09$). The estimated proportion of items cheaper in the second store tended to increase with the discount level when the second store was cheaper and to decrease with discount level when the second store was more expensive, indicating that participants were more aware of the experimental manipulation in the higher discount conditions. However, even in the 30% discount condition only 7 out of 32 participants correctly identified that all products were cheaper in Jones'. Across all conditions, only 66 out of 320 participants correctly identified the number of items cheaper in the second store. The central tendency and spread of estimates, as well as the proportion of respondents identifying the correct response category, are summarized in Table 3.32.

TABLE 3.32
Summary of the estimated frequency of price advantages in the second store (Experiment 3)

Discount Condition	Smith's *M*	Smith's *SD*	Jones' *M*	Jones' *SD*	Correct Responses
0%	21%	22%	23%	15%	11 / 31 (35%)
1%	31%	20%	35%	25%	4 / 38 (11%)
2%	29%	19%	28%	23%	2 / 26 (8%)
3%	15%	13%	27%	22%	8 / 35 (23%)
4%	22%	16%	26%	21%	5 / 32 (16%)
5%	24%	20%	25%	28%	5 / 26 (19%)
7.5%	23%	27%	27%	23%	5 / 24 (21%)
10%	12%	8%	40%	28%	4 / 25 (16%)
15%	15%	14%	41%	31%	9 / 26 (35%)
20%	21%	17%	38%	26%	6 / 25 (24%)
30%	17%	18%	60%	26%	7 / 32 (22%)

3.4 Discussion

3.4.1 Experimental Paradigm

Measures of participants' behaviour during the experiment indicated that they attempted to complete the shopping tasks as requested and that they behaved in line with expectations for a real-world shopping task. Although basket size and spend varied widely between participants, each participant did not vary their baskets greatly between the two stores, leading to significantly cheaper baskets in the discounted store in high discount conditions. Participants took significantly less time to complete their second shopping task, probably due to familiarity with the interface and task, but were typically exposed to the item prices for 4-5 minutes in each store. Hence, although the shopping tasks were completed far more quickly than one would expect in the real world, all other observed behaviour supports the ecological validity of the experimental paradigm and participants experienced no problems in completing the experimental tasks. The online comparative shopping task paradigm successfully achieved the desired improvements over the simpler experimental paradigm used in Experiments 1 and 2.

3.4.2 Price Judgment Hypotheses

The results of Experiment 3 support both hypotheses concerning comparative price judgments in the online shopping task. Participants' judgments of the prices in the discounted store relative to the control store were more favourable in high discount conditions. This was true for both direct between-store comparisons of the prices and also for differences between separate price judgments of the prices in each store. Hence Hypothesis 1 was supported: participants' price judgments are positively related to the discount level in the manipulated store. However, planned comparisons between the discount levels showed that mean comparative price judgments only differed significantly from the no discount condition when the discount was 7.5% or greater. Inter-store differences in absolute price judgments only differed significantly from the no discount condition when the discount was 10% or greater. Although a more sensitive analysis using binary logistic regression indicated that more than 50% of participants would correctly identify the cheaper store if the average discount was 3% or greater, all analyses showed that participants could not detect small differences between the two stores, even when the discount applied to every single item. Hence Hypothesis 2 was also supported: inter-store price differences exhibit a 'just-noticeable difference' below which prices in the two stores are judged to be indistinguishable.

As suggested in the previous chapter, a distinction should be drawn between participants' price judgments and their basket cost estimates, once an ecologically-valid shopping task is adopted. Although basket cost estimates did show some response to the discount condition, this response was less sensitive to inter-store price differences than the price judgment measures. In addition, there was evidence that participants' basket cost estimates were anchored on the basket cost from the first store. The fact that basket cost estimates were more strongly correlated with the basket cost in the first store than the actual basket cost suggests that participants failed to adjust sufficiently from the anchor in response to price differences between the two stores.

Although participants' comparative price judgments became more favourable toward the cheaper store as the discount level increased, few participants were aware of the experimental manipulation. This suggests either that participants only paid attention to a limited subset of the presented items or that they struggled to recall individual pairs of prices from memory. Thus, similar to the pooled presentation of Experiment 2, the frequency cue appears to be neither salient nor easily available in the online comparative shopping paradigm. The online comparative shopping task is therefore ideally suited to testing for a persistence of the frequency effect found in Experiment 2, in an ecologically valid setting. In the following chapter I describe the results of this test in Experiment 4.

CHAPTER 4

TESTING FOR A FREQUENCY BIAS IN PRICE DISCRIMINATION JUDGMENTS IN AN ONLINE COMPARATIVE SHOPPING TASK (EXPERIMENT 4)

4.1 Introduction

Experiments 1 and 2 demonstrated the persistence of a 'frequency effect' in comparative price estimates and judgments when switching from a paired to a pooled format for price information: with the average item price held constant across two stores, the store with frequent small price advantages over the other is perceived as cheaper, while the store with a few large price advantages is perceived as more expensive. However, limitations in the experimental design, in particular concerns about the ecological validity of the presentation and task, reduced the external validity of the findings. The online comparative shopping paradigm of Experiment 3 overcame those limitations and tested the sensitivity of the outcome variables to changes in the input prices. In this chapter the results from Experiment 3 are used to calibrate a robust test of the frequency effect (Experiment 4) and to select a sample size with appropriate statistical power.

As well as aiming to increase the ecological and external validity of the previous findings, Experiment 4 extends and improves upon Experiments 1 and 2 by disentangling the impacts of the frequency and magnitude cues. In the prior experiments only two test stores were used: a store with a high frequency of low magnitude price advantages and a store with a low frequency of high magnitude price advantages. Although holding average price constant between the control and test stores necessarily implies a negative relationship between the frequency and magnitude cues (as the number of price advantages in the test store is increased, the magnitude of those advantages must fall or the magnitude of the disadvantages must rise in order to offset the fall in average item price) it is possible to hold the frequency constant whilst varying the magnitude of both the price advantages and disadvantages in the test store. The two-way between subjects design of Experiment 4 means that the main effects of frequency and magnitude can be tested separately, as well as any interaction between the two cues.

If the findings of Experiments 1 and 2 extend to the comparative shopping paradigm of Experiment 3, one would hypothesize that a test store with a high frequency of low magnitude price advantages over a control store with the same average item price would be perceived more favourably than a test store with a low frequency of large magnitude price advantages (Hypothesis 1). Holding the frequency of price advantages and disadvantages constant, increasing the magnitude of price differences between the control and test stores should enhance the salience of both positive and negative price differences. Hence, one would hypothesize that increasing the magnitude of price differences whilst holding the frequency of price advantages constant should have no impact upon comparative price judgments (Hypothesis 2). However, assuming that small price differences are less salient or less likely to be accurately recalled by participants, one would expect to see a moderating influence of magnitude upon the frequency cue. When the test store has a high frequency of small price advantages (and therefore a low frequency of large price disadvantages) then increasing the magnitude of all price differences should enhance the salience and availability of the price advantages to a greater extent than the disadvantages. Similarly, when the test store has a low frequency of large price advantages (and therefore a high frequency of small price disadvantages) then increasing the magnitude of all price differences should enhance the salience and availability of the price advantages to a lesser extent than the disadvantages. Hence, one would hypothesize that increasing the magnitude of price differences should strengthen the frequency effect: the difference in price judgments between a low and a high frequency test store should be greater when the magnitude of price differences is larger (Hypothesis 3).

4.2 Calibration of Experimental Design

4.2.1 Magnitude of Price Differences

The results of Experiment 3 showed that when all 150 items were discounted, a statistically significant difference in comparative price judgments - relative to the no discount condition – was observed for a price difference of 7.5% or greater with a sample size of 25-30 participants per condition. A more sensitive test using binary logistic regression indicated that the just-noticeable difference in average item price was lower at just 3%. Given that the frequency of price advantages in Experiment 4 is less than 100%, then the just-noticeable difference in item prices may be higher. Hence, an item price difference of 5% was chosen for the price advantages in the small magnitude conditions. An item price difference of 20% was chosen for the price advantages in the large magnitude conditions.

4.2.2 Power Analysis

A priori power analysis was used to determine an appropriate sample size for Experiment 4, with sufficient power to detect an effect of the experimental manipulation with 95% probability ($\beta = 0.05$). Although the effect size of varying the frequency or magnitude was unknown, the results of Experiment 3 were used to estimate an effect size which would be sufficiently large to be considered a meaningful effect. A linear regression model was estimated for the data from Experiment 3, using the arithmetic mean of the pre- and post-checkout comparative price judgment ratings (*M* = 3.57, *SD* = 1.35) as the dependent variable and the percentage change in average item price between the first and second store as the independent (predictor) variable, using a least-squares procedure. The model fit (R^2 = 0.417) differed significantly from zero ($F(1,318) = 227.06$, $p<0.001$) and the effect of the predictor variable was highly significant (b = 0.062, $t(319) = -15.068$, $p<0.001$). A 10% reduction in average item price in the test store leads to a 0.62 point improvement on the seven-point comparative price judgment rating scale. This corresponds to an effect size (Cohen's f^2) of 0.715[15], which by convention is a large effect.[16] The results of the regression model are summarized in Table 4.1.

TABLE 4.1
Linear regression predicting comparative price judgment rating from magnitude of price change between stores (Experiment 3).

Predictor	B	*SE*	Standardized β	*t*	*p*
Price Change	-0.062	0.004	-0.645	-15.07	<0.001
Constant	3.599	0.058		62.23	<0.001

The effect size of the frequency and magnitude cues was hypothesized to be smaller than the average item price effect as the overall price level of the store is unaffected by changes in either variable. Nonetheless, for an effect to be of practical or substantive interest, it would need to be within an order of magnitude of

[15] Cohen's f^2 is calculated as $\frac{R^2}{1-R^2}$ and tests for a significant deviation of R^2 from zero in multiple regression.
[16] By convention, $f^2 = 0.02$ is a small effect, $f^2 = 0.15$ is a medium effect, and $f^2 = 0.35$ is a large effect, as defined by Cohen (1988, p. 412).

the effect size of a change in the average item price. Hence the appropriate effect size was set at 5% of the average item price effect size, i.e. at $f^2 = 0.0358$, which by convention is a small effect. Power analysis was conducted using the G*Power 3 software (Faul, Erdfelder, Lang, & Buchner, 2007), using the omnibus test for multiple regression. The error probability was set at $\alpha = 0.05$ and the power at $(1-\beta) = 0.95$. Power analysis was conducted for a model with three predictor variables: frequency, magnitude and the interaction frequency*magnitude. The power analysis indicated a minimum sample size of 484 participants with a critical F of 2.623. Given that power analysis for F-tests assumes equal group sizes in the case of ANOVA and Experiment 4 was expected to have unequal group sizes due to randomization and dropouts, a 25% margin was added to the minimum sample size, to give a minimum sample size of 604 participants. The power obtained for a range of sample sizes and effect sizes for a multiple regression with three predictor variables is shown in Figure 4.1.

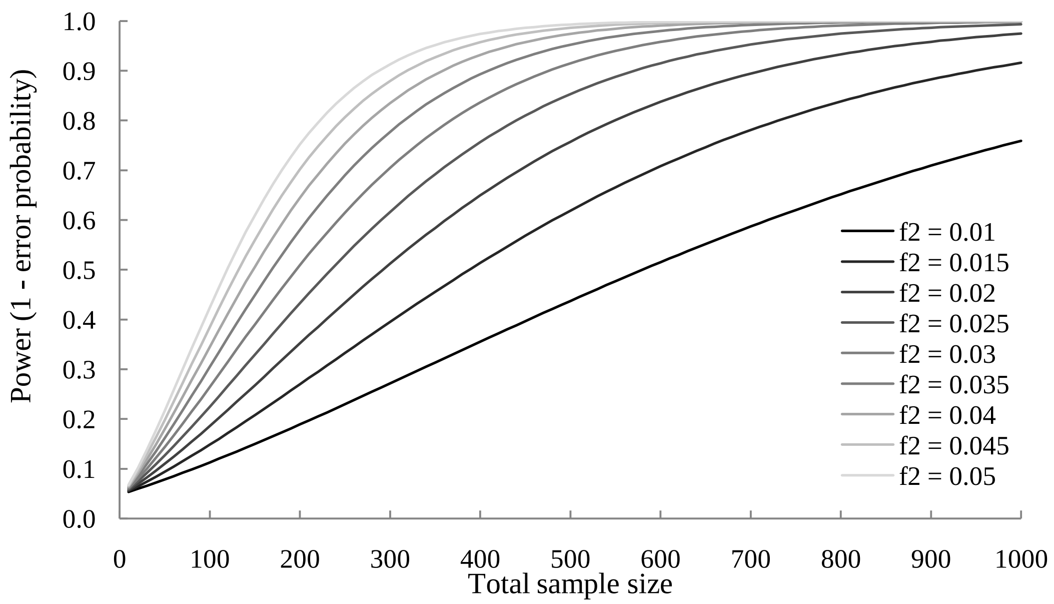

Figure 4.1: Experimental power of an omnibus test of a multiple regression with three predictor variables for a representative range of effect sizes and sample sizes.

4.3 Method

4.3.1 Participants

626 participants (253 men and 373 women, aged between 18 and 60 years, with a mean age of 38 years) took part in the web-based experiment. Participation was restricted to UK citizens. The recruitment was conducted via the iPoints reward scheme (www.ipoints.co.uk). Points collected on the scheme can be exchanged for CDs, flights and other goods, enabling iPoints to maintain a large panel with a good spread across a wide range of demographic variables. A sample of iPoints members were selected at random to receive an e-mail invitation to participate in the experiment. Participants each received 20 iPoints (worth approximately £2) for taking part.

4.3.2 Stimuli

The same 150 commonly purchased items and prices were used as in Experiment 3, with fifteen items from ten product categories. As in Experiment 3, each participant saw two sets of prices for two different stores. Both stores contained the same 150 items. All participants saw the same prices in the control store (“Smith’s Supermarket”) but the item prices in the test store (“Jones’ Supermarket”) varied between subjects. The item prices in the control store were between 11p and £14.98 with a mean of £1.95; the full item descriptions and prices are included in Appendices A1 and A2.

4.3.3 Design and Procedure

The experiment was set up as a 2x2 between-subjects design with two independent variables, *Frequency* and *Magnitude*, systematically varied. The frequency with which the test store was cheaper than the control store was either *Low* (20% of items cheaper in the test store) or *High* (80% of items cheaper in the test store). The magnitude of those price advantages was also either *Small* (5% cheaper) or *Large* (20% cheaper). The reduced items were selected at random, with the same number of items reduced within each product category. In each case, the average item price across the 150 items in the control store (Smith's) was replicated in the test store (Jones') by increasing all the non-reduced item prices by a constant percentage. A summary of the prices in the test store and the relative price distribution across the control and test stores is shown in Table 4.2. Assignment to the four groups was random. The presentation order of the two stores (Smith's-Jones' or Jones'-Smith's) was also randomized. The effect of drop-outs and randomization was that data was not evenly spread across the groups, which weakened the power of t-tests for significant differences. The allocation of participants to groups is shown in Table 4.3.

TABLE 4.2
Experimental design and price distributions for the test stores used in Experiment 4.

Frequency / Magnitude	Low/Small	Low/Large	High/Small	High/Large
Mean item price	£1.95	£1.95	£1.95	£1.95
SD of item prices	£2.52	£2.45	£2.57	£2.87
Minimum item price	£0.11	£0.12	£0.10	£0.09
Maximum item price	£14.23	£11.98	£14.23	£19.40
Number of items cheaper	30	30	120	120
Number of items more expensive	120	120	30	30
Mean price advantage	£0.16	£0.63	£0.10	£0.40
Mean price disadvantage	£0.04	£0.16	£0.40	£1.61

TABLE 4.3
Allocation of participants to frequency and magnitude conditions and store orders in Experiment 4.

Frequency / Magnitude	Smith's first	Jones' first	Total
Low / Small	70	73	143
Low / Large	70	81	151
High / Small	96	77	173
High / Large	82	77	159
	318	308	626

The experimental procedure was identical to Experiment 3. The experiment was implemented in Adobe Flash, embedded in a standard HTML page. Participants were told that they were taking part in a shopping simulation in two fictional stores, and were told to imagine that they have no provisions in their house and that they must buy everything they need for a week. Participants were instructed to try and buy the items that they would usually purchase, or the closest matching item stocked, but only to buy an item if the price was such that they would buy it on a normal shop given their normal budget. Participants were also informed that they would be unable to progress to the checkout until they had chosen a reasonable number and selection of items and were given detailed instructions on how to use the online store.

The store was laid out with the store name at the top of the screen and the ten product departments listed down the left-hand side in a fixed order. When a product department was selected then the fifteen items in that department were listed in the centre of the screen in a fixed order, with an item description, unit price, quantity selection buttons and a purchase button. Selected items were added to a shopping basket at the bottom of the screen. The item descriptions and quantities were shown in the basket but not the prices. After clicking on the 'Go To Checkout' button, participants were asked for their pre-checkout judgment of the prices in the first store, using a five-item scale. Participants were then shown a checkout receipt with the details of their purchases and the total cost was also given in a large red font below the receipt. After viewing the checkout receipt and total basket cost, participants were asked for their post-checkout judgment of the prices in the first store using the same five-item response scale as before the checkout.

Participants were given the same task description for the second store and then repeated the shopping task in the second store, which was laid out identically to the first store. After clicking on the ‘Go To Checkout’ button, participants were asked for their pre-checkout judgment of the prices in the second store *relative to the first store* using a seven-item scale and for their pre-checkout judgment of the prices in the second store using the same five-item response scale used for the first store. Participants were also asked to estimate what their basket of items in the second store would cost, in pounds and pence. As before, participants were then shown their checkout receipt and total basket cost. The seven-item comparative price judgment and five-item price judgment were repeated after the checkout in the second store. After completing the two shopping tasks and making price judgments about both stores, participants were asked a manipulation check question concerning their estimate of the frequency cue for the second store. Finally, participants supplied their gender, age, income, education level, employment status and location. The participant’s e-mail address was collected in order to credit their iPoints account with their incentive payment.

In addition to the price judgments and manipulation check already described, various behavioural process measures were recorded. Participants’ shopping baskets in each store were recorded, as well as the time spent browsing each of the ten product categories in each store. The IP address of each participant was also collected in order to check for multiple entries.

4.4 Results

4.4.1 Shopping Behaviour

4.4.1.1 Total Spend

There was a wide spread of total spends in both Smith's (*M* = £50.74, *SD* = £26.16) and Jones' (*M* = £49.56, *SD* = £24.24). Three-way ANOVA models were used to test for differences in mean spend for each store between frequency and magnitude conditions and between presentation orders (Smith's-Jones' or Jones'-Smith's). The central tendency and spread of total spends are summarized in Table 4.4. No significant effect of frequency condition, magnitude condition or presentation order was found for either store. A summary of the two ANOVA models is shown in Tables 4.5 (Smith's) and 4.6 (Jones').

TABLE 4.4
Summary of total spends in Smith's and Jones' (Experiment 4)

	Smith's		Jones'	
Frequency / Magnitude	*M*	*SD*	*M*	*SD*
Low / Small	£47.79	£22.61	£48.01	£23.01
Low / Large	£51.38	£24.51	£50.58	£22.75
High / Small	£49.70	£23.42	£49.11	£25.36
High / Large	£53.93	£32.56	£50.49	£25.56

TABLE 4.5
Three-way ANOVA model of total spend in Smith's (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	845.32	1	845.32	1.23	0.27	0.00
Magnitude	2327.47	1	2327.47	3.40	0.07	0.01
Order	641.04	1	641.04	0.94	0.33	0.00
Freq*Mag	19.51	1	19.51	0.03	0.87	0.00
Freq*Order	203.03	1	203.03	0.30	0.59	0.00
Mag*Order	274.21	1	274.21	0.40	0.53	0.00
Freq*Mag*Order	31.19	1	31.19	0.05	0.83	0.00
Error	423345.36	618	685.03			

TABLE 4.6
Three-way ANOVA model of total spend in Jones' (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	42.51	1	42.51	0.07	0.79	0.00
Magnitude	669.91	1	669.91	1.14	0.29	0.00
Order	298.65	1	298.65	0.51	0.48	0.00
Freq*Mag	49.27	1	49.27	0.08	0.77	0.00
Freq*Order	1101.39	1	1101.39	1.87	0.17	0.00
Mag*Order	436.13	1	436.13	0.74	0.39	0.00
Freq*Mag*Order	919.21	1	919.21	1.56	0.21	0.00
Error	363849.61	618	588.75			

The total spends in each store were highly correlated, $r(624) = 0.826$, $p<0.001$. A three-way ANCOVA model was used to partial out the variance in total spend in Jones' explained by the total spend in Smith's, the effect of which was large and highly significant ($R^2 = 0.69$; $F(1,617) = 1337.98$, $p<0.001$, $\eta^2=0.68$). Levene's test for Equality of Variances was used to check that the assumption of homogeneity of variances was satisfied ($F(7,618) = 0.262$, $p = 0.97$). The adjusted mean total spend in Jones' did not differ significantly between frequency conditions ($F(1,617) = 1.335$, $p = 0.25$), magnitude conditions ($F(1,617) = 0.661$, $p = 0.42$) or presentation orders ($F(1,617) = 0.24$, $p = 0.88$). The results of the ANCOVA model are summarized in Table 4.7.

TABLE 4.7
ANCOVA model of total spend in Jones' including total spend in Smith's as a covariate (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Spend Smith's	249016.81	1	249016.81	1337.98	<0.001	0.68
Frequency	248.48	1	248.48	1.34	0.25	0.00
Magnitude	122.93	1	122.93	0.66	0.42	0.00
Order	4.56	1	4.56	0.02	0.88	0.00
Freq*Mag	108.30	1	108.30	0.58	0.45	0.00
Freq*Order	495.23	1	495.23	2.66	0.10	0.00
Mag*Order	66.93	1	66.93	0.36	0.55	0.00
Freq*Mag*Order	677.78	1	677.78	3.64	0.06	0.01
Error	114832.80	617	186.12			

4.4.1.2 Basket Size

There was a wide spread of basket sizes in both Smith's ($M = 39.1$, $SD = 19.4$) and Jones' ($M = 38.8$, $SD = 16.9$). Three-way ANOVA models were used to test for differences in mean basket size for each store between frequency and magnitude conditions and between presentation orders. The central tendency and spread of basket sizes are summarized in Table 4.8. No significant effect of frequency condition, magnitude condition or presentation order was found for either store. A summary of the two ANOVA models is shown in Tables 4.9 (Smith's) and 4.10 (Jones').

TABLE 4.8
Summary of basket sizes in Smith's and Jones' (Experiment 4)

	Smith's		Jones'	
Frequency / Magnitude	*M*	*SD*	*M*	*SD*
Low / Small	37.9	18.8	37.1	17.8
Low / Large	39.2	18.2	39.2	17.8
High / Small	38.4	15.9	38.3	16.3
High / Large	40.7	23.9	40.6	15.7

TABLE 4.9
Three-way ANOVA model of basket size in Smith's (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	175.80	1	175.80	0.47	0.50	0.00
Magnitude	513.79	1	513.79	1.36	0.24	0.00
Order	433.83	1	433.83	1.15	0.28	0.00
Freq*Mag	42.50	1	42.50	0.11	0.74	0.00
Freq*Order	6.32	1	6.32	0.02	0.90	0.00
Mag*Order	20.23	1	20.23	0.05	0.82	0.00
Freq*Mag*Order	129.33	1	129.33	0.34	0.56	0.00
Error	233116.79	618	377.21			

TABLE 4.10
Three-way ANOVA model of basket size in Jones' (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	295.07	1	295.07	1.03	0.31	0.00
Magnitude	761.76	1	761.76	2.66	0.10	0.00
Order	246.92	1	246.92	0.86	0.35	0.00
Freq*Mag	1.08	1	1.08	0.00	0.95	0.00
Freq*Order	51.40	1	51.40	0.18	0.67	0.00
Mag*Order	6.38	1	6.38	0.02	0.88	0.00
Freq*Mag*Order	204.89	1	204.89	0.72	0.40	0.00
Error	176775.41	618	286.04			

The basket sizes in each store were highly correlated, $r(624) = 0.813$, $p<0.001$. A three-way ANCOVA model was used to partial out the variance in basket size in Jones' explained by the basket size in Smith's, the effect of which was large and highly significant (R^2 = 0.66; $F(1,617)$ = 1195.07, p<0.001, η^2=0.66). Levene's test for Equality of Variances was used to check that the assumption of homogeneity of variances was satisfied ($F(7,618)$ = 0.492, p = 0.84). The adjusted mean basket size in Jones' did not differ significantly between frequency conditions ($F(1,617)$ = 0.623, p = 0.43), magnitude conditions ($F(1,617)$ = 1.369, p = 0.24) or presentation orders ($F(1,617)$ = 0.010, p = 0.92). The results of the ANCOVA model are summarized in Table 4.11.

TABLE 4.11
ANCOVA model of basket size in Jones' including basket size in Smith's as a covariate (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Basket Smith's	116584.43	1	116584.43	1195.07	<0.001	0.66
Frequency	60.81	1	60.81	0.62	0.43	0.00
Magnitude	133.58	1	133.58	1.37	0.24	0.00
Order	0.97	1	0.97	0.01	0.92	0.00
Freq*Mag	12.75	1	12.75	0.13	0.72	0.00
Freq*Order	29.06	1	29.06	0.30	0.59	0.00
Mag*Order	0.43	1	0.43	0.00	0.95	0.00
Freq*Mag*Order	39.32	1	39.32	0.40	0.53	0.00
Error	60190.98	617	97.55			

4.4.1.3 Total Shopping Time

Repeated-measures ANOVA, with trip (first or second) as a within-subjects factor and frequency and magnitude conditions and presentation order as between-subjects factors, was used to test for differences in the mean time spent on each shopping trip. Outlier values >1000 seconds were removed from the analysis.[17] There was a large and significant difference between the mean times spent on each trip ($F(1,609)$ = 459.94, p<0.001, η^2=0.43). The mean duration of the first trip (*M* =

[17] The web-based experiment had no time limit, so participants could have been interrupted or taken a break during the shopping task and returned to it later.

312 seconds) was longer than the mean duration of the second trip (*M* = 219 seconds). There was a small and marginally significant effect of the frequency condition ($F(1,609) = 4.313$, $p<0.05$, $\eta^2=0.01$). The mean duration in the low frequency conditions (*M* = 256 seconds) was shorter than the mean duration in the high frequency durations (*M* = 275 seconds). The mean shopping time did not differ significantly between magnitude conditions ($F(1,609) = 1.063$, $p = 0.30$) or presentation orders ($F(1,609) = 0.064$, $p = 0.80$). The results of the within-subjects tests are shown in Table 4.12 and the between-subjects tests are shown in Table 4.13.

TABLE 4.12
Within-subjects tests of repeated-measures ANOVA of total shopping time (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Trip	2666716.6	1	2666716.6	459.94	<0.001	0.43
Trip*Freq	1229.0	1	1229.0	0.21	0.65	0.00
Trip*Mag	1581.4	1	1581.4	0.27	0.60	0.00
Trip*Order	5223.1	1	5223.1	0.90	0.34	0.00
Trip*Freq*Mag	28848.0	1	28848.0	4.98	<0.05	0.01
Trip*Freq*Order	792.3	1	792.3	0.14	0.71	0.00
Trip*Mag*Order	241.1	1	241.1	0.04	0.84	0.00
Trip*Freq*Mag*Order	1440.8	1	1440.8	0.25	0.62	0.00
Error	3530993.1	609	5798.0			

TABLE 4.13
Between-subjects tests of repeated-measures ANOVA of total shopping time (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	107708.8	1	107708.8	4.31	<0.05	0.01
Magnitude	26547.3	1	26547.3	1.06	0.30	0.00
Order	1592.1	1	1592.1	0.06	0.80	0.00
Freq*Mag	18165.7	1	18165.7	0.73	0.39	0.00
Freq*Order	17710.2	1	17710.2	0.71	0.40	0.00
Mag*Order	5346.7	1	5346.7	0.21	0.64	0.00
Freq*Mag*Order	236.2	1	236.2	0.01	0.92	0.00
Error	15207385.4	609	24971.1			

Participants' basket costs and basket sizes were similar across frequency and magnitude conditions, with no evidence of a change in shopping behaviour in response to the changes in item prices. Similarly, the mean duration of each shopping trip was not strongly influenced by the experimental manipulation of the prices. Overall, participants' shopping behaviour appears to have been the same across the different price conditions.

4.4.2 Price Judgments

4.4.2.1 Comparative Price Judgments

Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and frequency and magnitude conditions and presentation order as between-subjects factors, was used to test for differences in the mean price judgment rating of the second store relative to the first store. There was no significant difference in mean rating before and after the checkout ($F(1,618) = 0.650$, $p = 0.42$). There was a significant effect of the presentation order ($F(1,618) = 15.537$, $p<0.001$, $\eta^2=0.03$) and the presentation order interacted significantly with both the frequency of price advantages in Jones' ($F(1,618) = 46.176$, $p<0.001$, $\eta^2=0.07$) and the magnitude of those price advantages ($F(1,618) = 1.869$, $p<0.001$, $\eta^2=0.02$). The hypothesized three-way interaction between presentation order, frequency and magnitude was not significant ($F(1,618) = 1.542$, $p = 0.22$). The results of the within-subjects tests are shown in Table 4.14 and the between-subjects tests are shown in Table 4.15.

TABLE 4.14
Within-subjects tests of repeated-measures ANOVA of comparative price judgments (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Checkout	0.802	1	0.802	0.65	0.42	0.00
Checkout*Freq	3.204	1	3.204	2.60	0.11	0.00
Checkout*Mag	0.644	1	0.644	0.52	0.47	0.00
Checkout*Order	0.082	1	0.082	0.07	0.80	0.00
Checkout*Freq*Mag	0.070	1	0.070	0.06	0.81	0.00
Checkout*Freq*Order	0.644	1	0.644	0.52	0.47	0.00
Checkout*Mag*Order	0.497	1	0.497	0.40	0.52	0.00
Checkout*Freq*Mag*Order	6.405	1	6.405	5.20	<0.05	0.01
Error	761.770	618	1.233			

TABLE 4.15
Between-subjects tests of repeated-measures ANOVA of comparative price judgments (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	7.356	1	7.356	3.27	0.07	0.01
Magnitude	0.435	1	0.435	0.19	0.66	0.00
Order	34.906	1	34.906	15.54	<0.001	0.03
Freq*Mag	0.001	1	0.001	0.00	0.98	0.00
Freq*Order	103.741	1	103.741	46.18	<0.001	0.07
Mag*Order	24.419	1	24.419	10.87	<0.001	0.02
Freq*Mag*Order	3.465	1	3.465	1.54	0.22	0.00
Error	1388.421	618	2.247			

The means plot in Figure 4.2 shows that, for participants who saw the control store (Smith's) followed by the test store (Jones'), comparative price judgments about the test store are more favourable (lower) when the test store has a high frequency of price advantages and are less favourable (higher) when the test store has a low frequency of price advantages. Similarly, the means plot in Figure 4.3 shows that, for participants who saw the test store (Jones') followed by the control store (Smith's), comparative price judgments about the test store are also more favourable (higher) when the test store has a high frequency of price advantages and are less favourable (lower) when the test store has a low frequency of price advantages. The mean relative price judgments in the two figures show that the second store was perceived to have lower prices than the first store (i.e. < 4) when that second store has a high frequency of price advantages but the same mean price. The second store was perceived to have the same or higher prices than the first store (i.e. ≥ 4) when that second store has a low frequency of price advantages and the same mean price.

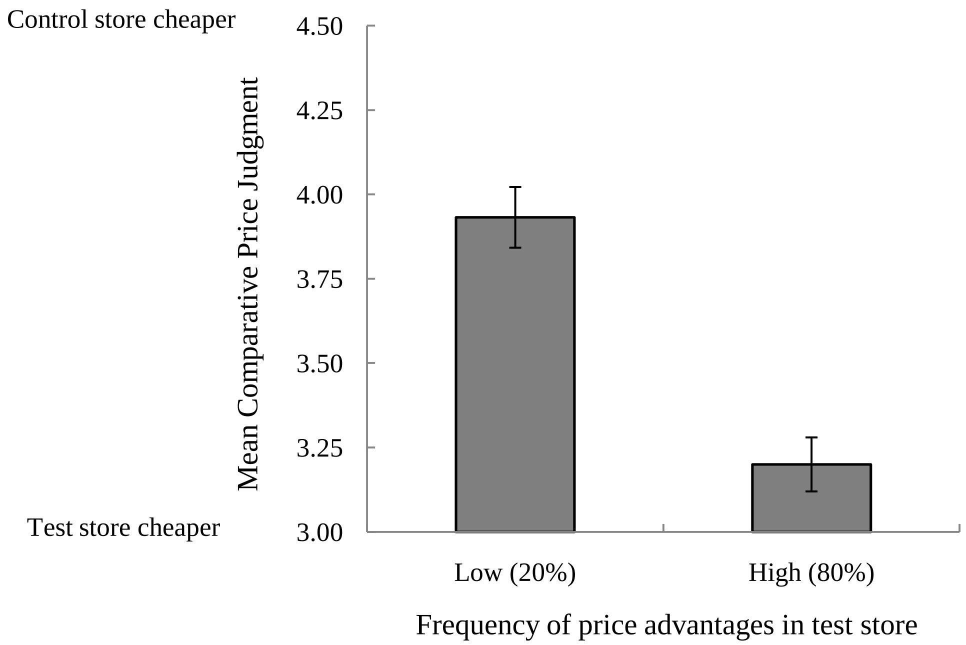

Figure 4.2: Interaction plot of the mean comparative price judgments of the second store relative to the first store across frequency conditions for participants who saw the control store followed by the test store (Experiment 4).

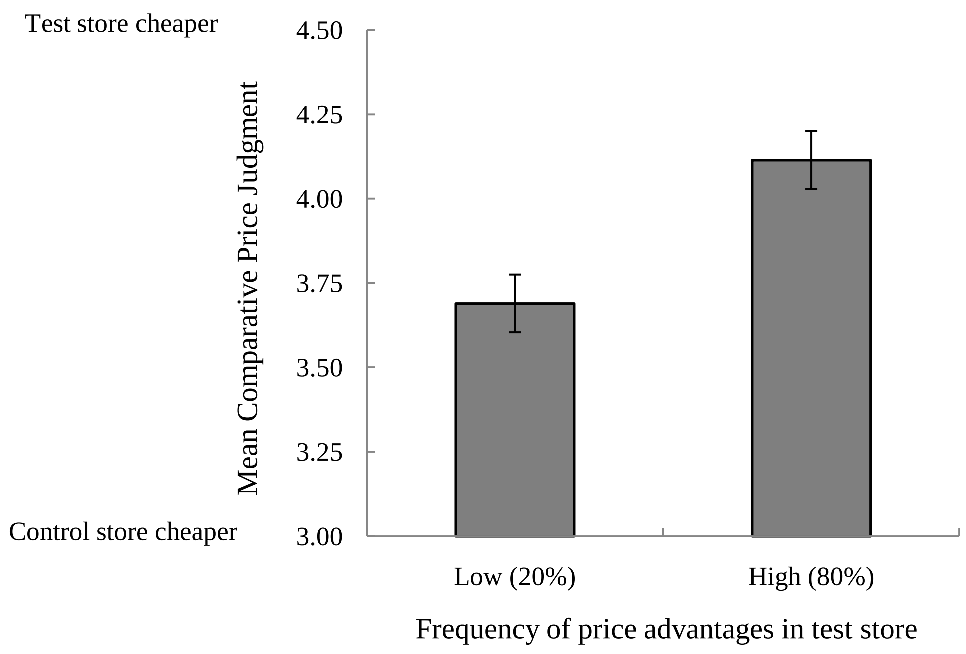

Figure 4.3: Interaction plot of the mean comparative price judgments of the second store relative to the first store across frequency conditions for participants who saw the test store followed by the control store (Experiment 4).

The means plot in Figure 4.4 shows that when price advantages in the test store are small in magnitude, comparative price judgments about the second store are the same regardless of whether the store being rated is the control store or the test store. When price advantages in the test store are large in magnitude, comparative price judgments about the second store are more favourable when the store being rated is the test store and less favourable when the store being rated is the control store.

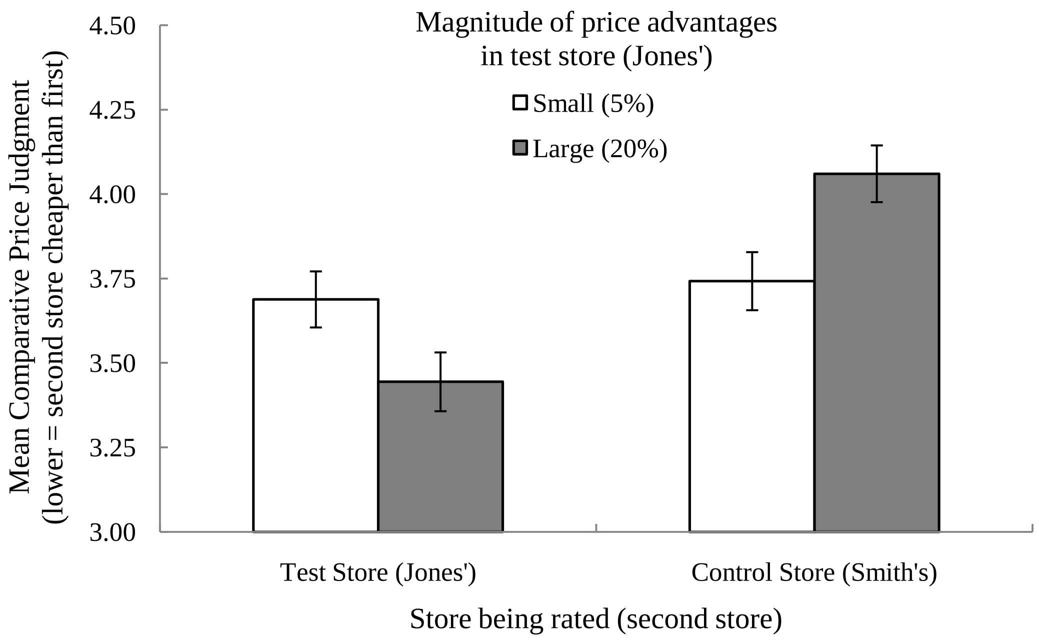

Figure 4.4: Interaction plot of the mean comparative price judgments of the second store relative to the first store across magnitude conditions and presentation orders (Experiment 4).

4.4.2.2 Absolute Price Judgments

The absolute price ratings of each store showed a similar pattern of results to the comparative price judgments, although the ratings were much noisier. Analysis was conducted on the change in ratings between the two stores (a negative difference score implies the second store's prices are perceived more favourably than the first store's prices) both before and after the checkout. The change in pre-checkout ratings was not significantly correlated with the pre-checkout comparative price judgments, $r(624) = 0.040$, $p = 0.31$. However, the change in post-checkout ratings is slightly and significantly correlated with the post-checkout comparative price judgments, $r(624) = 0.092$, $p<0.05$.

Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and frequency and magnitude conditions and presentation order as between-subjects factors, was used to test for differences in the mean change in price judgment ratings between the two stores. There was no significant difference in rating changes before and after the checkout ($F(1,618) = 2.150$, $p = 0.14$). There was a significant effect of the presentation order ($F(1,618) = 7.412$, $p<0.01$, $\eta^2=0.01$) and the frequency condition interacted significantly with the effect of the presentation order ($F(1,618) = 9.620$, $p<0.01$, $\eta^2=0.02$). The magnitude condition did not interact significantly with the presentation order ($F(1,618) = 1.781$, $p = 0.18$) and the hypothesized three-way interaction between frequency, magnitude and presentation order was also insignificant ($F(1,618) = 0.641$, $p = 0.42$). The results of the within-subjects tests are shown in Table 4.16 and the between-subjects tests are shown in Table 4.17.

TABLE 4.16
Within-subjects tests of repeated-measures ANOVA of inter-store changes in absolute price judgments (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Checkout	1.045	1	1.045	2.15	0.14	0.00
Checkout*Freq	0.012	1	0.012	0.02	0.88	0.00
Checkout*Mag	0.035	1	0.035	0.07	0.79	0.00
Checkout*Order	0.612	1	0.612	1.26	0.26	0.00
Checkout*Freq*Mag	0.223	1	0.223	0.46	0.50	0.00
Checkout*Freq*Order	0.322	1	0.322	0.66	0.42	0.00
Checkout*Mag*Order	0.039	1	0.039	0.08	0.78	0.00
Checkout*Freq*Mag*Order	0.843	1	0.843	1.74	0.19	0.00
Error	300.226	618	0.486			

TABLE 4.17
Between-subjects tests of repeated-measures ANOVA of inter-store changes in absolute price judgments (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	13.71	1	13.71	12.86	<0.001	0.02
Magnitude	1.99	1	1.99	1.87	0.17	0.00
Order	7.90	1	7.90	7.41	<0.01	0.01
Freq*Mag	0.65	1	0.65	0.61	0.43	0.00
Freq*Order	10.25	1	10.25	9.62	<0.01	0.02
Mag*Order	1.90	1	1.90	1.78	0.18	0.00
Freq*Mag*Order	0.68	1	0.68	0.64	0.42	0.00
Error	658.50	618	1.066			

The means plot in Figure 4.5 shows that changes in absolute price judgments between the two stores tend to be more favourable towards the second store when the second store is the test store and has a high frequency of price advantages.

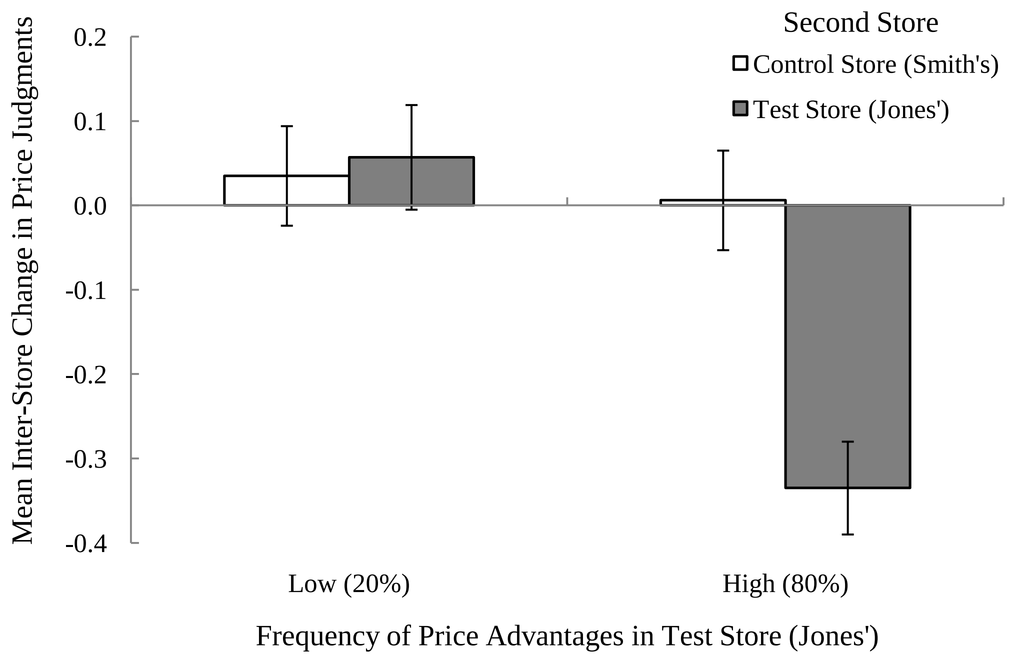

Figure 4.5: Interaction plot of the mean change in absolute price judgments between the two stores across frequency conditions and presentation orders (Experiment 4).

4.4.2.3 Basket Cost Estimates

There was a wide spread of basket cost estimates in the second store (*M* = £49.25, *SD* = £23.12). A three-way ANOVA model was used to test for differences in mean basket cost estimate between frequency and magnitude conditions and presentation orders. The central tendency and spread of basket cost estimates are summarized in Table 4.18. There were no significant differences between stores ($F(1,618) = 1.876$, $p = 0.17$), frequency conditions ($F(1,618) = 0.515$, $p = 0.47$) or magnitude conditions ($F(1,618) = 1.180$, $p = 0.28$). A summary of the ANOVA model is shown in Table 4.19.

TABLE 4.18
Summary of basket cost estimates in the second store (Experiment 4)

	Smith's		Jones'	
Frequency / Magnitude	*M*	*SD*	*M*	*SD*
Low / Low	£48.09	£22.70	£47.83	£25.02
Low / High	£49.85	£19.02	£48.72	£25.53
High / Low	£50.21	£20.25	£47.00	£19.91
High / High	£54.10	£26.91	£48.53	£25.63

TABLE 4.19
Three-way ANOVA model of basket cost estimates in the second store (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	276.27	1	276.27	0.52	0.47	0.00
Magnitude	632.75	1	632.75	1.18	0.28	0.00
Order	1005.56	1	1005.56	1.88	0.17	0.00
Freq*Mag	73.93	1	73.93	0.14	0.71	0.00
Freq*Order	531.48	1	531.48	0.99	0.32	0.00
Mag*Order	101.20	1	101.20	0.19	0.66	0.00
Freq*Mag*Order	21.55	1	21.55	0.04	0.84	0.00
Error	331326.14	618	536.13			

The basket cost estimates in the second store were strongly correlated with the actual basket cost in the first store, $r(624) = 0.904$, $p<0.001$. A three-way ANCOVA model was used to partial out the variance in basket cost estimates in the second store explained by the actual basket cost in the first store, the effect of which was large and highly significant ($R^2 = 0.82$; $F(1,617) = 2797.70$, $p<0.001$, $\eta^2=0.82$). The adjusted mean basket cost estimates did not differ significantly between any of the price conditions. The results of the ANCOVA model are summarized in Table 4.20.

TABLE 4.20
ANCOVA model of basket cost estimate in second store including basket cost in first store as a covariate (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Cost First Store	271458.98	1	271458.98	2797.70	<0.001	0.82
Frequency	60.18	1	60.18	0.62	0.43	0.00
Magnitude	98.16	1	98.16	1.01	0.32	0.00
Order	632.79	1	632.79	6.52	<0.05	0.01
Freq*Mag	1.69	1	1.69	0.02	0.90	0.00
Freq*Order	142.18	1	142.18	1.47	0.23	0.00
Mag*Order	10.99	1	10.99	0.11	0.74	0.00
Freq*Mag*Order	39.67	1	39.67	0.41	0.52	0.00
Error	59867.16	617	97.03			

The basket cost estimates in the second store were also strongly correlated with the actual basket cost in the second store, $r(624) = 0.849$, $p<0.001$. A three-way ANCOVA model was used to partial out the variance in basket cost estimates in the second store explained by the actual basket cost in the second store, the effect of which was large and highly significant ($R^2 = 0.72$; $F(1,617) = 1595.25$, $p<0.001$, $\eta^2=0.72$). Again, the adjusted mean basket cost estimates did not differ significantly between any of the price conditions. The results of the ANCOVA model are summarized in Table 4.21.

TABLE 4.21
ANCOVA model of basket cost estimates in the second store including the actual basket cost in the second store as a covariate (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Cost Second Store	238918.85	1	238918.85	1595.25	<0.001	0.72
Frequency	111.06	1	111.06	0.74	0.39	0.00
Magnitude	0.11	1	0.11	0.00	0.98	0.00
Order	35.28	1	35.28	0.24	0.63	0.00
Freq*Mag	345.11	1	345.11	2.30	0.13	0.00
Freq*Order	6.26	1	6.26	0.04	0.84	0.00
Mag*Order	138.49	1	138.49	0.93	0.34	0.00
Freq*Mag*Order	159.71	1	159.71	1.07	0.30	0.00
Error	92407.29	617	149.77			

The results of the two ANCOVA models suggest that participants tended to form their basket cost estimates for the second store by anchoring on the basket cost in the first store. Neither the degree of adjustment from this anchor nor the accuracy of participants' estimates differed systematically between price conditions.

4.4.3 Manipulation Check

A three-way ANOVA model was used to test for differences in participants' mean estimate of the number of items that were cheaper in the second store between frequency and magnitude conditions and presentation orders. The mean estimate differed significantly between the two stores ($F(1,618) = 5.618$, $p<0.05$, $\eta^2=0.01$) and the frequency of price advantages in the test store interacted significantly with the presentation order ($F(1,618) = 8.097$, $p<0.01$, $\eta^2=0.01$). When the second store was the test store, the estimated proportion of items cheaper in the second store was higher when the actual frequency was high ($M = 30.4\%$) than when the actual frequency was low ($M = 22.6\%$). When the second store was the control store, the estimated proportion of items cheaper in the second store did not differ significantly between frequency conditions. The results indicate that few participants were aware of the experimental manipulation. Even in the extreme case where the second (test) store was cheaper on 80% of the items by 20%, only 7 out of 82 participants correctly identified that between 105 and 134 products were cheaper in the second store. Across all conditions, only 127 out of 626 participants correctly identified the number of items cheaper in the second store. This level of accuracy (20%) is no higher than one would expect if participants had chosen one of the two correct answers at random from the ten available response categories. The central tendency and spread of estimates, as well as the proportion of respondents identifying the correct response category, are summarized in Table 4.22.

TABLE 4.22
Summary of the estimated frequency of price advantages in the second store (Experiment 4)

	Smith's		Jones'		
Frequency / Magnitude	*M*	*SD*	*M*	*SD*	Correct Responses
Low / Low	22%	19%	20%	16%	26 / 143 (18%)
Low / High	25%	18%	25%	18%	28 / 151 (19%)
High / Low	21%	17%	27%	21%	38 / 173 (22%)
High / High	23%	19%	34%	23%	35 / 159 (22%)

4.4.4 Impact of Basket Size

As outlined in Chapter 1, Bell and Lattin (1998) hypothesize that large basket shoppers will tend to prefer stores with an EDLP pricing format (a low average price with little variation) whilst small basket shoppers will tend to prefer stores with a PROMO pricing format (a higher average price with a few deep discounts). Against an intermediate store, an EDLP strategy would result in a high frequency of small magnitude price advantages while a PROMO strategy would result in a low frequency of large magnitude price advantages, analogous to the high and low frequency conditions in Experiment 4. Bell and Lattin argue that (controlling for household size) large basket shoppers shop less frequently and have a higher probability of purchase for any given product category, and as such are more captive to prices across the store as a whole. Small basket shoppers, on the other hand, shop more frequently and are more able to respond to prices in individual product categories. The authors find support for this hypothesis in their empirical analysis of scanner panel data from 1,042 US households.

Assuming that basket size differences in Experiment 4 reflect real-world differences in purchasing behaviour, one might expect to observe a difference in the comparative price judgments of small and large basket shoppers in the online shopping paradigm. Specifically, one would hypothesize that small basket shoppers are better able to respond to specific large magnitude price advantages and hence reduce their overall basket cost, so they should judge the prices in a low frequency, large magnitude store more favourably than a large basket shopper. Conversely, large basket shoppers should judge the prices in a high frequency, small magnitude store more favourably than small basket shoppers.

Basket size was defined as the total number of items purchased in the control store, Smith's. A median split (>37 items) was used to separate participants into small basket shoppers (n = 328) and large basket shoppers (n = 298). The mean number of items purchased in Smith's by small basket shoppers (M = 26.3, SD = 7.4) was smaller than the mean number of items purchased by large basket shoppers (M = 53.2, SD = 18.7) and the difference was highly significant ($t(624) = -24.123$, $p<0.001$, two-tailed). Small basket shoppers' mean spend in Smith's was also significantly lower than large basket shoppers' (£35.49 vs. £67.52, $t(624) = -19.329$, $p<0.001$, two-tailed) and their mean total time spent shopping in Smith's was significantly shorter (239 seconds vs. 324 seconds, $t(624) = -6.164$, $p<0.001$, two-tailed).

Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and frequency condition, presentation order and basket size as between-subjects factors, was used to test for differences in the mean price judgment rating of the second store relative to the first store. As already shown previously, there was no significant difference between mean comparative price judgments before and after the checkout, there was a significant difference in mean ratings between the two stores and the frequency of price advantages in the second store interacted significantly with the presentation order. However, there was no significant difference in mean ratings between basket size groups ($F(1,618) = 2.833$, $p = 0.09$) and the hypothesized interaction between presentation order, frequency condition and basket size was not statistically significant ($F(1,618) = 2.213$, $p = 0.14$). The results of the within-subjects tests are shown in Table 4.23 and the between-subjects tests are shown in Table 4.24.

TABLE 4.23
Within-subjects tests of repeated-measures ANOVA of comparative price judgments including basket size as a between-subjects factor (Experiment 4).

Source	*SS*	*Df*	*MS*	*F*	*p*	η2
Checkout	0.996	1	0.996	0.81	0.37	0.00
Checkout*Freq	2.634	1	2.634	2.14	0.14	0.00
Checkout*Order	0.076	1	0.076	0.06	0.80	0.00
Checkout*Basket Size	1.183	1	1.183	0.96	0.33	0.00
Checkout*Freq*Order	0.848	1	0.848	0.69	0.41	0.00
Checkout*Freq*BSize	0.859	1	0.859	0.70	0.40	0.00
Checkout*Order*BSize	6.237	1	6.237	5.06	<0.05	0.01
Checkout*Freq*Order*BSize	0.042	1	0.042	0.03	0.85	0.00
Error	761.197	618	1.232			

TABLE 4.24
Between-subjects tests of repeated-measures ANOVA of comparative price judgments including basket size as a between-subjects factor (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	7.455	1	7.455	3.29	0.07	0.01
Order	32.475	1	32.475	14.32	<0.001	0.02
Basket Size	6.425	1	6.425	2.83	0.09	0.01
Freq*Order	100.760	1	100.760	44.43	<0.001	0.07
Freq*Basket Size	1.499	1	1.499	0.66	0.42	0.00
Order*Basket Size	3.253	1	3.253	1.43	0.23	0.00
Freq*Order*Basket Size	5.019	1	5.019	2.21	0.14	0.00
Error	1401.677	618	2.268			

4.4.5 Strategic Purchasing

Although no significant difference between large and small basket shoppers' reactions to the price distributions was observed in Experiment 4, the behavioural measures collected allow for a more direct test of the hypothesized relationship between purchasing behaviour and price judgments. Instead of using basket size as a proxy for the ability (or tendency) to vary purchases between stores in response to observed price differences, strategic purchasing behaviour can be directly observed and quantified using the experimental data. Specifically, differences between purchase choices in the two stores can be correlated with item-level inter-store price differences to test for evidence of price-triggered strategic purchasing behaviour. The hypothesized relationships described by Bell and Lattin (1998) would manifest themselves in two ways. Firstly, one would observe a greater degree of strategic additional purchasing when the second store has a low frequency of large magnitude price advantages, as participants respond to the opportunity to make significant savings from the deep discounts. Secondly, one would observe that strategic additional purchasers judge the prices in a low frequency, large magnitude store more favourably while the remaining participants judge the prices in a high frequency, small magnitude store more favourably.

Price-triggered strategic purchasing can occur in two forms. A participant may notice a product that they did not buy on their first trip is cheaper in the second store, and decide to buy it on their second trip (*Addition*). Alternatively a participant may notice that a product that they bought on their first trip is more expensive in the second store and choose not to purchase it on their second trip (*Exclusion*). There are two important points to note. Firstly, if differences between participants' baskets are not driven by price differences but are simply random, one would still expect to see a greater proportion of any additional items being cheaper when the second store has a high frequency of price advantages. Similarly, one would also expect a greater proportion of any excluded items to be more expensive when the second store has a low frequency of price advantages. Secondly, the expected probability of an additional or excluded item being cheaper or more expensive is determined by the basket of items chosen in the first store, rather than the frequency of price advantages in the second store as a whole. For instance, if a participant happened to select a set of items on their first trip that were all more expensive in the second store then it is impossible to prove whether or not a product was discarded because of price differences or for some other reason. If instead only a few of the items selected on the first trip happen to be more expensive in the second store but these particular items are the ones that are not purchased again, then it is more likely that the item selection differences are driven by inter-store price differences.

Two variables were created to measure the degree to which differences between the two baskets of items selected by a participant are driven by each of the two strategies described above. The first measure is the difference between the *observed* and *expected* probabilities of selecting an additional item on the second trip that is also cheaper in the second store. The observed probability for participant *i* is given by the number of additional purchases that are cheaper in the second store divided by the total number of additional items:

$$P_{obs,i} = \frac{\sum_{k=1}^{N}\left(1 - \mathrm{First}_{i,k}\right) \times \mathrm{Second}_{i,k} \times \mathrm{Cheaper}_{k}}{\sum_{k=1}^{N}\left(1 - \mathrm{First}_{i,k}\right) \times \mathrm{Second}_{i,k}}$$

Where:

$$\mathrm{First}_{i,k} = \begin{cases} 1, & \text{if participant } i \text{ purchased item } k \text{ in first store} \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{Second}_{i,k} = \begin{cases} 1, & \text{if participant } i \text{ purchased item } k \text{ in second store} \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{Cheaper}_{k} = \begin{cases} 1, & \text{if item } k \text{ is cheaper in second store} \\ 0, & \text{otherwise} \end{cases}$$

The expected probability for participant *i* is the probability that an item selected at random from all items not purchased on the first trip is cheaper in the second store. This is given by the number of products not selected in the first store that are also cheaper in the second store divided by the total number of items not selected in the first store:

$$P_{exp,i} = \frac{\sum_{k=1}^{N}\left(1 - \mathrm{First}_{i,k}\right) \times \mathrm{Cheaper}_{k}}{\sum_{k=1}^{N}\left(1 - \mathrm{First}_{i,k}\right)}$$

The difference between these two probabilities is an indication of the degree to which additional purchases are driven by price differences rather than random variation:

$$\mathrm{Addition}_i = P_{obs,i} - P_{exp,i}$$

A score that is significantly different from zero means that the observed behaviour is unlikely to have occurred by chance. A positive value indicates price-triggered strategic additional purchases. A negative value indicates irrational or random behaviour. For participants who made no additional purchases in the second store ($P_{\mathrm{obs},i} = 0$) the score was set to zero.

Similarly, the second measure is the difference between the observed and expected probabilities of excluding a previously purchased item on the second trip that is also more expensive in the second store. The observed probability for participant i is the number of non-repeated purchases that are more expensive in the second store divided by the total number of excluded items:

$$P_{obs,i} = \frac{\sum_{k=1}^{N} \mathrm{First}_{i,k} \times (1 - \mathrm{Second}_{i,k}) \times \mathrm{Expensive}_k}{\sum_{k=1}^{N} \mathrm{First}_{i,k} \times (1 - \mathrm{Second}_{i,k})}$$

Where:

$$\mathrm{First}_{i,k} = \begin{cases} 1, & \text{if participant } i \text{ purchased item } k \text{ in first store} \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{Second}_{i,k} = \begin{cases} 1, & \text{if participant } i \text{ purchased item } k \text{ in second store} \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{Expensive}_{k} = \begin{cases} 1, & \text{if item } k \text{ is more expensive in second store} \\ 0, & \text{otherwise} \end{cases}$$

The expected probability for participant i is the probability that an item selected at random from all items purchased on the first trip is more expensive in the second store. This is given by the number of products selected in the first store that are also more expensive in the second store divided by the total number of items selected in the first store:

$$P_{exp,i} = \frac{\sum_{k=1}^{N} \text{First}_{i,k} \times \text{Expensive}_k}{\sum_{k=1}^{N} \text{First}_{i,k}}$$

The difference between these two probabilities is an indication of the degree to which non-repeated (excluded) purchases are driven by price differences rather than random variation:

$$\text{Exclusion}_i = P_{obs,i} - P_{exp,i}$$

A score that is significantly different from zero means that the observed behaviour is unlikely to have occurred by chance. A positive value indicates price-triggered strategic non-repeated purchases. A negative value indicates irrational or random behaviour. For participants who did not exclude any previously-purchased products in the second store ($P_{\text{obs},i} = 0$) the score was set to zero.

Across all participants in Experiment 4, the mean value of Addition (M = 0.020, SD = 0.187) was significantly greater than zero ($t(625) = 2.617$, $p<0.01$, two-tailed). Similarly, the mean value of Exclusion (M = 0.042, SD = 0.201) was significantly greater than zero ($t(625) = 5.290$, $p<0.001$, two-tailed). Participants appear to have engaged in both forms of strategic purchasing behaviour, and were more likely to exclude an item that was more expensive in the second store than to purchase additional items that were cheaper in the second store. This could be because participants were better able to remember the prices of items they actually purchased in the first store, or because they were more sensitive to losses than gains (Tversky & Kahneman, 1992). The two strategic purchasing scores were uncorrelated, $r(624) = 0.029$, $p = 0.48$. The Addition scores were uncorrelated with comparative price judgments about the second store, both before the checkout ($r(624) = -0.042$, $p = 0.29$) and after the checkout ($r(624) = -0.033$, $p = 0.41$). The

Exclusion scores were significantly correlated with comparative price judgments about the second store, both before the checkout ($r(624) = -0.132$, $p<0.001$) and after the checkout ($r(624) = -0.134$, $p<0.001$). The negative correlation implies that, perhaps counter-intuitively, participants who excluded previously-purchased items from their purchases in the second store because those items were more expensive nonetheless gave a more favourable price judgment of the second store relative to the first store.

Three-way ANOVA models were used to test for differences in participants' mean degree of strategic purchasing (Addition and Exclusion) between frequency and magnitude conditions and presentation orders. The mean Addition scores did not vary significantly across any of the price conditions. The mean Exclusion scores differed significantly between the two stores ($F(1,618) = 21.004$, $p<0.001$, $\eta^2=0.03$) and the frequency of price advantages in the test store interacted significantly with the presentation order ($F(1,618) = 6.812$, $p<0.01$, $\eta^2=0.01$). The mean degree of strategic exclusion of previously-purchased items was greatest when the second store was the test store with a high frequency of small magnitude price advantages (M = 0.10). The results of the two ANOVA models are summarized in Table 4.25 (Addition) and Table 4.26 (Exclusion), and the interaction plot in Figure 4.6 shows how the Exclusion scores varied across frequency conditions and presentation orders. This result explains the previously observed negative correlation between Exclusion behaviour and comparative price judgments of the second store: when the second store has a high frequency of small magnitude price advantages (and hence a low frequency of large price disadvantages) the participants are more likely to exclude specific items that are significantly more expensive but also to give a more favourable comparative rating of the store overall. This suggests that participants' comparative price judgments are based on a wider selection of item prices than just the few large item price differences that the participant noticed and adjusted their purchasing behaviour in response to.

TABLE 4.25
Three-way ANOVA model of Addition measure of strategic purchasing (Experiment 4).

Source	*SS*	*Df*	*MS*	*F*	*p*	η2
Frequency	0.119	1	0.119	3.41	0.07	0.01
Magnitude	0.034	1	0.034	0.97	0.33	0.00
Order	0.118	1	0.118	3.40	0.07	0.01
Freq*Mag	0.001	1	0.001	0.04	0.85	0.00
Freq*Order	0.021	1	0.021	0.61	0.44	0.00
Mag*Order	0.063	1	0.063	1.80	0.18	0.00
Freq*Mag*Order	0.000	1	0.000	0.00	0.98	0.00
Error	21.525	618	0.035			

TABLE 4.26
Three-way ANOVA model of Exclusion measure of strategic purchasing (Experiment 4).

Source	*SS*	*Df*	*MS*	*F*	*p*	η2
Frequency	0.012	1	0.012	0.31	0.58	0.00
Magnitude	0.138	1	0.138	3.62	0.06	0.01
Order	0.804	1	0.804	21.00	<0.001	0.03
Freq*Mag	0.130	1	0.130	3.40	0.07	0.01
Freq*Order	0.261	1	0.261	6.81	<0.01	0.01
Mag*Order	0.021	1	0.021	0.56	0.45	0.00
Freq*Mag*Order	0.096	1	0.096	2.51	0.11	0.00
Error	23.658	618	0.038			

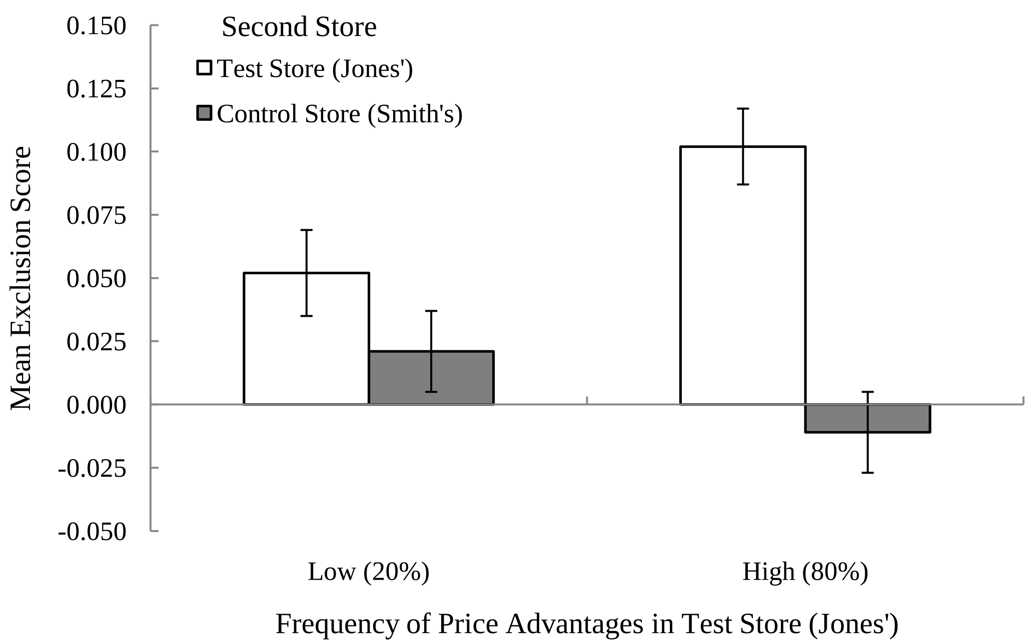

Figure 4.6: Interaction plot of the mean Exclusion measure of strategic purchasing across frequency conditions and presentation orders (Experiment 4).

The mean Addition scores for small basket shoppers (M = 0.013, SEM = 0.011) and large basket shoppers (M = 0.027, SEM = 0.010) did not differ significantly ($t(624) = -0.985$, $p = 0.33$, two-tailed). The mean Exclusion scores for small basket shoppers (M = 0.050, SEM = 0.012) and large basket shoppers (M = 0.034, SEM = 0.011) were also not significantly different ($t(624) = 1.002$, $p = 0.32$, two-tailed). Adding basket size as a covariate in the previous three-way ANOVA models showed no improvement in fit and the basket size variable was not statistically significant in either ANCOVA model. Overall, there was no evidence for the posited relationship between basket size and strategic purchasing behaviour. Participants strategically altered their purchasing behaviour in response to inter-store item-level price differences and the degree of strategic purchasing behaviour varied depending upon the distribution of inter-store price differences, but small basket shoppers do not appear to have taken advantages of inter-store price differences to a greater extent than large basket shoppers in Experiment 4.

In order to test for the second hypothesized effect, participants were divided into two equal size groups based upon the Addition measure, with participants who scored less than or equal to 0.021 classified as non-Additional purchasers (n = 313) and participants who scored more than 0.021 classified as Additional purchasers (n = 313). The mean Addition score of the non-Additional purchasers (M = -0.12, SEM = 0.008) was lower than the mean score of the Additional purchasers (M = 0.16, SEM = 0.007) and the difference was highly significant ($t(624) = -26.442$, $p<0.001$, two-tailed). Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and frequency condition, presentation order and Addition purchasing group as between-subjects factors, was used to test for differences in the mean price judgment rating of the second store relative to the first store. As already shown previously, there was no significant difference between mean comparative price judgments before and after the checkout, but there was a significant difference in mean ratings between the two stores and the frequency of price advantages in the second store interacted significantly with the presentation order. Furthermore, the Addition group interacted significantly with the presentation order ($F(1,618) = 4.984$, $p<0.05$, $\eta^2=0.01$) and there was a significant three-way interaction between the presentation order, frequency condition and Addition group ($F(1,618) = 4.763$, $p<0.05$, $\eta^2=0.01$). The interaction plots show that, as before, the mean comparative price judgments are always more favourable when the second store has a high frequency of price advantages than when the second store has a low frequency of price advantages. For both presentation orders the difference in comparative price judgment rating between high and low frequency conditions is more extreme for

Additional purchasers than for non-Additional purchasers. The results of the within-subjects tests are shown in Table 4.27, the between-subjects tests are shown in Table 4.28, and the three-way interaction is shown in Figure 4.7.

TABLE 4.27
Within-subjects tests of repeated-measures ANOVA of comparative price judgments including Addition purchasing group as a between-subjects factor (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Checkout	0.537	1	0.537	0.43	0.51	0.00
Checkout*Freq	2.641	1	2.641	2.13	0.15	0.00
Checkout*Order	0.189	1	0.189	0.15	0.70	0.00
Checkout*Addition Group	0.002	1	0.002	0.00	0.97	0.00
Checkout*Freq*Order	0.839	1	0.839	0.68	0.41	0.00
Checkout*Freq*AdGrp	1.416	1	1.416	1.14	0.29	0.00
Checkout*Order*AdGrp	0.132	1	0.132	0.11	0.74	0.00
Checkout*Freq*Order*AdGrp	2.321	1	2.321	1.87	0.17	0.00
Error	765.369	618	1.238			

TABLE 4.28
Between-subjects tests of repeated-measures ANOVA of comparative price judgments including Addition purchasing group as a between-subjects factor (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	6.527	1	6.527	2.89	0.09	0.01
Order	32.327	1	32.327	14.31	<0.001	0.02
Addition Group	0.514	1	0.514	0.23	0.63	0.00
Freq*Order	101.935	1	101.935	45.13	<0.001	0.07
Freq*Addition Group	0.091	1	0.091	0.04	0.84	0.00
Order*Addition Group	11.259	1	11.259	4.98	<0.05	0.01
Freq*Order*AdGrp	10.760	1	10.760	4.76	<0.05	0.01
Error	1396.018	618	2.259			

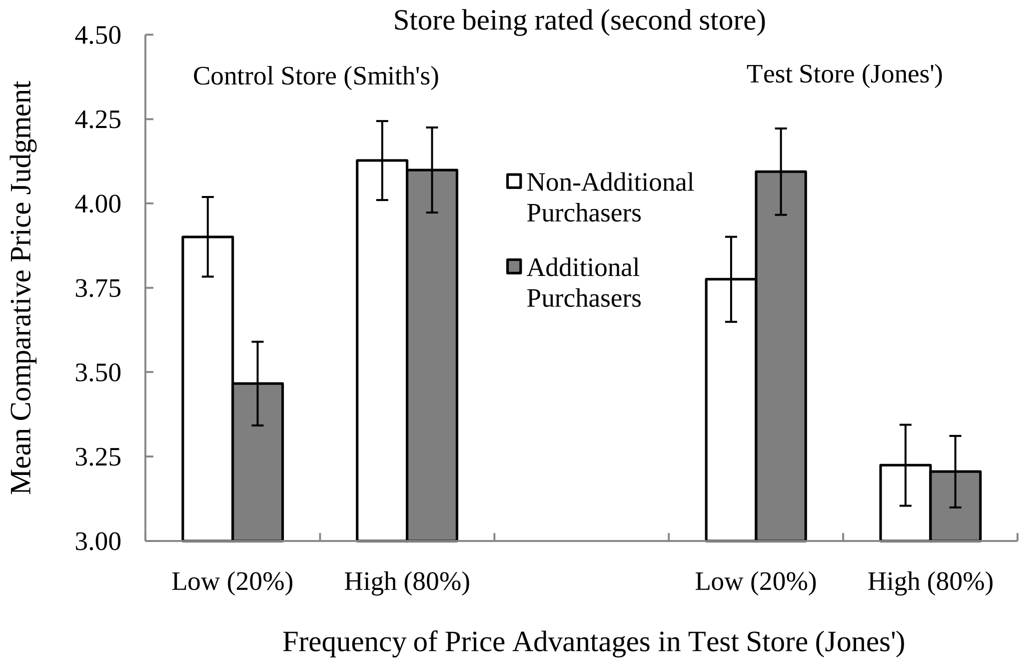

Figure 4.7: Interaction plot of the mean comparative price judgments of the second store relative to the first store across frequency conditions, presentation orders and Additional purchase groups (Experiment 4).

Similarly, participants were divided into two equal size groups based upon the Exclusion measure, with participants who scored less than or equal to 0.029 classified as non-Excluders (n = 313) and participants who scored more than 0.029 classified as Excluders (n = 313). The mean Exclusion score of the non-Excluders (M = -0.10, SEM = 0.008) was lower than the mean score of the Excluders (M = 0.18, SEM = 0.008) and the difference was highly significant ($t(624) = -25.098$, $p<0.001$, two-tailed). Repeated-measures ANOVA, with checkout (pre- or post-checkout) as a within-subjects factor and frequency condition, presentation order and Exclusion group as between-subjects factors, was used to test for differences in the mean price judgment rating of the second store relative to the first store. As already shown previously, there was no significant difference between mean comparative price judgments before and after the checkout, but there was a significant difference in mean ratings between the two stores and the frequency of price advantages in the second store interacted significantly with the presentation order. In addition, there was a significant three-way interaction between the presentation order, frequency condition and Exclusion group ($F(1,618) = 5.946$, $p<0.05$, $\eta^2=0.01$). The interaction plots show that, once again, the mean comparative price judgments are always more favourable when the second store has a high frequency of price advantages than when the second store has a low frequency of price advantages. For both presentation orders the difference in comparative price judgment rating between high and low frequency conditions is more extreme for Excluders than for non-Excluders. The results of the within-subjects tests are shown in Table 4.29, the between-subjects tests are shown in Table 4.30, and the three-way interaction is shown in Figure 4.8.

TABLE 4.29
Within-subjects tests of repeated-measures ANOVA of comparative price judgments including Exclusion group as a between-subjects factor (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Checkout	0.158	1	0.158	0.13	0.72	0.00
Checkout*Freq	3.601	1	3.601	2.91	0.09	0.01
Checkout*Order	0.280	1	0.280	0.23	0.63	0.00
Checkout*Exclusion Group	0.618	1	0.618	0.50	0.48	0.00
Checkout*Freq*Order	0.422	1	0.422	0.34	0.56	0.00
Checkout*Freq*ExGrp	0.304	1	0.304	0.25	0.62	0.00
Checkout*Order*ExGrp	4.293	1	4.293	3.47	0.06	0.01
Checkout*Freq*Order*ExGrp	0.270	1	0.270	0.22	0.64	0.00
Error	764.262	618	1.237			

TABLE 4.30
Between-subjects tests of repeated-measures ANOVA of comparative price judgments including Exclusion group as a between-subjects factor (Experiment 4).

Source	*SS*	*df*	*MS*	*F*	*p*	η2
Frequency	3.631	1	3.631	1.61	0.21	0.00
Order	27.693	1	27.693	12.26	<0.001	0.02
Exclusion Group	5.074	1	5.074	2.25	0.13	0.00
Freq*Order	105.278	1	105.278	46.62	<0.001	0.07
Freq*Exclusion Group	3.303	1	3.303	1.46	0.23	0.00
Order*Exclusion Group	0.016	1	0.016	0.01	0.93	0.00
Freq*Order*ExGrp	13.428	1	13.428	5.95	<0.05	0.01
Error	1395.540	618	2.258			

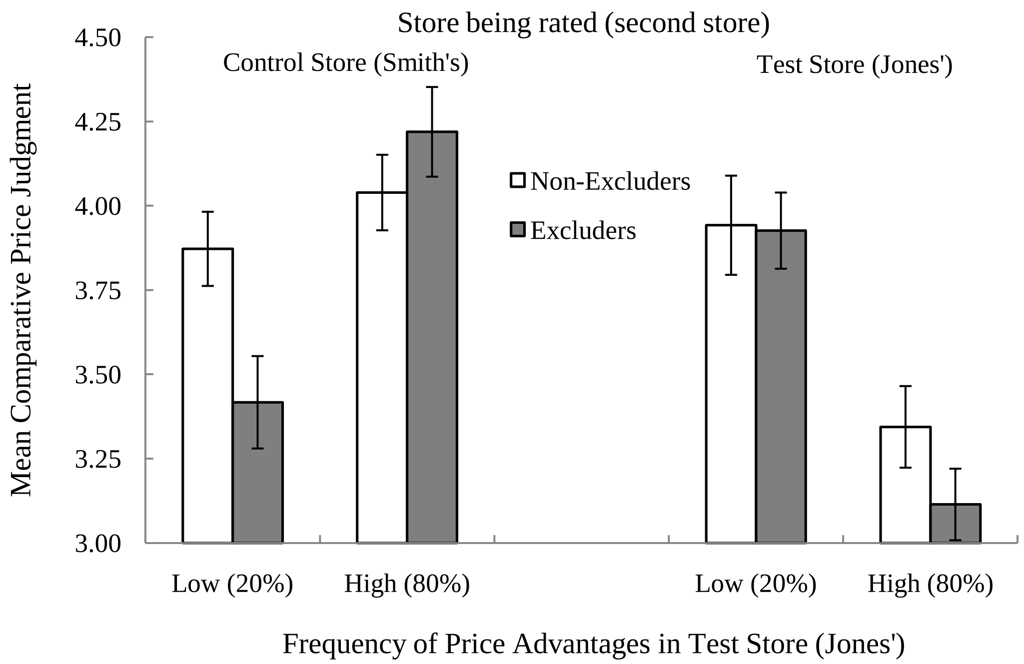

Figure 4.8: Interaction plot of the mean comparative price judgments of the second store relative to the first store across frequency conditions, presentation orders and Exclusion groups (Experiment 4).

Neither of the hypotheses drawn from the work of Bell and Lattin (1998) were supported. Strategic additional purchasing did not vary between high and low frequency stores. Rather, strategic exclusion of previously-purchased items was more prevalent when the second store had a small number of large price disadvantages. Although it might be argued that this is simply the reverse of the strategic purchasing behaviour described by Bell and Lattin, the participants in the high frequency store were both more likely to exclude previously purchased items and also to judge the prices in the second store more favourably relative to the first store. In this case the strategic purchasers were *more* favourable towards the high frequency (EDLP) store than the low frequency (PROMO) store, not *less* favourable as was hypothesized. In fact, the results of Experiment 4 suggest that strategic purchasing behaviour is correlated with more extreme differences in comparative price judgments between low frequency and high frequency stores. The direction of causality is unclear – more extreme comparative price judgments and strategic purchasing behaviour may both result from a greater awareness of price differences; alternatively, a tendency to engage in strategic purchase behaviour may lead to a greater awareness of price differences and hence more extreme comparative price judgments – but in either case there was no evidence that strategic purchasers prefer a different pricing structure to non-strategic purchasers.

4.5 Discussion

Support for the experimental hypotheses was mixed. Hypothesis 1 was strongly supported, with both comparative price judgments and absolute price judgments being more favourable when the second store had a high frequency of (small) price advantages rather than a low frequency of (large) price advantages. However, Hypotheses 2 and 3 were not supported: comparative price judgments varied significantly between the control and test store when the magnitude of price differences was large but not when they were small, but the difference did not systematically vary with the frequency of price advantages in the test store. No significant effects of magnitude were observed in the (much noisier) absolute price ratings. Overall, the results showed a frequency effect in comparative price judgments but did not show – or the experiment lacked sufficient power to find – a moderating effect of the magnitude of price differences. Nonetheless, the findings from Experiments 1 and 2 were successfully replicated within the more ecologically valid experimental paradigm of Experiments 3 and 4. Holding the average item price constant between a control and test store, comparative price judgments are more favourable when the test store has a high frequency of small price advantages and a low frequency of large price disadvantages and less favourable when the test store has a low frequency of large price advantages and a high frequency of small price disadvantages.

As the results of Experiment 3 had suggested, Experiment 4 also showed that price judgments are distinct from basket cost estimates. Basket cost estimates were not systematically biased by the distribution of inter-store price differences, while comparative price judgments about the two stores were. This supports the idea that comparative price judgments are made across all item prices – or as many as can be recalled – not just the specific basket of items that each participant chose. On the other hand, the fact that strategic exclusion of expensive previously-purchased items was more prevalent than strategic addition of cheaper extra items suggests that participants were better able to recall and respond to the prices of items they had purchased in the first store relative to other items. The case for a link between awareness of item price differences and comparative price judgments is further strengthened by the correlation between strategic purchasing behaviour and the strength of observed differences in price judgments between low and high frequency stores. However, participants were unaware of – or could not accurately estimate – the frequency cue. Although item-level inter-store price differences appear to have influenced comparative price judgments, participants were not keeping count of the number of items that were cheaper and more expensive in each store and using that information to form a comparative price judgment.

So how *were* participants making comparative price judgments about the two stores? In this chapter, linear statistics were used to test differences between groups of participants, defined by *a priori* hypotheses and tested as factors in the between-subjects experiment design. However, participants also exhibited individual variability in their behaviour in each store – e.g. time spent in each product department; items purchased; inter-store differences in purchases – which may yield further clues about the judgment process. Additionally, linear statistics do not easily allow one to explore the underlying judgment process: how the available inputs are weighted and combined to form an overall judgment. In the following chapter I shall adopt the alternative approach of fitting different families of judgment models suggested in the previously-reviewed psychological literature to the available data from Experiments 3 and 4 using a maximum-likelihood methodology. Information-theoretic measures of goodness of fit will then be used to decide which model or models provide the best[18] description of the comparative price judgment process.

[18] "Best" using information criteria such as AIC or BIC implies the best trade-off between fitting to the observed data and generalizability of predictions to new data, usually through parsimony in the number of model variables.

CHAPTER 5

FITTING AND COMPARING COGNITIVE PROCESS MODELS OF INTUITIVE COMPARATIVE PRICE JUDGMENTS

5.1 Introduction

The findings from Experiments 3 and 4 suggest that intuitive comparative price judgments are sensitive to differences in mean price between two paired-item distributions, but are biased by the frequency of paired-item advantages and disadvantages. When the mean price is identical in two stores, the store with a high number of small price advantages is perceived to be cheaper than the store with a low number of large price advantages. In this chapter a number of plausible cognitive process models are compared in order to understand how these intuitive statistical judgments are made. In particular, I explore which process models predict the observed frequency effect and what the underlying cognitive mechanism is in each case. Two families of process model are compared: *representative participant* models that are based only on the set of prices faced by each participant and *sampling differences* models that incorporate additional information about the process each individual participant followed in order to make their judgment. Each process model incorporates a stochastic response function, in which the judged price difference determines the probability of a participant choosing each of the seven available responses. This enables model parameters to be fitted in a maximum-likelihood procedure and allows for comparison between models using model selection statistics such as the Akaike Information Criterion (AIC). Because of the importance of the response function, I begin the chapter with a Signal Detection analysis of the data from Experiment 3, in order to determine how participants used the rating scale and whether the pattern of observed responses is consistent with an SDT model of the judgment process.

5.2 Signal Detection Analysis

5.2.1 Estimating Hit Rates and False Alarm Rates

As outlined in Chapter 1, Signal Detection Theory (SDT) assumes that discrimination judgments are equivalent to statistical inference, with a subjective decision criteria required to decide whether or not two observed signals will be judged as the same or different. Varying the decision criteria implies varying the trade-off between correct identification of differences (hits) and incorrect identification of a difference when the two signals are the same (false alarms). In order to determine the *Receiver Operating Characteristic* (ROC) of a participant, it is necessary to obtain judgments at different levels of decision criteria, i.e. different levels of confidence. The seven-point rating scale used for the comparative price judgments in Experiments 3 and 4 can be interpreted as a confidence rating: a rating of 1 implies the participant was extremely confident that the second store was cheaper whilst a rating of 3 implies that the participant was only weakly confident that the second store was cheaper. By aggregating data across all the responses to Experiment 3, the ratings can be transformed into binary judgments of whether the second store was cheaper or the same price as the first store with differing degrees of confidence.

For high confidence judgments, participants who responded 1 ("Compared to the first supermarket, I thought that the second store was a lot cheaper") were coded as having identified a difference and all other responses (2-7) were coded as perceiving the two stimuli as identical. For the next level of confidence, all participants who responded 1 or 2 were coded as having identified a difference and all other responses (3-7) were coded as perceiving the two stimuli as identical. In this way, each rating was transformed into six binary responses. Where participants saw the stores in reverse order, the ratings were also reversed (i.e. 1 mapped to 7, 2 mapped to 6, etc.) in order to ensure consistency across judgments. Each participant contributed two ratings, before and after the checkout in the second store. Hit rates were calculated as the proportion of participants correctly identifying that the test store was cheaper, at each discount level and each degree of confidence. False alarm rates were calculated as the proportion of participants incorrectly identifying that the second store was cheaper in the 0% discount condition, for each degree of confidence. The observed pattern of hits and false alarms is shown in Table 5.1 and the ROC chart is shown in Figure 5.1.

TABLE 5.1
Observed hit rates and false alarm rates for comparative price judgments with varying degrees of confidence in Experiment 3.

	Degree of Confidence					
Discount	Highest (1)	(1-2)	(1-3)	(1-4)	(1-5)	Lowest (1-6)
0% (False alarms)	0%	8%	21%	65%	87%	98%
1%	3%	11%	37%	66%	86%	97%
2%	13%	23%	56%	85%	98%	100%
3%	1%	16%	57%	74%	90%	97%
4%	3%	19%	55%	77%	97%	98%
5%	0%	12%	37%	75%	88%	96%
7.5%	6%	21%	56%	81%	85%	94%
10%	18%	44%	70%	84%	94%	98%
15%	29%	60%	79%	85%	92%	98%
20%	18%	56%	82%	88%	90%	100%
30%	47%	72%	91%	98%	98%	100%

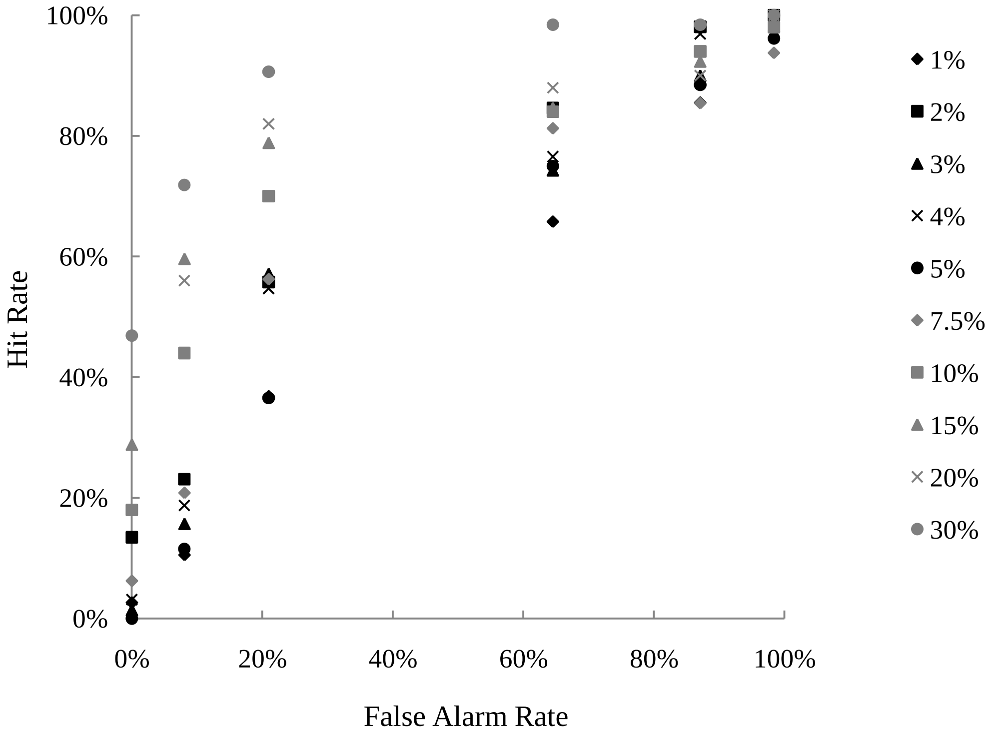

Figure 5.1: Receiver Operating Characteristic plot for judgments discriminating whether the second store is cheaper than the first store at different discount levels (Experiment 3).

5.2.2 Estimating Sensitivity

The ROC plot shows that the observed judgments are more accurate than random guesses, as the majority of points lie in the top left of the diagram. It also shows that for a given false alarm rate, the hit rate tends to increase with the discount in the test store, implying increasing sensitivity. SDT usually assumes that activity caused by each stimulus is normally distributed on a subjective psychological scale, and the sensitivity (usually denoted as *d*') is measured by the distance between the mean of the two distributions divided by their standard deviation, as described in Chapter 1 (see Figures 1.3 and 1.4). Hence, the sensitivity *d*' is equal to *Z*(hit rate) – *Z*(false alarm rate). If the false alarm rates and hit rates on the ROC plot are transformed using a standardized cumulative normal probability distribution, the resultant points should lie on a straight line. Furthermore, the intersection of the line with the y-axis (normalized hit rate) is the sensitivity, *d*'. In order to transform extreme values of hit rate and false alarm rate, zeroes were replaced with 1/N, where N is the number of ratings for that test stimulus, and 100 was replaced with N-1/N. Because so few participants used the highest response categories of 6 and 7, the final two columns of Table 5.1 were excluded. An ordinary least squares procedure was used to fit a straight line through each set of points using Microsoft Excel. The resulting normalized plot is shown in Figure 5.2 and the estimated values of *d*' for each discount level are shown in Figure 5.3.

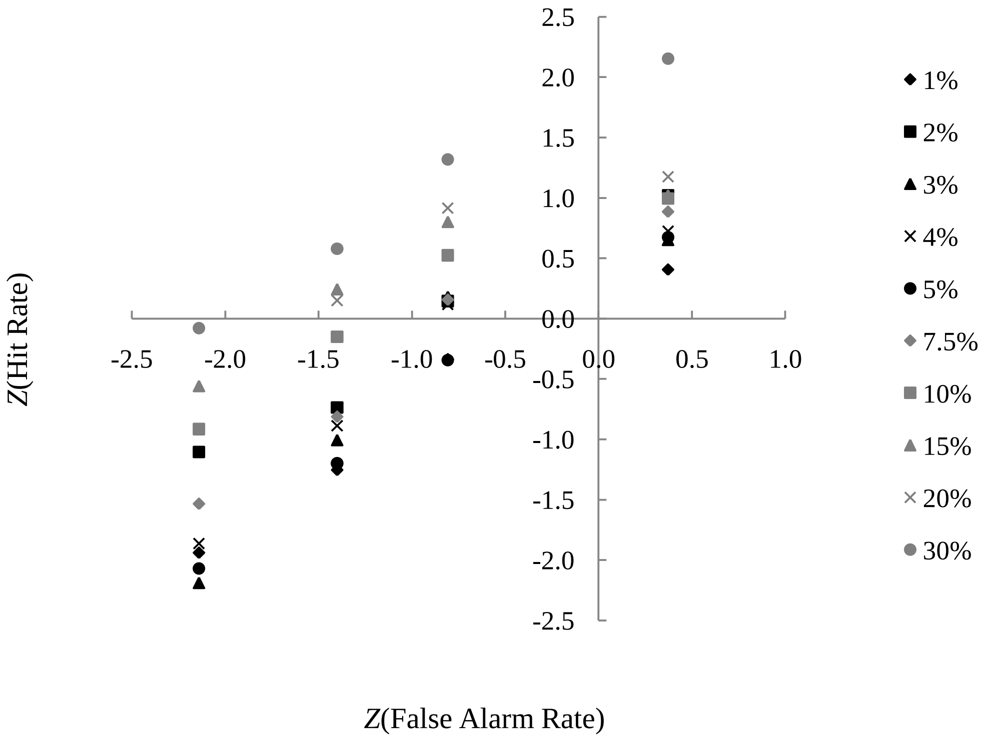

Figure 5.2: Normalized ROC plot for judgments discriminating whether the second store is cheaper than the first store at different discount levels (Experiment 3).

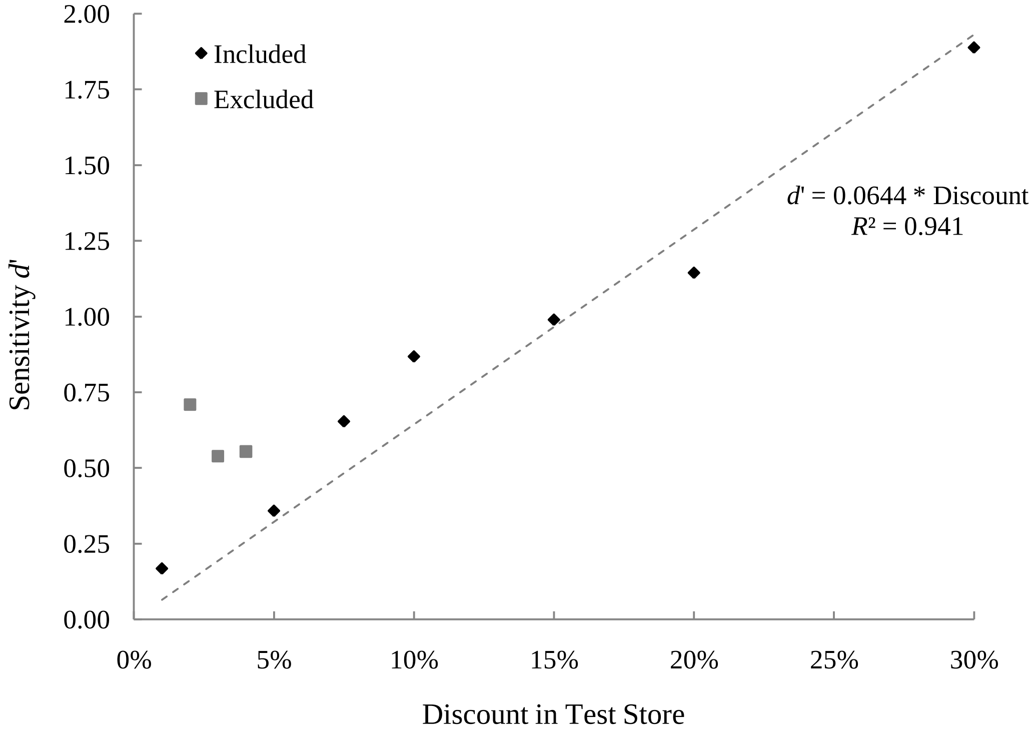

Figure 5.3: Relationship between sensitivity *d*' and the percentage difference in mean price showing the values excluded and included in the subsequent linear fit (Experiment 3).

As can be seen from Figure 5.3, participants' sensitivity increased approximately linearly as the difference in mean price between the two stores was increased. An OLS procedure was used to fit a straight line through the observed values of *d*' and the origin using Microsoft Excel, excluding three points as indicated due to noise at low discount levels. The linear fit indicated that *d*' increases by 0.0644 for every 1% increase in the mean price difference. The implied ROC plot for each level of mean price difference is shown in Figure 5.4.

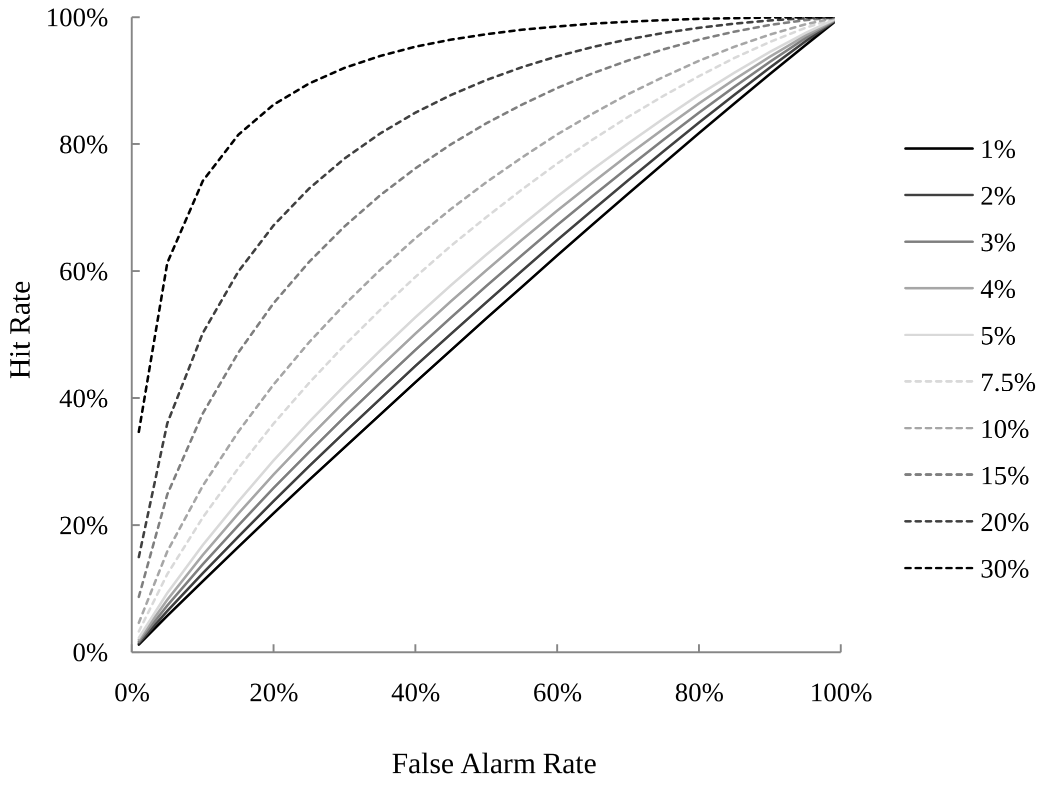

Figure 5.4: ROC plot for judgments discriminating whether the second store is cheaper than the first store at different discount levels using fitted values of *d*' (Experiment 3).

5.2.3 Estimating Decision Criteria

The decision criterion for each discrimination judgment determines the false alarm rate, equivalent to setting the value of α for Type I errors in ANOVA. The sensitivity *d*' then determines the hit rate, equivalent to the power of an ANOVA analysis. Each confidence level, i.e. each threshold between neighbouring values on the seven-point response scale, corresponds to a decision criterion with an associated false alarm rate. Given the fitted values of *d*', the expected hit rate for any false alarm rate can be calculated. The expected hit rates for each discount condition were compared with the observed hit rates and an OLS procedure was used to find the decision criterion, *Z*(false alarm rate), which best predicts the observed data for each confidence level. The estimated value of each decision criterion, *Z*(false alarm rate), is shown in Figure 5.5.

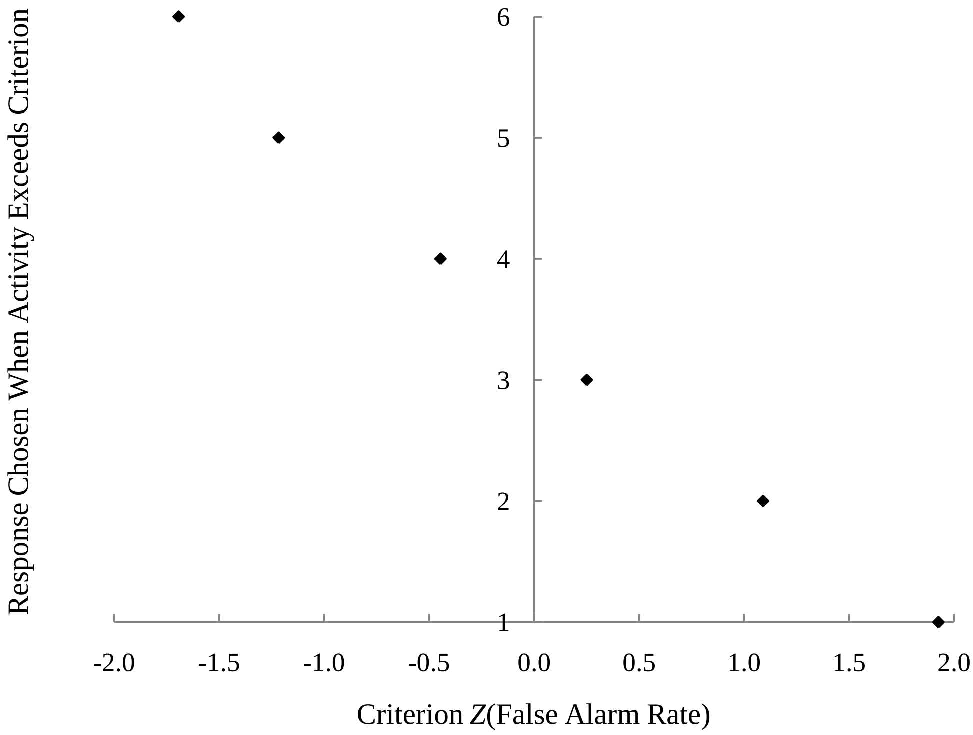

Figure 5.5: Fitted values of decision criteria for the threshold between each point on the seven-point response scale (Experiment 3).

The decision criteria are approximately linearly spaced, suggesting that participants used the response scale sensibly to differentiate between different levels of confidence in discriminating a difference in mean price between the two stores. The most extreme response value of 1 (“Compared to the first supermarket, I thought that the second store was a lot cheaper”) has a false alarm rate of just 2.7%. Selecting one of the first three response values, i.e. responding that the second store is cheaper than the first store, has a false alarm rate of 40%. In order for the hit rate for the first three response values to be at least 50%, the mean price difference between the two stores has to be between 3% and 4%, which agrees with the 3% estimate for the just-noticeable difference made in Chapter 3 using binary logistic regression. The estimated false alarm and hit rates for each level of mean price difference are shown in Table 5.2.

TABLE 5.2
Fitted hit rates and false alarm rates for comparative price judgments with varying degrees of confidence in Experiment 3.

	Degree of Confidence					
Discount	Highest (1)	(1-2)	(1-3)	(1-4)	(1-5)	Lowest (1-6)
0% (False alarms)	3%	14%	40%	67%	89%	96%
1%	3%	15%	43%	69%	90%	96%
2%	4%	17%	45%	72%	91%	97%
3%	4%	18%	48%	74%	92%	97%
4%	5%	20%	50%	76%	93%	97%
5%	5%	22%	53%	78%	94%	98%
7.5%	7%	27%	59%	82%	96%	99%
10%	10%	33%	65%	86%	97%	99%
15%	17%	45%	76%	92%	99%	100%
20%	26%	58%	85%	96%	99%	100%
30%	50%	80%	95%	99%	100%	100%

5.2.4 Implications and Limitations

The fitted values of false alarm and hit rate emphasize how poorly participants performed at making the paired-item mean discrimination judgment in Experiment 3. Only half of participants would be confident enough to choose the most extreme response value when the mean price difference between the two stores was 30%. Although 95% of participants should correctly identify that the second store is cheaper in that condition, they would also have a 40% false alarm rate when there is no price difference between the two stores. Overall, the Signal Detection analysis shows that participants are insensitive to small differences in mean price and lack the confidence to use extreme responses on the scale. However, it also shows that participants used the rating scale sensibly, with the response values being approximately linearly spaced along the subjective psychological scale.

The SDT approach has a number of limitations for this data and judgment task. Firstly, using the seven-point response scale values as estimates of confidence is a crude methodology for eliciting judgments with different decision criteria. In particular, the interpretation of false alarms is counter-intuitive as correct responses of 4 (the two stores have the same prices) are treated as false alarms for low confidence judgments. For a more accurate SDT analysis, a binary response should be elicited (second store is/is not cheaper than first store) and the decision criterion manipulated by encouraging respondents only to indicate a difference when their confidence exceeds a certain level. Secondly, the sparseness of the data and the need to fit sensitivity and decision criteria values generated some large discrepancies between observed and fitted performance, especially false alarm rates. This indicates a poor model fit, which means either the data contains a lot of noise or participants' judgments were only approximated by a signal detection process. Thirdly, and most fundamentally, the standard SDT model cannot easily explain the frequency bias observed in Experiment 4.

In the standard SDT model each item (store) is mapped onto a psychological response subject to equal Gaussian noise. If the mean item price in each store is accurately estimated (and there is nothing inherent in SDT to suggest the signal would be sensitive to skew or other higher moments of the price distribution), then the two responses would completely overlap on the psychological response scale and hence would not be discriminable. The frequency bias indicates either that participants' estimate of the mean price in each store is biased by the shape of the price distribution or that participants' judgment process is something other than an SDT-like discrimination between two perceived mean prices. If the former is the case, then an additional process is required to explain why mean price estimates are biased, for example incomplete sampling of the available information or recall errors when estimating mean prices from memory. If the latter is the case, then an alternative process model is required, such as the "count the frequency of price advantages" heuristic suggested by Alba et al (1994). In that particular case, SDT cannot explain the salience of frequency information, as it suggests that large inter-store item price differences should be more discriminable and hence more salient.

Thus, two different families of process model could explain the observed frequency effect. The first family of models involve a two-stage decision in which the mean item price in each store is estimated, and the estimates then compared in a signal detection discrimination judgment to determine the response value. The common feature of these models is that they assume the mean price for each store is estimated independently and hence I shall refer to them as *pairing-independent* models. The second family of models assumes that the decision process involves some comparison of individual item prices between stores, such as counting the frequency of price advantages, with an alternative process to determine the response value. I shall refer to these as *pairing-dependent* models. These two types of model can make quite divergent predictions, especially in the context of real stores with different ranges. For example, imagine one store (A) that is cheaper on items sold in both stores but that also stocks additional items that gave it a higher average item price than another store (B). In this situation a pairing-dependent model would predict Store A would be judged as cheaper, while a pairing-independent model would predict Store B would be judged as cheaper.[19]

Both types of model can be either *representative participant* models (based

[19] This is a form of "Simpson's Paradox", described by Edward H. Simpson (1951), in which statistical conclusions which hold for separate groups are reversed when those groups are combined.

on the total set of prices faced by each participant, regardless of whether or not every price was observed) or *sampling differences* models (incorporating information about the process each individual participant followed in order to make their judgment). In the following sections I shall fit and compare cognitive process models of each possible type to the comparative judgment data observed in Experiments 3 and 4.

5.3 Cognitive Model Fitting

5.3.1 Modelling Approach

5.3.1.1 Ordered Probit Model

In the experimental setting participants are presented with stimuli (the prices in each store) and then asked to respond to a judgment task using a rating scale. Any model of the cognitive processes involved must therefore have two stages. Firstly, the comparative judgment process must take the stimuli as inputs and transform them into an internal psychological value (the *Valuation Function*). Secondly, that internal value must be transformed into a response on the provided scale (the *Response Function*). In order to compare and choose between alternative cognitive models I chose a common response function, allowing me to focus upon identifying the valuation function that best predicts the observed data. I assume that the valuation functions all have the general form:

$$R_{AB}^{*} = f(\text{prices}_{\text{A}}, \text{prices}_{\text{B}}, \text{behaviour}_{\text{A}}, \text{behaviour}_{\text{B}}) + \varepsilon_{AB}$$

Where ε is a zero-mean Gaussian noise term to capture unexplained variance, $\varepsilon_{AB} \sim N(0, \sigma^2)$. I also assume that the internal psychological value, R_{AB}^{*}, is transformed into the observed response, R_{AB}, in an ordered and monotonically-increasing fashion:

$$R_{AB} \begin{cases} 1 & if \quad R_{AB}^* \leq c_1 \\ 2 & if \quad c_1 < R_{AB}^* \leq c_2 \\ 3 & if \quad c_2 < R_{AB}^* \leq c_3 \\ 4 & if \quad c_3 < R_{AB}^* \leq c_4 \\ 5 & if \quad c_4 < R_{AB}^* \leq c_5 \\ 6 & if \quad c_5 < R_{AB}^* \leq c_6 \\ 7 & if \quad R_{AB}^* > c_6 \end{cases}$$

The thresholds $c_{1\text{-}6}$ are parameters to be jointly identified with the parameters of the valuation function. Specifically, I chose an ordered probit model, a commonly-used response function for stochastic choice models:

$$p(R_{AB} = 1) = \Phi[(c_1 - f(\cdot))/\sigma]$$

$$p(R_{AB} = 2) = \Phi[(c_2 - f(\cdot))/\sigma] - \Phi[(c_1 - f(\cdot))/\sigma]$$

$$p(R_{AB} = 3) = \Phi[(c_3 - f(\cdot))/\sigma] - \Phi[(c_2 - f(\cdot))/\sigma]$$

$$p(R_{AB} = 4) = \Phi[(c_4 - f(\cdot))/\sigma] - \Phi[(c_3 - f(\cdot))/\sigma]$$

$$p(R_{AB} = 5) = \Phi[(c_5 - f(\cdot))/\sigma] - \Phi[(c_4 - f(\cdot))/\sigma]$$

$$p(R_{AB} = 6) = \Phi[(c_6 - f(\cdot))/\sigma] - \Phi[(c_5 - f(\cdot))/\sigma]$$

$$p(R_{AB} = 7) = 1 - \Phi[(c_6 - f(\cdot))/\sigma]$$

Where $\Phi[x]$ is the standardized normal cumulative distribution function. The scale of the model was fixed using the normalization $\sigma^2=1$. Parameter values for the valuation and response functions were found using a maximum-likelihood estimation procedure, with the log likelihood function:

$$\ln(L) = \sum\nolimits_{participants} \sum_{i=1}^{i=7} \delta_{AB,i} \ \ln[\, p(R_{AB} = i)]$$

Where $\delta_{m,n} = 1$ if $m = n$ and $\delta_{m,n} = 0$ otherwise. Model parameters were fit in Matlab using the constrained non-linear multivariate optimization routine *fmincon*. The probit boundaries were constrained such that $c_1 \leq c_2 \leq c_3 \leq c_4 \leq c_5 \leq c_6$.

5.3.1.2 Model Selection Criterion

In this chapter I shall be comparing models with different numbers of parameters. In general, models with more parameters are more flexible and able to provide a better fit to any arbitrary set of data, without necessarily having any greater predictive power to generalize to new data. It is therefore important to correct for this effect when comparing models. In order to compare between different valuation functions an information-theoretic model selection criterion, the Akaike Information Criterion (AIC), was calculated for each model:

$$\text{AIC} = 2k - 2\ln(L)$$

This model selection criterion penalizes models with additional free parameters, *k*, in order to prevent over-fitting to the observed data. A lower score implies a model which better predicts the observed data whilst retaining generalizability to new data. The AIC penalizes additional free parameters to a lesser extent than alternative model selection criteria such as BIC. In cases where the number of free parameters in competing models is identical, the AIC reduces to a likelihood ranking. Information-theoretic model selection criteria like the AIC have two major advantages over alternative methodologies. Firstly, they can be used to compare non-nested parametric models, unlike alternatives such as a likelihood ratio test.

Secondly, they allow models to be fit to an entire dataset without requiring a holdout sample to test the generalizability of each fitted model to new data.

Each model was fit twice: once to the pre-checkout comparative price judgments and once to the post-checkout comparative price judgments. A baseline model (Model 0) was fit to the dataset, to determine the degree to which the response function alone could predict the observed data, using a constant valuation function:

$$f(\cdot) = 0$$

The goodness-of-fit measures for the baseline models are shown in Table 5.3 and the fitted parameter values are shown in Table 5.4. The AIC scores obtained for the pre- and post-checkout rating models are 3132.9 and 3440.2 respectively. In order to provide a better explanation of the observed data, any other valuation function must have a lower AIC score.

TABLE 5.3
Goodness-of-fit measures for a baseline model (Model 0) with a constant valuation function.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
0: Baseline	-1560.5	3132.9	-1714.1	3440.2

TABLE 5.4
Fitted parameter values for Model 0.

Parameter	Pre-Checkout	Post-Checkout
c_1	-1.713	-1.250
c_2	-0.992	-0.595
c_3	-0.114	-0.027
c_4	0.554	0.394
c_5	1.424	1.183
c_6	2.389	2.010

5.3.2 Pairing-Independent Judgments

5.3.2.1 Representative Participant Models

As described earlier, one plausible set of cognitive process models assumes that participants estimated the mean item price in each store separately and then compared the estimated mean prices in order to determine the cheaper store. Five basic types of valuation function were compared to determine which best describes how participants might have made that comparison. The first type (A) is based upon the linear difference between the two estimated means:

$$f(\cdot) = \alpha(\mu_B - \mu_A)$$

The second type (B) is based upon the difference in estimated mean prices raised to a power, as suggested by Stevens' Law and described in Chapter 1 (Stevens, 1957):

$$f(\cdot) = \alpha\left({\mu_B}^\beta - {\mu_A}^\beta\right)$$

The third type (C) is based upon the natural logarithm of the ratio of the estimated mean prices, as suggested by Fechner's Law and described in Chapter 1 (Gigerenzer & Murray, 1987):

$$f(\cdot) = \alpha \ln\left(\frac{\mu_B + \beta}{\mu_A + \beta}\right)$$

The fourth type (D) is based upon the Luce choice rule, motivated by the discussion in Chapter 1 of the tendency for human judges to probability match in repeated decisions (Shanks et al., 2002):

$$f(\cdot) = \alpha\left(\frac{{\mu_B}^\beta}{{\mu_A}^\beta + {\mu_B}^\beta}\right)$$

The fifth type (E) is based upon the logistic function, with a sigmoid relationship between the valuation and the difference in estimated mean prices in each store, as suggested by Buyukkurt and described in Chapter 1 (Buyukkurt, 1986):

$$f(\cdot) = \alpha\left(\frac{e^{\beta\mu_B}}{e^{\beta\mu_A} + e^{\beta\mu_B}}\right)$$

In all cases, the shape parameter was constrained so that $\beta \geq 0$. Using one or more of these five types of valuation function it is possible to predict judgments that exhibit increasing, decreasing, or constant sensitivity to differences in estimated mean price.

Different families of pairing-independent model were fitted. Some of those models assume that all participants make the comparative judgment in the same manner (*representative participant*), while others assume that the process by which each individual sampled the price information presented influences the manner in which they make their subsequent judgment (*sampling differences*). The simplest form of representative participant model is one in which participants are able to accurately determine the mean price in each store (Model 1), with the price P of each item i being equally weighted in the estimated mean price in store S:

$$\mu_S = \frac{1}{150}\sum_{i=1}^{i=150} P_{S,i}$$

Clearly this family of models cannot predict the observed frequency effect from Experiment 4, as the mean item price was identical in the two stores. Nonetheless, it serves as a useful yardstick against which to compare more complex cognitive process models.

As before, each type of Model 1 was fit twice, to both pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.5 and the fitted parameter values are shown in Table 5.6. All five model types offer an improvement over the baseline model. The best-fitting models for pre-checkout ratings were the power law comparison (1B), the Luce rule comparison (1D), and the logistic comparison (1E). The log-likelihood of all five model types was almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. A log-likelihood score of -1481.7 implies that the mean likelihood of the observed ratings is 21%, compared to a random choice probability of 1 in 7 or 14%. For the pre-checkout rating models, all five models predict an increased probability of a high rating ("Compared to the first supermarket, I thought that the second store was a lot more expensive") as the mean price in the first store decreases or as the mean price in the second store increases ($\alpha > 0$).

TABLE 5.5
Goodness-of-fit measures for Model 1.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
1A: Linear	-1484.6	2983.1	-1643.5	3301.0
1B: Stevens	-1481.7	2979.5	-1643.5	3302.9
1C: Fechner	-1484.6	2985.1	-1643.5	3302.9
1D: Luce	-1481.7	2979.4	-1643.5	3303.0
1E: Logistic	-1481.7	2979.4	-1643.5	3303.0

TABLE 5.6
Fitted parameter values for Model 1.

	Pre-Checkout				
Parameter	1A	1B	1C	1D	1E
c_1	-1.904	-1.907	-1.904	-0.321	-0.246
c_2	-1.082	-1.085	-1.082	0.502	0.577
c_3	-0.141	-0.141	-0.141	1.447	1.521
c_4	0.570	0.575	0.570	2.162	2.237
c_5	1.505	1.514	1.505	3.102	3.176
c_6	2.638	2.643	2.638	4.231	4.306
α	3.109	0.045	85 633	3.174	3.324
β		5.557	27 540	10.76	5.487
	Post-Checkout				
Parameter	1A	1B	1C	1D	1E
c_1	-1.372	-1.372	-1.372	2.342	84.85
c_2	-0.651	-0.651	-0.651	3.063	85.57
c_3	-0.043	-0.043	-0.043	3.672	86.18
c_4	0.406	0.406	0.406	4.121	86.63
c_5	1.252	1.252	1.252	4.966	87.48
c_6	2.138	2.138	2.138	5.852	88.36
α	3.051	4.839	19.20	7.429	172.5
β		0.729	4.574	2.971	0.071

The second mean estimation process tested (Model 2) assumes that rather than making an accurate estimate of the mean price in each store, participants' estimates were biased, with more weight being placed on either large or small prices:

$$\mu_S = \left(\frac{1}{150}\sum_{i=1}^{i=150} {P_{S,i}}^{w}\right)^{1/w}$$

If $w < 1$, then proportionately more weight is placed on small prices. This would bias the mean estimate in the frequency store downwards, as it contains many small prices and a few large prices. Hence, if Model 2 predicts a frequency effect, a weighting factor less than unity should be observed.

The same five types of mean price comparison (A-E) were once again compared, and each model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.7 and the fitted parameter values are shown in Table 5.8. All five model types offer an improvement over both the baseline model and the equivalent versions of Model 1. The best-fitting model for pre-checkout ratings was the logistic comparison (2E). The log-likelihoods of all five model types were again almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. For the pre-checkout rating models, all five models predict an increased probability of a high rating as the mean price in the first store decreases or as the mean price in the second store increases ($\alpha > 0$) and three of the five models predict a lower estimated mean price in a frequency store compared to a magnitude store ($w < 1$).

TABLE 5.7
Goodness-of-fit measures for Model 2.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
2A: Linear	-1470.7	2957.4	-1630.3	3276.6
2B: Stevens	-1466.8	2951.6	-1630.3	3278.5
2C: Fechner	-1470.7	2959.4	-1630.3	3278.6
2D: Luce	-1466.1	2950.1	-1630.2	3278.3
2E: Logistic	-1466.2	2950.3	-1630.2	3278.5

TABLE 5.8
Fitted parameter values for Model 2.

	Pre-Checkout				
Parameter	2A	2B	2C	2D	2E
c_1	-1.939	-1.945	-1.939	-0.390	-0.336
c_2	-1.106	-1.111	-1.106	0.444	0.498
c_3	-0.147	-0.147	-0.147	1.409	1.464
c_4	0.575	0.581	0.575	2.137	2.192
c_5	1.525	1.538	1.525	3.096	3.150
c_6	2.670	2.675	2.670	4.233	4.287
α	5.763	0.276	193 620	3.112	3.221
β		0.359	33 597	12.12	0.260
w	-0.070	6.036	-0.070	0.228	9.385
	Post-Checkout				
Parameter	2A	2B	2C	2D	2E
c_1	-1.399	-1.399	-1.399	1.267	2.925
c_2	-0.667	-0.667	-0.667	1.998	3.656
c_3	-0.047	-0.047	-0.047	2.619	4.277
c_4	0.410	0.410	0.410	3.076	4.734
c_5	1.266	1.267	1.266	3.932	5.590
c_6	2.161	2.161	2.162	4.827	6.485
α	5.885	4.824	109 320	5.330	8.646
β		-0.137	18 537	4.598	-0.127
w	-0.169	1.218	-0.173	-0.194	2.777

There is no strong theoretical justification for participants' estimated mean prices to be biased toward small prices rather than large prices (although knowledge of small and large prices could be tested empirically). In order to estimate the mean price in each store, participants would have to recall a sample of those prices from memory, and this is a more plausible source of bias in the estimates. As discussed in Chapter 1, memory trace theories suggest that the most easily recallable memories are those that are both salient and typical (Hintzmann, 1988). Saliency is related to factors such as attention and repetition at the moment of storage. Typicality is related to the similarity of a memory trace to the other traces in memory storage.

The only available attribute for describing each item in the experimental store is the price, hence the distance between any two items *i* and *j* is given by:

$$d_{ij} = \left|P_i - P_j\right|$$

The similarity between any pair of prices is inversely related to the distance, and depends upon the similarity gradient ψ:

$$S_{ij} = e^{-\psi d_{ij}}$$

The similarity is constrained so that $\psi \geq 0$. If ψ is large then only very close items are judged as similar. If ψ is small, then even distant items are judged as somewhat similar. When $\psi = 0$ then all items are judged as equally similar. The typicality of an item is given by the summed similarity with all items (including itself):

$$T_i = \sum_{j=1}^{j=150} S_{ij}$$

Model 3 assumes that the probability of recall for a price is proportional to its typicality, so the estimated mean price in each store is given by:

$$\mu_S = \sum_{i=1}^{i=150} \left(\frac{T_{S,i}}{\sum_{i=1}^{i=150} T_{S,i}} P_{S,i} \right)$$

This weighting could explain the frequency effect, as small prices are more typical in a frequency store, so the estimated mean price would be biased downwards. When ψ=1, the estimated mean price in the baseline store is £1.03 (actual mean = £1.95). The estimated mean prices in the magnitude stores are £1.04 (F20 M5) and £1.06 (F20 M20). The estimated mean prices in the frequency stores are £1.03 (F80 M5) and £0.98 (F80 M20). Hence, if Model 3 predicts a frequency effect, a positive similarity gradient should be observed.

Again, five types of mean price comparison (A-E) were compared, and each model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.9 and the fitted parameter values are shown in Table 5.10. All five model types offer an improvement over both the baseline model and the equivalent versions of Models 1 and 2. The best-fitting models for pre-checkout ratings were the power law comparison (3B), the Luce rule comparison (3D) and the logistic comparison (3E). The log-likelihoods of all five model types were again almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. For the pre-checkout rating models, all five models predict an increased probability of a high rating as the mean price in the first store decreases or as the mean price in the second store increases ($\alpha > 0$) and a lower estimated mean price in a frequency store compared to a magnitude store ($\psi > 0$).

TABLE 5.9
Goodness-of-fit measures for Model 3.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
3A: Linear	-1456.0	2928.0	-1625.3	3266.6
3B: Stevens	-1451.1	2920.2	-1625.3	3268.5
3C: Fechner	-1456.0	2930.0	-1625.3	3268.5
3D: Luce	-1451.3	2920.6	-1625.2	3268.5
3E: Logistic	-1451.2	2920.5	-1625.3	3268.6

TABLE 5.10
Fitted parameter values for Model 3.

	Pre-Checkout				
Parameter	3A	3B	3C	3D	3E
c_1	-1.970	-1.976	-1.969	-0.356	-0.251
c_2	-1.130	-1.141	-1.130	0.480	0.585
c_3	-0.156	-0.159	-0.156	1.462	1.567
c_4	0.581	0.588	0.581	2.207	2.313
c_5	1.550	1.566	1.550	3.186	3.291
c_6	2.709	2.711	2.709	4.332	4.437
α	4.819	0.928	168 870	3.239	3.450
β		4.457	35 213	8.927	7.424
ψ	42.61	45.29	43.02	46.01	45.74
	Post-Checkout				
Parameter	3A	3B	3C	3D	3E
c_1	-1.409	-1.409	-1.409	1.771	109.7
c_2	-0.675	-0.675	-0.675	2.505	110.4
c_3	-0.051	-0.050	-0.050	3.130	111.0
c_4	0.411	0.411	0.411	3.591	111.5
c_5	1.273	1.273	1.273	4.453	112.3
c_6	2.170	2.170	2.170	5.350	113.2
α	5.090	6.746	17.29	6.360	222.1
β		0.740	2.466	3.141	0.092
ψ	24.93	24.55	24.53	24.26	24.93

5.3.2.2 Sampling Differences Models

Whilst the previous (representative participant) models assume the same comparative judgment process is followed by all participants, it is plausible that each individual followed a slightly different judgment process because their browsing behaviour led them to sample the price information in a unique way. Three measures of each participant's behaviour can be used to characterise their personal sampling process: the basket of items purchased in each store; the time spent browsing each department in each store; and the total cost of the basket of items purchased in each store, which was shown at the checkout.

As described in Chapter 1, memory traces are stronger when greater attention is paid to a stimulus at the moment of encoding and storage. It is likely that greater attention was paid to purchased items than to non-purchased items, as the purchased items were selected and placed into the basket by the participant. Hence, the memory traces for the prices of purchased items should be stronger than those for the prices of un-purchased items. If participants attempt to estimate a mean price for each store by recalling item prices, the estimated mean prices should then be biased towards purchased item prices due to the higher probability of recall. A binary variable $X_{S,i}$ was created for each item i in each store S, such that $X_{S,i} = 1$ when the item was purchased and $X_{S,i} = 0$ otherwise. A weighting factor for each price was then created, so that:

$$W_{S,i} = 1 + \pi X_{S,i}$$

$$\mu_S = \sum_{i=1}^{i=150} \left(\frac{W_{S,i}}{\sum_{i=1}^{i=150} W_{S,i}} P_{S,i} \right)$$

If the prices of purchased items are more salient than the prices of non-purchased items, and hence bias the estimated mean price in each store, then a positive contribution $\pi > 0$ should be observed.

Purchase-weighting (PW) was applied to each of the five types of mean price comparison (A-E), and each model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.11 and the fitted parameter values are shown in Table 5.12. Purchase-weighting all five model types improves the model fits relative to the equivalent un-weighted versions of Model 1. The best-fitting models for pre-checkout ratings were the Luce rule comparison (1D+PW) and the logistic comparison (1E+PW). The log-likelihoods of all five model types were again almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. For the pre-checkout rating models, all five models predict an increased probability of a high rating as the mean price in the first store decreases or as the mean price in the second store increases ($\alpha > 0$) and mean price estimates biased toward the prices of purchased items ($\pi > 0$).

TABLE 5.11
Goodness-of-fit measures for purchase-weighted versions of Model 1.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
1A+PW: Linear	-1483.4	2982.8	-1608.8	3233.7
1B+PW: Stevens	-1480.7	2979.3	-1608.4	3234.8
1C+PW: Fechner	-1483.4	2984.8	-1608.5	3234.9
1D+PW: Luce	-1478.8	2975.5	-1607.6	3233.1
1E+PW: Logistic	-1479.0	2976.0	-1608.3	3234.7

TABLE 5.12
Fitted parameter values for purchase-weighted versions of Model 1.

	Pre-Checkout				
Parameter	1A+PW	1B+PW	1C+PW	1D+PW	1E+PW
c_1	-1.907	-1.911	-1.906	-0.391	-0.348
c_2	-1.083	-1.086	-1.082	0.436	0.478
c_3	-0.139	-0.139	-0.139	1.387	1.428
c_4	0.573	0.577	0.573	2.104	2.145
c_5	1.508	1.517	1.507	3.046	3.087
c_6	2.636	2.642	2.634	4.170	4.210
α	3.163	0.052	395 640	3.047	3.130
β		5.428	126 030	11.90	6.334
π	0.312	0.168	0.312	0.393	0.351
	Post-Checkout				
Parameter	1A+PW	1B+PW	1C+PW	1D+PW	1E+PW
c_1	-1.416	-1.418	-1.418	1.012	1.330
c_2	-0.668	-0.670	-0.670	1.761	2.078
c_3	-0.035	-0.037	-0.037	2.396	2.712
c_4	0.431	0.431	0.430	2.864	3.180
c_5	1.312	1.312	1.312	3.747	4.061
c_6	2.238	2.237	2.237	4.670	4.985
α	3.218	14.19	6.641	4.865	5.493
β		0.304	0.541	4.521	2.553
π	3.868	4.960	5.026	5.043	3.755

In a similar fashion, it is likely that greater attention was paid to items in product departments where the participant spent more time browsing. Hence, the memory traces for the prices of such items should be stronger than those for the prices of other items. If participants attempt to estimate a mean price for each store by recalling item prices, the estimated mean prices should then be biased towards item prices from those departments that were browsed for the longest time, due to the higher probability of recall. Browsing time could not be measured for each individual item, but only for each of the ten product departments, each of which contained fifteen items. A variable $Y_{S,i}$ was created to represent the approximate percentage of total browsing time in store S dedicated to each item i in product department d:

$$Y_{S,i} = \frac{t_{S,d}(i)}{15 \sum_{d=1}^{d=10} t_{S,d}(i)}$$

A weighting factor for each price was then created, so that:

$$W_{S,i} = 1 + \tau Y_{S,i}$$

$$\mu_S = \sum_{i=1}^{i=150} \left(\frac{W_{S,i}}{\sum_{i=1}^{i=150} W_{S,i}} P_{S,i} \right)$$

If the prices of items from departments that were browsed for a longer time are more salient, and hence bias the estimated mean price in each store, then a positive contribution $\tau > 0$ should be observed.

Time-weighting (TW) was applied to each of the five types of mean price comparison (A-E), and each model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.13 and the fitted parameter values are shown in Table 5.14. Time-weighting all five model types improved the model fits relative to the equivalent un-weighted versions of Model 1 for pre-checkout ratings, but not for post-checkout ratings. The best-fitting models for pre-checkout ratings were the Luce choice rule comparison (1D+TW) and the logistic comparison (1E+TW). The log-likelihoods of all five model types were again almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. For the pre-checkout rating models, all five models predict an increased probability of a high rating as the mean price in the first store decreases or as the mean price in the second store increases ($\alpha > 0$) but – unexpectedly - mean price estimates biased toward the prices of items that were browsed for a shorter time ($\tau < 0$).

TABLE 5.13
Goodness-of-fit measures for time-weighted versions of Model 1.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
1A+TW: Linear	-1482.4	2980.9	-1643.2	3302.3
1B+TW: Stevens	-1480.3	2978.6	-1643.2	3304.3
1C+TW: Fechner	-1482.4	2982.9	-1643.2	3304.3
1D+TW: Luce	-1480.1	2978.1	-1643.2	3304.4
1E+TW: Logistic	-1480.1	2978.2	-1643.2	3304.3

TABLE 5.14
Fitted parameter values for time-weighted versions of Model 1.

	Pre-Checkout				
Parameter	1A+TW	1B+TW	1C+TW	1D+TW	1E+TW
c_1	-1.906	-1.909	-1.906	-0.255	-0.163
c_2	-1.082	-1.085	-1.082	0.570	0.662
c_3	-0.140	-0.140	-0.140	1.515	1.607
c_4	0.573	0.577	0.573	2.232	2.324
c_5	1.511	1.519	1.511	3.174	3.266
c_6	2.646	2.648	2.646	4.305	4.396
α	3.100	0.067	422 330	3.308	3.492
β		5.008	136 250	9.903	5.007
τ	-7.935	-4.441	-7.936	-5.495	-5.239
	Post-Checkout				
Parameter	1A+TW	1B+TW	1C+TW	1D+TW	1E+TW
c_1	-1.373	-1.373	-1.373	2.228	87.87
c_2	-0.652	-0.652	-0.652	2.948	88.59
c_3	-0.044	-0.044	-0.044	3.557	89.20
c_4	0.405	0.405	0.405	4.006	89.65
c_5	1.251	1.251	1.251	4.852	90.50
c_6	2.139	2.139	2.139	5.740	91.385
α	3.059	4.066	31.73	7.201	178.49
β		0.826	8.658	3.077	0.069
τ	3.293	3.341	3.339	3.437	3.293

Finally, it is likely that the total cost of the basket of items purchased in each store (C_S) will also influence judgments of relative price, especially post-checkout once the total cost has been presented to the participant. Hence, in addition to the impact of item prices already discussed, the basket totals may have an additional influence on price judgments, with participants making a comparison between the basket costs in each store and then integrating this information into their overall judgment of relative price. As suggested in the Information Integration Theory literature described in Chapter 1, an additive integration process is likely to perform best (N. H. Anderson, 1965):

$$f(\cdot) = f(\cdot)_{item\,prices} + f(\cdot)_{basket\,costs}$$

In later models the item price information and basket price comparisons were combined, but initially the basket cost effect was considered in isolation, using a constant item price valuation function:

$$f(\cdot)_{item\,prices} = 0$$

Five different types of basket cost valuation functions were modelled, based upon the same comparison processes already tested for the mean price estimate comparisons. The first type (A) is based upon the linear difference between the two basket costs:

$$f(\cdot)_{basket\,costs} = \gamma(C_B - C_A)$$

The second type (B) is based upon the difference in basket costs raised to a power, as suggested by Stevens' Law:

$$f(\cdot)_{basket\,costs} = \gamma\left({C_B}^{\delta} - {C_A}^{\delta}\right)$$

The third type (C) is based upon the natural logarithm of the ratio of the basket costs, as suggested by Fechner's Law:

$$f(\cdot)_{basket\ costs} = \gamma \ln\left(\frac{C_B + \delta}{C_A + \delta}\right)$$

The fourth type (D) is based upon the Luce choice rule:

$$f(\cdot)_{basket\ costs} = \gamma\left(\frac{{C_B}^{\delta}}{{C_A}^{\delta} + {C_B}^{\delta}}\right)$$

The fifth type (E) is based upon the logistic function:

$$f(\cdot)_{basket\ costs} = \gamma\left(\frac{e^{\delta C_B}}{e^{\delta C_A} + e^{\delta C_B}}\right)$$

In all cases, the shape parameter was constrained so that $\delta \geq 0$. Using one or more of these five types of valuation function it is possible to predict judgments that exhibit increasing, decreasing, or constant sensitivity to differences in basket cost.

Each basket cost comparison (BC) model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.15 and the fitted parameter values are shown in Table 5.16. Adding basket cost comparisons does not improve the model fits relative to the baseline model (Model 0) for pre-checkout ratings, but significantly improves the fit for post-checkout ratings using any of the five types of basket cost comparison. The best-fitting model for basket cost comparisons was the logistic comparison (BC(E)). The post-checkout log-likelihoods of all five BC models are similar in size to the pre-checkout log-likelihoods, suggesting that the previously observed difference was largely due to participants making a basket cost comparison after the checkout. For the post-checkout rating models, all five models predict an increased probability of a high rating as the basket cost in the first store decreases or as the basket cost in the second store increases ($\gamma > 0$).

TABLE 5.15
Goodness-of-fit measures for basket cost comparison models.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
BC(A): Linear	-1560.1	3134.2	-1625.6	3265.2
BC(B): Stevens	-1560.1	3136.2	-1598.5	3213.1
BC(C): Fechner	-1560.1	3136.1	-1593.3	3202.6
BC(D): Luce	-1560.3	3136.7	-1540.6	3097.2
BC(E): Logistic	-1559.4	3134.7	-1534.9	3085.8

TABLE 5.16
Fitted parameter values for basket cost comparison models.

	Pre-Checkout				
Parameter	BC(A)	BC(B)	BC(C)	BC(D)	BC(E)
c_1	-1.713	-1.713	-1.714	-1.677	-1.516
c_2	-0.991	-0.991	-0.991	-0.955	-0.792
c_3	-0.113	-0.113	-0.113	-0.077	0.087
c_4	0.556	0.556	0.556	0.592	0.756
c_5	1.426	1.426	1.426	1.461	1.627
c_6	2.390	2.390	2.389	2.425	2.589
γ	0.002	0.002	0.598	0.075	0.401
δ		0.982	256.4	4.866	0.039
	Post-Checkout				
Parameter	BC(A)	BC(B)	BC(C)	BC(D)	BC(E)
c_1	-1.358	-1.425	-1.433	-0.643	-0.612
c_2	-0.636	-0.677	-0.681	0.152	0.184
c_3	-0.017	-0.037	-0.037	0.858	0.891
c_4	0.445	0.439	0.441	1.379	1.417
c_5	1.320	1.338	1.346	2.332	2.381
c_6	2.258	2.302	2.316	3.302	3.361
γ	0.026	1.300	3.170	1.853	1.918
δ		0.357	28.51	15.06	0.329

The fitted parameter values for the three sampling difference models are generally consistent with a plausible causal process of sampling and judgment. Purchased items presumably receive more attention than un-purchased items, making them easier to recall subsequently, so the prices of purchased items receive greater weight in the overall comparative judgment. Before the checkout the second basket cost is unknown, so differences in basket cost could only have a large impact after the checkout. Only the impact of browsing time is counter-intuitive, with longer-browsed items receiving less and not more weight in the overall judgment. Given that all three sampling differences – purchase-weighting, time-weighting, and basket cost comparison – appear to offer an improvement in fit over a simple un-weighted mean price model, the logistic versions of each model were combined in different permutations to determine whether any incremental improvement in fit is observed. Firstly, the purchase-weighting and time-weighting were combined by adjusting the weighting parameters:

$$W_{S,i} = 1 + \pi X_{S,i} + \tau Y_{S,i}$$

Secondly, the purchase-weighting and time-weighting models were combined with the basket cost comparisons using additive integration, such that:

$$f(\cdot) = f(\cdot)_{item\,prices} + \gamma\left(\frac{e^{\delta C_B}}{e^{\delta C_A} + e^{\delta C_B}}\right)$$

Each combination model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.17 and the fitted parameter values are shown in Table 5.18. Combining sampling difference terms improves the fit for three of the four pre-checkout ratings relative to the best single-term model (1E+PW). The best-fitting model for pre-checkout ratings was the logistic estimated mean comparison model, with purchase-weighting, time-weighting, and a basket cost comparison (1E+PW+TW+BC(E)) with an AIC of 2973.9. Combining sampling difference terms also improves the fit for three of the four post-checkout ratings relative to the best single-term model (BC(E)). The best-fitting model for post-checkout ratings was also the logistic estimated mean comparison model, with purchase-weighting and a basket cost comparison (1E+PW+BC(E)) with an AIC of 2966.6. Adding time-weighting did not further improve the fit of the post-checkout model.

TABLE 5.17
Goodness-of-fit measures for Model 1E with sampling differences.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
1E+PW+TW	-1477.2	2974.4	-1608.2	3236.4
1E+PW+BC(E)	-1476.6	2975.3	-1472.3	2966.6
1E+TW+BC(E)	-1478.9	2979.8	-1477.2	2976.4
1E+PW+TW+BC(E)	-1475.0	2973.9	-1472.2	2968.4

TABLE 5.18
Fitted parameter values for Model 1E with sampling differences.

	Pre-Checkout			
Parameter	1E+PW+TW	1E+PW+BC(E)	1E+TW+BC(E)	1E+PW+TW+BC(E)
c_1	-0.286	-0.373	-0.158	-0.311
c_2	0.544	0.454	0.667	0.519
c_3	1.494	1.405	1.612	1.471
c_4	2.213	2.124	2.330	2.191
c_5	3.157	3.070	3.274	3.140
c_6	4.281	4.203	4.412	4.273
α	3.259	3.244	3.610	3.368
β	5.847	6.246	4.878	5.796
π	0.364	0.439		0.447
τ	-5.541		-5.028	-5.285
γ		-0.167	-0.111	-0.161
δ		1.440	1.876	1.515
	Post-Checkout			
Parameter	1E+PW+TW	1E+PW+BC(E)	1E+TW+BC(E)	1E+PW+TW+BC(E)
c_1	1.316	1.069	1.775	1.097
c_2	2.063	1.933	2.637	1.962
c_3	2.697	2.687	3.386	2.716
c_4	3.165	3.249	3.943	3.277
c_5	4.047	4.278	4.964	4.307
c_6	4.972	5.308	5.989	5.335
α	5.466	3.853	5.102	3.904
β	2.571	3.777	2.454	3.703
π	3.801	0.741		0.738
τ	3.127		-1.397	-1.557
γ		1.681	1.831	1.686
δ		0.424	0.358	0.421

These results suggest that adding sampling difference terms – purchase-weighting, time-weighting, and basket cost comparisons – improves the model fit relative to the representative participant models. However, this was with an otherwise un-weighted mean price estimate for each store, whereas the earlier results showed that weighting the mean price estimates either by the magnitude of each price or by the typicality of each price gives better representative participant rating predictions. Hence, the three sampling difference terms were also added to the logistic typicality-weighted model (3E), both separately and in combination, to test whether including sampling difference terms continues to offer improved predictions. In order to combine purchase-weighting or time-weighting with typicality-weighting of mean price estimates, a weighting factor for each price was created, so that:

$$\mu_S = \sum_{i=1}^{i=150} \left(\frac{W_{S,i}}{\sum_{i=1}^{i=150} W_{S,i}} P_{S,i} \right)$$

The weighting factor used for purchase-weighting was:

$$W_{S,i} = \left(1 + \pi X_{S,i}\right) \times T_{S,i}$$

The weighting factor used for time-weighting was:

$$W_{S,i} = \left(1 + \tau Y_{S,i}\right) \times T_{S,i}$$

The weighting factor used for both purchase-weighting and time-weighting was:

$$W_{S,i} = \left(1 + \pi X_{S,i} + \tau Y_{S,i}\right) \times T_{S,i}$$

As before, the basket cost comparisons were integrated in an additive fashion:

$$f(\cdot) = f(\cdot)_{item\ prices} + \gamma \left(\frac{e^{\delta \mathrm{C}_B}}{e^{\delta \mathrm{C}_A} + e^{\delta \mathrm{C}_B}} \right)$$

All seven permutations of typicality-weighted logistic models were fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.19 and the fitted parameter values are shown in Table 5.20. For the pre-checkout rating models, six of the seven models have improved AIC scores relative to the representative participant model (3E). The best-fitting model for pre-checkout ratings was the typicality-weighted logistic comparison model, with purchase-weighting and a basket cost comparison (3E+PW+TW+BC(E)), with an AIC of 2916.7. For the post-checkout rating models, six of the seven models have improved AIC scores relative to the representative participant model (3E). The best-fitting model for post-checkout ratings was the typicality-weighted logistic comparison model, with purchase-weighting and a basket cost comparison (3E+PW+BC(E)), with an AIC of 2948.8. Adding time-weighting did not improve the AIC of either model. These two models are also the best fitting of all the pairing-independent models compared, so represent the benchmark against which the pairing-dependent models must be judged.

TABLE 5.19
Goodness-of-fit measures for Model 3E with sampling differences.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
3E+PW	-1450.5	2921.0	-1595.8	3211.5
3E+TW	-1449.3	2918.6	-1625.0	3270.1
3E+BC(E)	-1448.1	2918.2	-1464.4	2950.7
3E+PW+TW	-1448.3	2918.5	-1595.7	3213.3
3E+PW+BC(E)	-1446.5	2916.9	-1462.4	2948.8
3E+TW+BC(E)	-1446.4	2916.7	-1464.3	2952.6
3E+PW+TW+BC(E)	-1445.4	2916.7	-1462.3	2950.6

TABLE 5.20
Fitted parameter values for Model 3E with sampling differences.

	Pre-Checkout						
Parameter	3E+PW	3E+TW	3E +BC(E)	3E+PW +TW	3E+PW +BC(E)	3E+TW +BC(E)	3E+PW +TW +BC(E)
c_1	-0.275	-0.203	-0.259	-0.235	-0.298	-0.210	-0.258
c_2	0.562	0.635	0.577	0.605	0.540	0.628	0.583
c_3	1.546	1.618	1.560	1.590	1.527	1.611	1.571
c_4	2.292	2.366	2.309	2.339	2.277	2.362	2.323
c_5	3.271	3.347	3.293	3.321	3.262	3.349	3.311
c_6	4.413	4.494	4.453	4.463	4.418	4.509	4.467
α	3.407	3.551	3.618	3.494	3.583	3.715	3.666
β	7.617	7.039	7.282	7.265	7.541	6.926	7.219
ψ	46.16	47.22	44.93	47.681	45.15	46.29	46.45
π	0.294			0.328	0.473		0.490
τ		-9.569		-10.80		-8.947	-10.49
γ			-0.185		-0.220	-0.179	-0.215
δ			1.581		1.336	1.647	1.420

	Post-Checkout						
Parameter	3E+PW	3E+TW	3E +BC(E)	3E+PW +TW	3E+PW +BC(E)	3E+TW +BC(E)	3E+PW +TW +BC(E)
c_1	100.5	100.6	3.493	100.5	3.300	3.555	3.299
c_2	101.2	101.3	4.367	101.2	4.174	4.428	4.174
c_3	101.9	101.9	5.130	101.9	4.939	5.191	4.940
c_4	102.4	102.4	5.697	102.4	5.510	5.759	5.510
c_5	103.3	103.3	6.733	103.3	6.548	6.795	6.549
c_6	104.2	104.2	7.768	104.2	7.584	7.829	7.583
α	203.9	204.0	8.608	203.9	8.336	8.725	8.329
β	0.100	0.100	2.212	0.100	2.338	2.179	2.339
ψ	33.48	25.10	25.62	33.61	26.75	25.53	26.60
π	8.684			8.950	0.973		0.955
τ		6.034		7.805		-3.138	-4.104
γ			1.826		1.723	1.832	1.729
δ			0.342		0.380	0.340	0.377

5.3.2.3 Interpreting Model Parameters

To understand and interpret the selected pairing-independent model (3E+PW+BC(E)) and the fitted parameter values, I have explored the predictions of the model at typical values of the input data as well as the sensitivity of those predictions to changes in the input variables. The model can be decomposed into its three constituent parts: typicality-weighting (i.e. the price stimuli), purchase-weighting (i.e. purchasing behaviour), and a basket cost comparison. Each part can be explored independently, by holding the other two parts constant at the mean values observed in the experimental data. In reality the three parts are not truly independent – for example, increased purchasing of large price items also increases the total basket cost – but these second-order interdependencies will be ignored here for the sake of clear exposition.

The impact of the input prices was explored for a representative participant, with mean purchasing probability for each item in each store, and with the mean basket costs C_A = £50.25 and C_B = £50.88. The prices in the first store were fixed as the control store, Jones. The prices in the second store were varied to test how the model predictions changed in response to those variations. Figure 5.6 shows how the probability distribution for the pre-checkout ratings varies as a function of the discount in the second store. The probability of rating the second store as cheaper (R_{AB} = 1, 2, or 3) increases from 46% when the two stores have the same mean price to 96% when the second store is 30% cheaper on all items. Figure 5.7 shows the probability distributions for the post-checkout ratings. The probability of rating the second store as cheaper increases from 42% when the two stores have the same mean price to 95% when the second store is 30% cheaper on all items. Figures 5.8 and 5.9 show the observed mean pre- and post-checkout comparative price judgment ratings from each discount condition in Experiment 3, and the expected values from the model:

$$\mathrm{E}(R_{AB}) = \sum_{i=1}^{i=7} i \cdot p(R_{AB} = i)$$

The post-checkout ratings are strongly influenced by the basket costs in each store, hence they are noisier and less well-explained by the price variations than the pre-checkout ratings. The shapes of the pre- and post-checkout response curves are similar, although the pre-checkout ratings exhibit a diminishing response to the discount level in the large discount conditions. The post-checkout model appears to under-predict the impact of discount level on comparative price judgments. This is likely to be due to the impact of the discount level on the basket cost in the second store, whereas the basket costs were held constant in these figures.

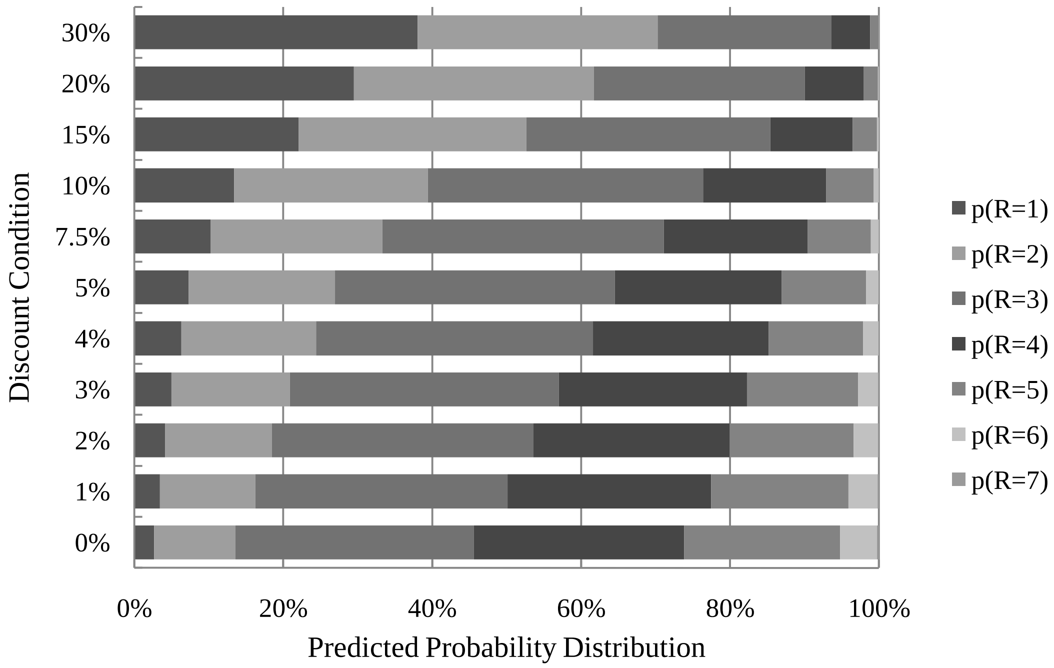

Figure 5.6: Predicted probability distribution for pre-checkout ratings as a function of the mean price difference between the two stores (Model 3E+PW+BC(E)).

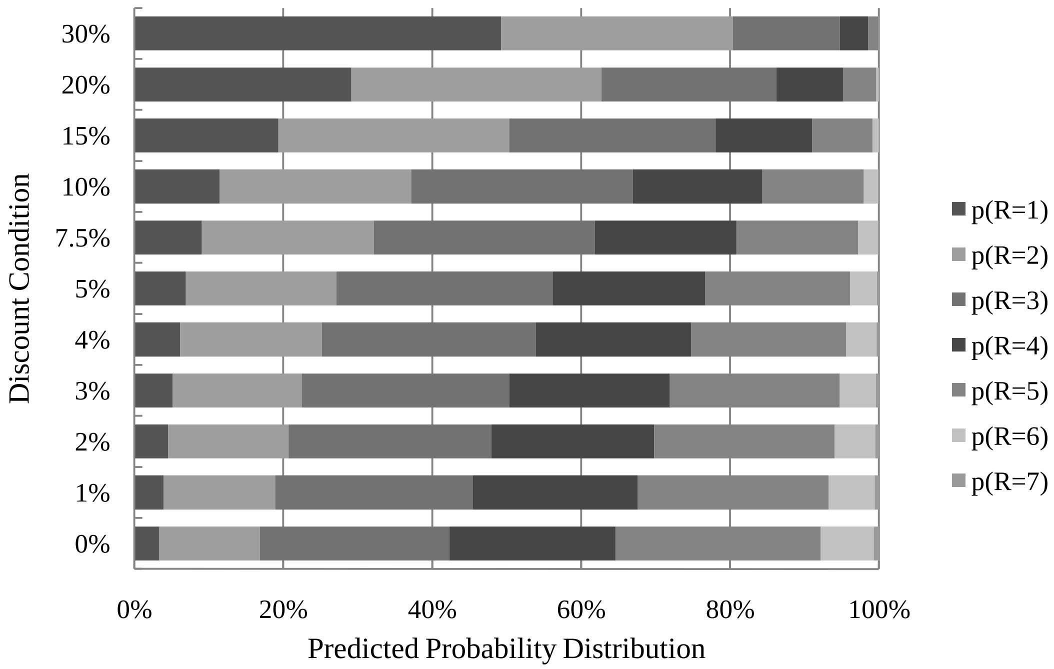

Figure 5.7: Predicted probability distribution for post-checkout ratings as a function of the mean price difference between the two stores (Model 3E+PW+BC(E)).

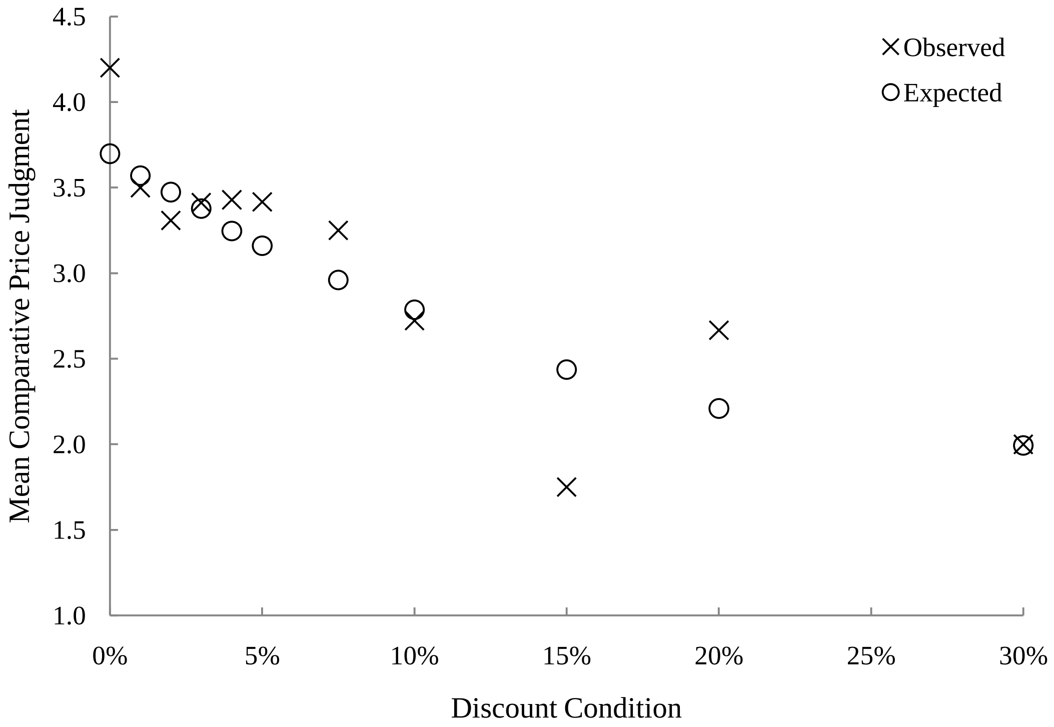

Figure 5.8: Observed and expected values of pre-checkout ratings as a function of the mean price difference between the two stores (Experiment 3 and Model 3E+PW+BC(E)).

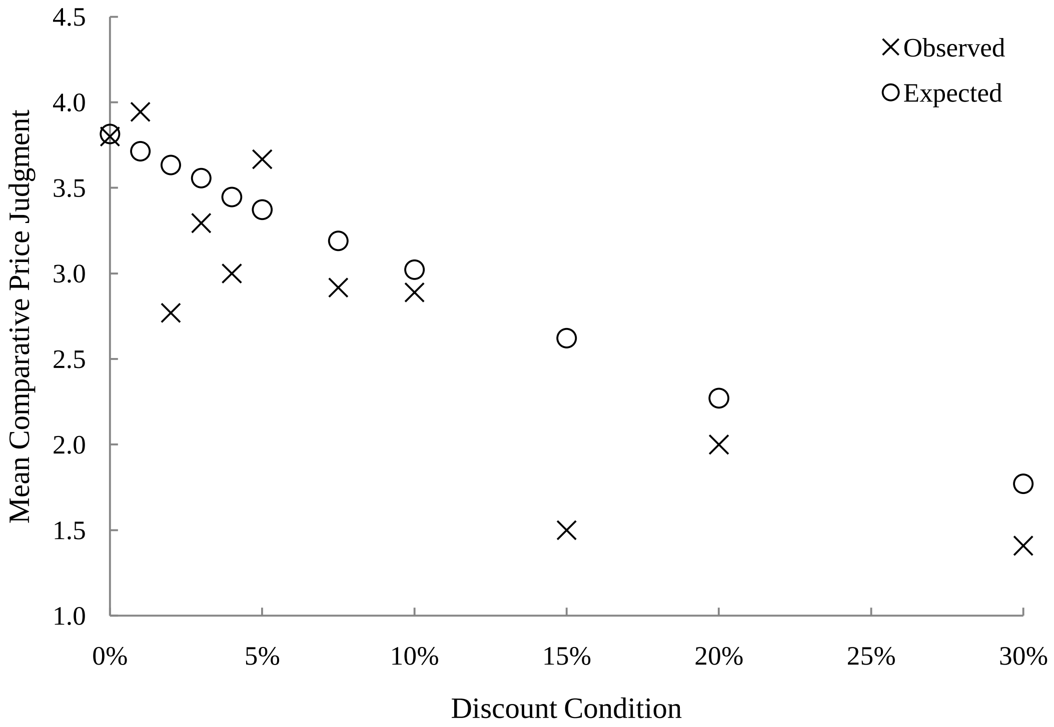

Figure 5.9: Observed and expected values of post-checkout ratings as a function of the mean price difference between the two stores (Experiment 3 and Model 3E+PW+BC(E)).

The model also predicts the pattern of comparative price judgments observed in the frequency and magnitude conditions of Experiment 4. Figure 5.10 shows how the predicted probability distributions for the pre-checkout ratings vary as a function of the frequency and magnitude of price advantages in the second store. When the magnitude of the price advantages in the second store is small (5%), the probability of rating the second store as cheaper (R_{AB} = 1, 2, or 3) increases from 42% when the second store is cheaper on 20% of the items to 58% when the second store is cheaper on 80% items. When the magnitude of the price advantages in the second store is large (20%), the probability of rating the second store as cheaper increases from 43% when the second store is cheaper on 20% of the items to 69% when the second store is cheaper on 80% items. Figure 5.11 shows the predicted probability distributions for the post-checkout ratings. The post-checkout ratings are less sensitive than the pre-checkout ratings to frequency increases: when the magnitude of the price advantages is small, the probability of rating the second store as cheaper increases from 40% to 51%; when the magnitude of the price advantages is large, the probability of rating the second store as cheaper increases from 41% to 60%. This is reflected in the typicality-weighting parameter values, with the similarity gradient in the pre-checkout model (ψ = 45) being much steeper than the similarity gradient in the post-checkout model (ψ = 27). Figures 5.12 and 5.13 show the observed mean pre- and post-checkout comparative price judgment ratings from each price condition in Experiment 4 and the expected values from the model. As before, the post-checkout ratings are noisier and less well-explained by the price variations than the pre-checkout ratings. The post-checkout model also appears to somewhat under-predict the impact of frequency on comparative price judgments.

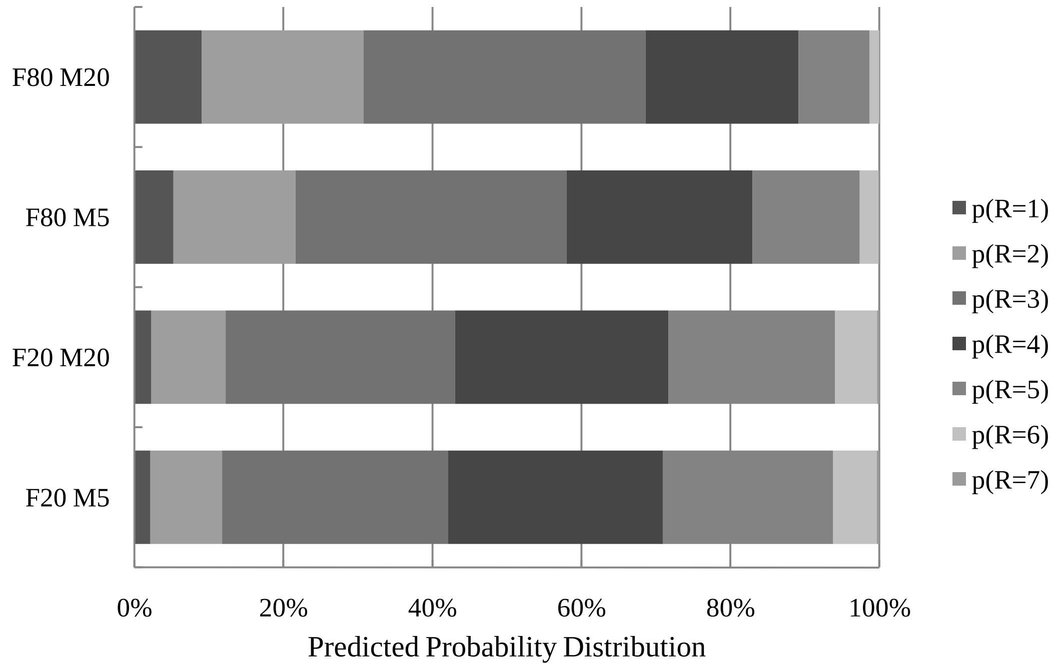

Figure 5.10: Predicted probability distribution for pre-checkout ratings as a function of the frequency and magnitude of price advantages in Store B (Model 3E+PW+BC(E)).

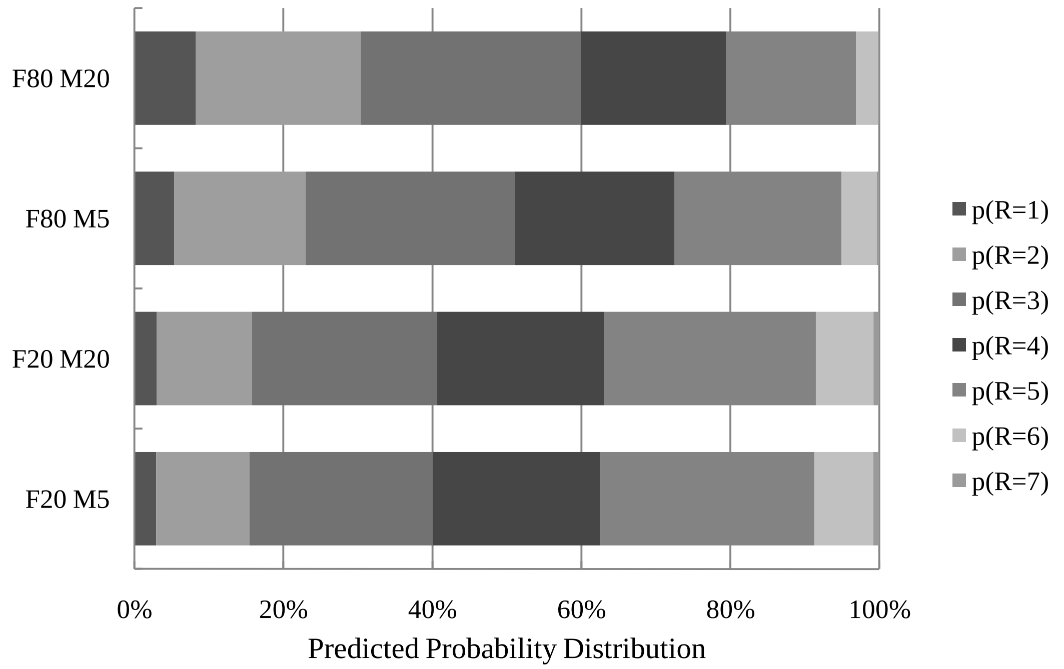

Figure 5.11: Predicted probability distribution for post-checkout ratings as a function of the frequency and magnitude of price advantages in Store B (Model 3E+PW+BC(E)).

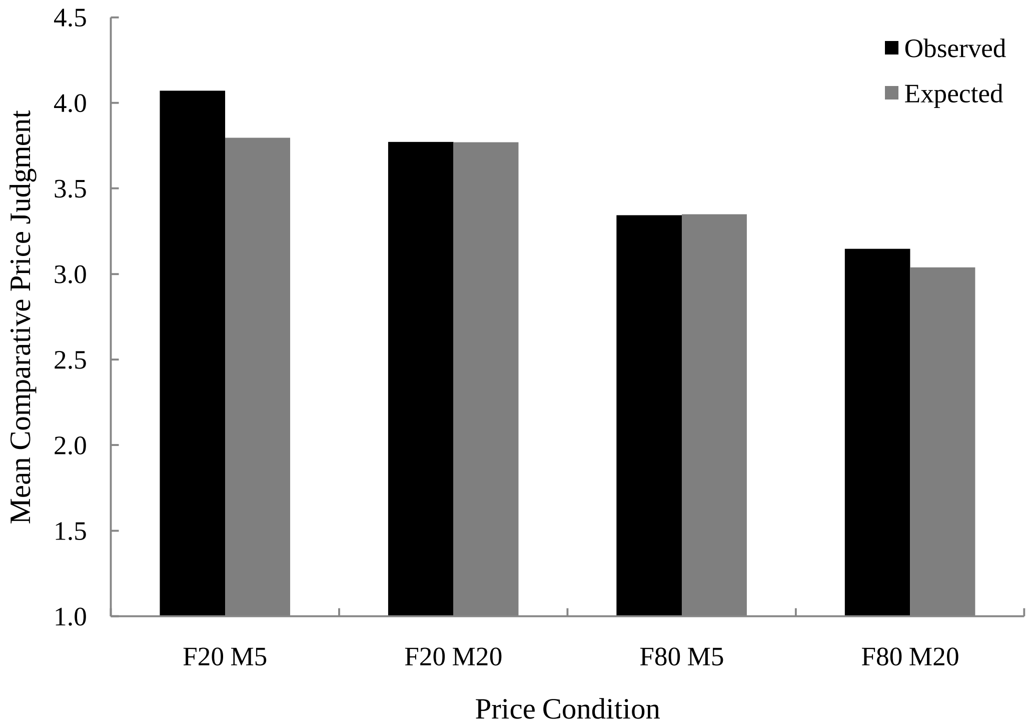

Figure 5.12: Observed and expected values of pre-checkout ratings as a function of the price distribution in each store (Experiment 4 and Model 3E+PW+BC(E)).

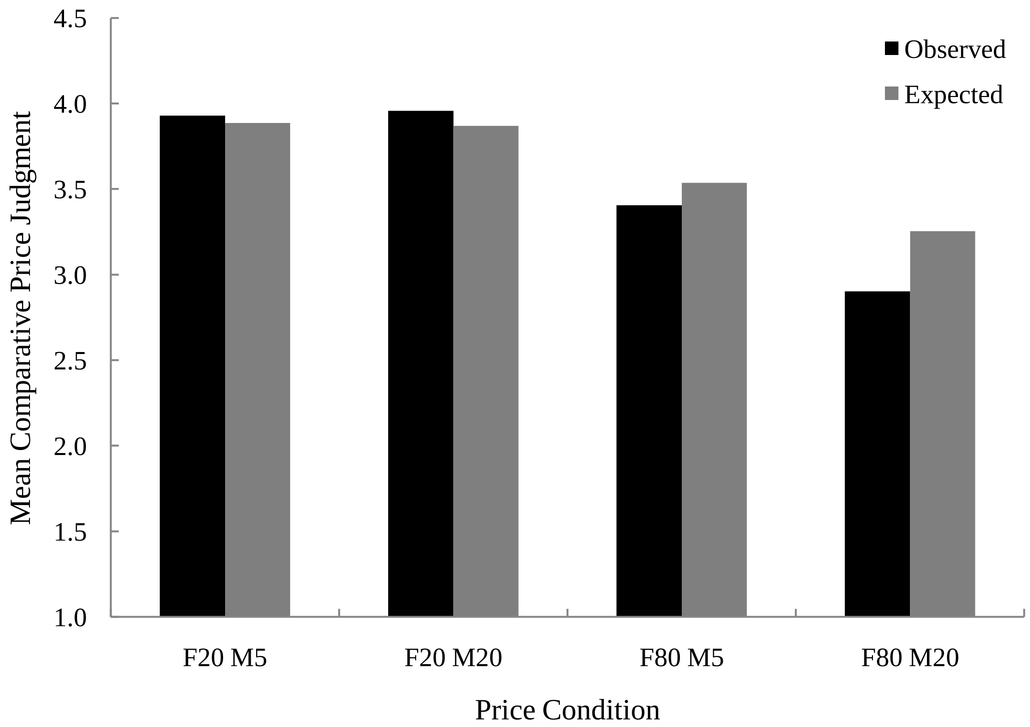

Figure 5.13: Observed and expected values of post-checkout ratings as a function of the price distribution in each store (Experiment 4 and Model 3E+PW+BC(E)).

The impact of the sampling differences terms can be explored in a similar fashion. By setting the item prices in the test store to match the control store, and by holding the total basket costs C_A and C_B at the mean observed values, the impact of variations in purchasing behaviour can be independently tested. The purchasing behaviour in the first store was also held constant at the mean observed item purchase probabilities. Figure 5.14 shows the predicted probability distributions for the pre-checkout ratings for five different patterns of purchasing behaviour in the second store: (i) the mean observed purchase probability for each item; (ii) no items purchased; (iii) all items purchased; (iv) only the 31 smallest price items purchased; and (v) only the 32 largest price items purchased. In the first three cases, the probability of rating the second store as cheaper (R_{AB} = 1, 2, or 3) is very similar, at 46%, 42% and 42% respectively. When only the 31 smallest price items are purchased in the second store, the probability of rating the second store as cheaper rises to 64%, whereas when only the 32 largest price items are purchased, the probability falls to just 13%. Figure 5.15 shows the predicted probability distributions for the post-checkout ratings, for the same five patterns of purchasing behaviour. Once again, in the first three cases, the probability of rating the second store as cheaper is very similar, at 42%, 37% and 37% respectively. The impact of purchasing only the smallest or largest price items is even more extreme for the post-checkout ratings, with the probability of rating the second store as cheaper rising to 67% when only the 31 smallest price items are purchased, and falling to just 4% when only the 32 largest price items are purchased. This is reflected in the parameter values, with purchased items having 50% more impact than un-purchased items in the pre-checkout model (π = 0.47), and almost 100% more impact than un-purchased items in the post-checkout model (π = 0.97).

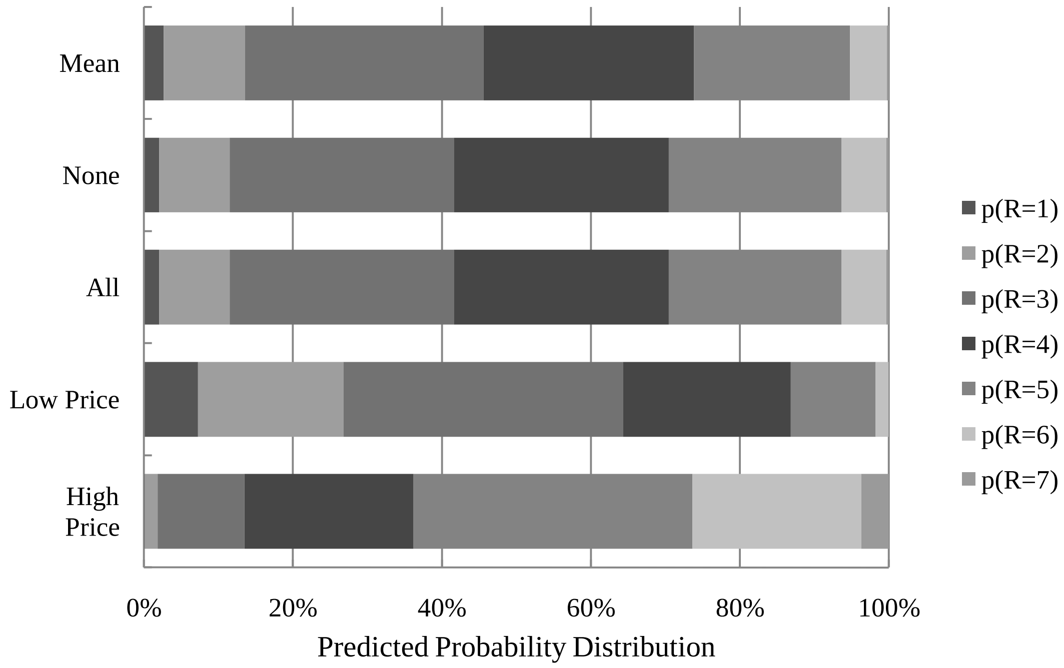

Figure 5.14: Predicted probability distribution for pre-checkout ratings as a function of the purchasing behaviour in Store B (Model 3E+PW+BC(E)).

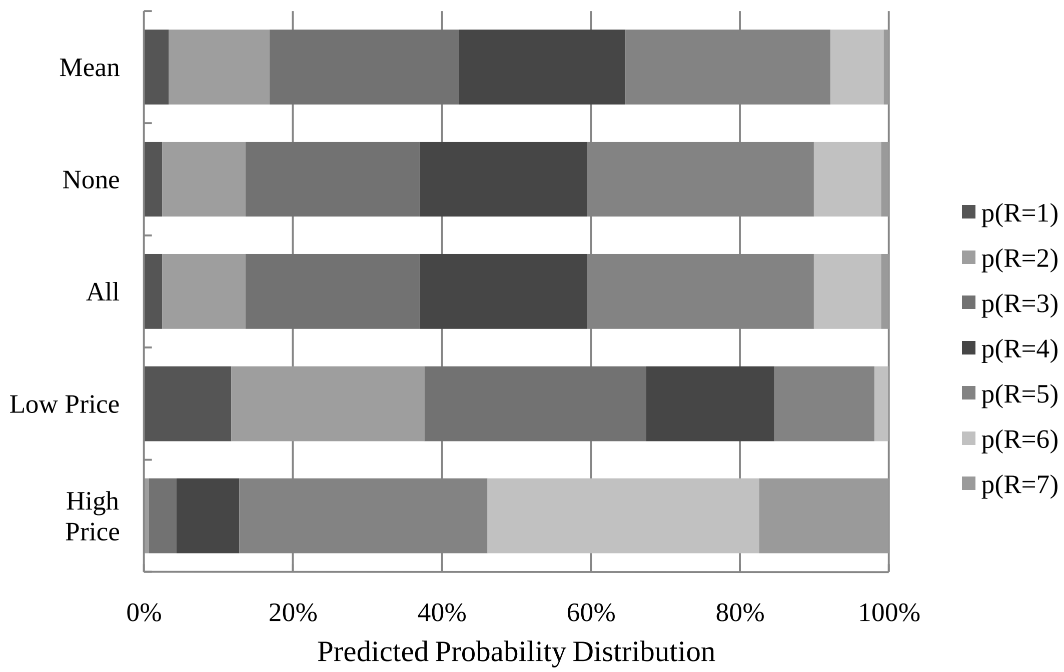

Figure 5.15: Predicted probability distribution for post-checkout ratings as a function of the purchasing behaviour in Store B (Model 3E+PW+BC(E)).

Finally, by setting the item prices in the test store to match the control store, and by holding the purchasing behaviour in each store at the mean observed values, the impact of variations in basket cost can be independently tested. The total basket cost in the first store was also held constant at the mean observed value of £50.25. Figure 5.16 shows the predicted probability distributions for the pre-checkout ratings for five different levels of basket cost in the second store: (i) £30; (ii) £40; (iii) £50; (iv) £60; and (v) £70. The impact of the basket cost on pre-checkout ratings is small and in a counter-intuitive direction, with the probability of rating the second store as cheaper (R_{AB} = 1, 2, or 3) increasing from 40% when C_B = £30 to 48% when C_B = £70. One possible explanation is therefore that participants purchased slightly more items when they perceived the second store to be cheaper, increasing the basket cost at the same time as increasing the probability of rating the second store as cheaper.[20] Figure 5.17 shows the predicted probability distributions for the post-checkout ratings, for the same five values of basket cost C_B. The impact of the basket cost on post-checkout ratings is much greater and in the intuitive direction, with the probability of rating the second store as cheaper falling from 78% when C_B = £30 to 17% when C_B = £70. This difference is reflected in the parameter values, with a much larger scaling parameter for the basket cost comparison term in the post-checkout rating model ($\gamma = 1.72$) compared to the pre-checkout rating model ($\gamma = -0.22$). Before the checkout, comparative price judgments are dominated by the impact of the price distributions in each store ($|\alpha| / |\gamma| = 16$), whilst after the checkout judgments are also strongly influenced by the basket cost comparison ($|\alpha| / |\gamma| = 5$). When the basket costs are correlated with price differences, as in Experiment 3, the basket cost comparison will intensify the impact of price

[20] This hypothesis could be tested directly in a future experiment, for example by fixing the purchases by providing participants with a shopping list to follow, or alternatively by providing false feedback on basket prices after the checkout.

differences on comparative judgments.

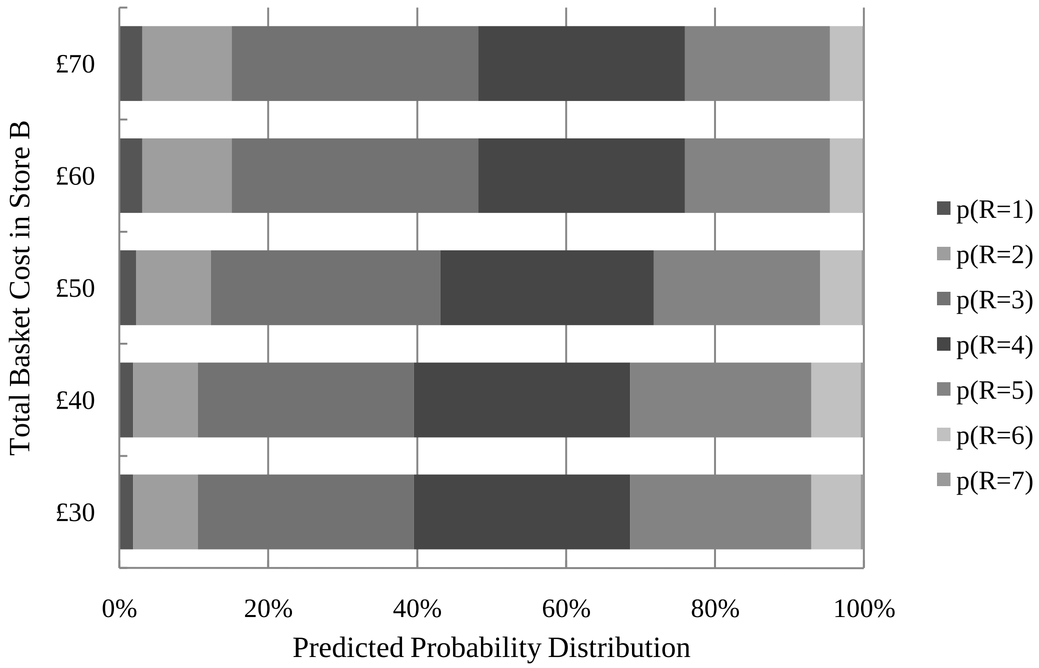

Figure 5.16: Predicted probability distribution for pre-checkout ratings as a function of the total basket cost in Store B (Model 3E+PW+BC(E)).

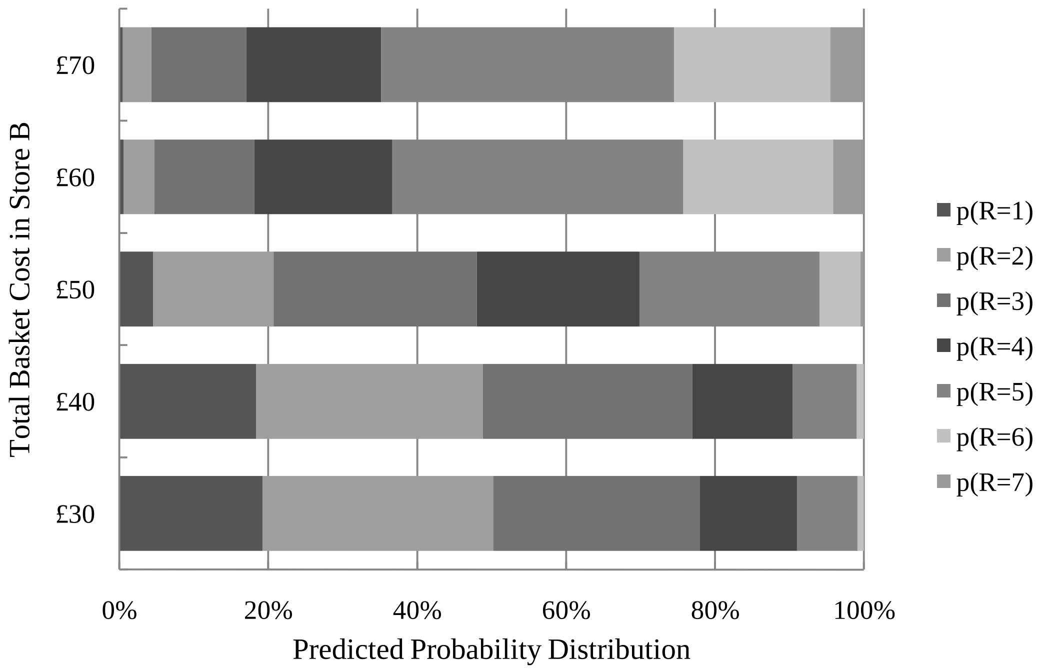

Figure 5.17: Predicted probability distribution for post-checkout ratings as a function of the total basket cost in Store B (Model 3E+PW+BC(E)).

5.3.3 Pairing-Dependent Judgments

5.3.3.1 Representative Participant Models

An alternative plausible set of cognitive process models assume that participants make pair-wise comparisons between the prices of each item in the two stores and then integrate those individual comparisons to make an overall comparative price judgment rating. The most extreme version of these pairing-dependent models (Model 4) assumes (i) that participants are only sensitive to the number of price differences between the two stores (i.e. frequency), regardless of the magnitude of those price differences and (ii) that all price differences are noticed and equally-weighted in estimating the frequency of price advantages in each store, F_S:

$$P_{A,i} < P_{B,i} \quad \rightarrow \quad F_{A,i} = 1 \; and \; F_{B,i} = 0$$

$$P_{A,i} > P_{B,i} \quad \rightarrow \quad F_{A,i} = 0 \; and \; F_{B,i} = 1$$

$$P_{A,i} = P_{B,i} \quad \rightarrow \quad F_{A,i} = F_{B,i} = 0$$

$$F_S = \sum_{i=1}^{i=150} F_{S,i}$$

The same five basic types of valuation function as before were compared to determine which best describes how participants compared the relative frequency of price advantages in each store. The first type (A) is based upon the linear difference between the two frequency counts:

$$f(\cdot) = \alpha(F_A - F_B)$$

The second type (B) is based upon the difference in frequency counts raised to a power, as suggested by Stevens' Law:

$$f(\cdot) = \alpha\left({F_A}^{\beta} - {F_B}^{\beta}\right)$$

The third type (C) is based upon the natural logarithm of the ratio of the frequency counts, as suggested by Fechner's Law:

$$f(\cdot) = \alpha \ln\left(\frac{F_A + \beta}{F_B + \beta}\right)$$

The fourth type (D) is based upon the Luce choice rule (adjusted so that the term in brackets is equal to 0.5 when the prices in both stores are identical, and hence $F_A = F_B = 0$):

$$f(\cdot) = \alpha\left(\frac{{F_A}^{\beta}}{{F_A}^{\beta} + {F_B}^{\beta}}\right)$$

The fifth type (E) is based upon the logistic function, with a sigmoid relationship between the valuation and the difference in frequency counts in each store, as suggested by Buyukkurt (1986):

$$f(\cdot) = \alpha\left(\frac{e^{\beta F_A}}{e^{\beta F_A} + e^{\beta F_B}}\right)$$

In all cases, the shape parameter was constrained so that $\beta \geq 0$. Using one or more of these five types of valuation function it is possible to predict judgments that exhibit increasing, decreasing, or constant sensitivity to differences in frequency. Clearly this family of models cannot predict the observed discount effect from Experiment 3, as the frequency, F_S, is always 150 in the discounted store regardless of the level of discount. Nonetheless, it serves as a useful yardstick against which to compare more complex cognitive process models.

Each type of Model 4 was fit twice, to both pre- and post-checkout ratings.

The goodness-of-fit measures for the models are shown in Table 5.21 and the fitted parameter values are shown in Table 5.22. All five model types offer an improvement over the baseline model. The best-fitting models for pre-checkout ratings were the power law comparison (4B) and the Fechner Law comparison (4C). The log-likelihood of all five model types was almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. A log-likelihood score of -1486.5 implies that the mean likelihood of the observed ratings is 21%, compared to a random choice probability of 1 in 7 or 14%. For the pre-checkout rating models, all five models predict an increased probability of a high rating ("Compared to the first supermarket, I thought that the second store was a lot more expensive") as the frequency of price advantages in the first store increases or as the frequency in the second store decreases ($\alpha > 0$).

TABLE 5.21
Goodness-of-fit measures for Model 4.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
4A: Linear	-1488.2	2990.5	-1670.6	3355.2
4B: Stevens	-1486.5	2988.9	-1667.7	3351.5
4C: Fechner	-1486.7	2989.5	-1667.9	3351.9
4D: Luce	-1487.6	2991.1	-1668.7	3353.4
4E: Logistic	-1488.2	2992.5	-1670.6	3357.2

TABLE 5.22
Fitted parameter values for Model 4.

	Pre-Checkout				
Parameter	4A	4B	4C	4D	4E
c_1	-1.874	-1.878	-1.877	-1.226	60.70
c_2	-1.098	-1.097	-1.096	-0.446	61.47
c_3	-0.148	-0.145	-0.145	0.505	62.42
c_4	0.582	0.585	0.585	1.234	63.15
c_5	1.526	1.529	1.529	2.177	64.10
c_6	2.568	2.577	2.577	3.224	65.14
α	0.004	0.061	0.242	1.298	125.1
β		0.476	10.64	0.651	0.0001
	Post-Checkout				
Parameter	4A	4B	4C	4D	4E
c_1	-1.329	-1.330	-1.330	-0.792	48.59
c_2	-0.643	-0.641	-0.641	-0.103	49.28
c_3	-0.044	-0.041	-0.041	0.496	49.88
c_4	0.402	0.405	0.405	0.942	50.32
c_5	1.230	1.235	1.235	1.771	51.15
c_6	2.082	2.090	2.090	2.625	52.00
α	0.003	0.105	0.136	1.073	99.84
β		0.328	2.876	0.476	0.0001

At the other extreme, participants could be sensitive only to the magnitude of paired-item price differences, regardless of whether those differences are distributed over a small or large number of items. In the case of Experiments 3 and 4, with an identical set of products in each store, this is indistinguishable from the case where participants make independent judgments of the mean price in each store, as already modelled. More plausibly, participants may be sensitive to the frequency of price differences but be more likely to notice and/or be able to recall them when the magnitude of the price difference is larger. Four different candidate cognitive judgment models were tested (Model 5). Each assumes that the comparison process occurs in two stages. Firstly, each pair of prices is compared. The greater the difference between the two prices, the greater the probability that the participant will notice the difference and/or the greater the impact of that item on the overall judgment. Secondly, each of the separate item price comparisons is integrated to make an overall comparative judgment about the two stores. The first pair-wise item price comparison process (A) is based on the difference between each item price raised to a power, as suggested by Stevens' Law:

$$F_i = {P_{B,i}}^{\beta} - {P_{A,i}}^{\beta}$$

The second pair-wise comparison process (B) is based on the natural logarithm of the ratio of the price of an item in each store, as suggested by Fechner's Law:

$$F_i = \ln\left(\frac{P_{B,i} + \beta}{P_{A,i} + \beta}\right)$$

The third pair-wise comparison process (C) is based on the Luce choice rule:

$$F_i = \frac{{P_{B,i}}^{\beta}}{{P_{A,i}}^{\beta} + {P_{B,i}}^{\beta}}$$

The fourth pair-wise comparison process (D) is based on the logistic function, as suggested by Buyukkurt:

$$F_i = \frac{e^{\beta P_{B,i}}}{e^{\beta P_{A,i}} + e^{\beta P_{B,i}}}$$

In each case, the pair-wise comparisons were integrated in an additive fashion, as suggested in the Information Integration Theory literature (N. H. Anderson, 1965):

$$f(\cdot) = \alpha \sum_{i=1}^{i=150} F_i$$

A linear model was not fitted again as it is equivalent to the pairing-independent, accurate mean estimation model (Model 1A) tested earlier, given the identical range of items in each of the two stores.

Each type of Model 5 was fit twice, to both pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.23 and the fitted parameter values are shown in Table 5.24. All four model types offer an improvement over both the baseline model and the equivalent versions of Model 4. The best-fitting model for pre-checkout ratings was the logistic pair-wise comparison process (5D). The best-fitting models for post-checkout ratings were the Luce choice rule pair-wise comparison process (5C) and the logistic pair-wise comparison process (5D). The log-likelihood of all four model types was a significantly worse fit for the pre-checkout ratings than for the pre-checkout ratings. For both the pre- and post-checkout rating models, all four types predict an increased probability of a high rating as the frequency of price advantages in the first store increases or as the frequency in the second store decreases ($\alpha > 0$).

TABLE 5.23
Goodness-of-fit measures for Model 5.

Model	Pre-Checkout Log-Likelihood	Pre-Checkout AIC	Post-Checkout Log-Likelihood	Post-Checkout AIC
5A: Stevens	-1473.0	2962.0	-1631.0	3278.0
5B: Fechner	-1472.6	2961.2	-1631.0	3278.0
5C: Luce	-1464.7	2945.5	-1626.9	3269.9
5D: Logistic	-1459.6	2935.2	-1626.8	3269.5

TABLE 5.24
Fitted parameter values for Model 5.

	Pre-Checkout			
Parameter	5A	5B	5C	5D
c_1	-1.934	-1.935	-0.041	1.308
c_2	-1.102	-1.102	0.789	2.141
c_3	-0.147	-0.147	1.759	3.116
c_4	0.573	0.573	2.489	3.851
c_5	1.520	1.521	3.450	4.818
c_6	2.665	2.667	4.581	5.956
α	1.931	0.048	0.025	0.044
β	0.019	0.233	6.789	3.004
	Post-Checkout			
Parameter	5A	5B	5C	5D
c_1	-1.397	-1.397	1.513	4.604
c_2	-0.665	-0.665	2.246	5.337
c_3	-0.047	-0.047	2.870	5.961
c_4	0.409	0.409	3.329	6.421
c_5	1.264	1.265	4.189	7.281
c_6	2.161	2.161	5.086	8.178
α	67.87	0.038	0.039	0.080
β	0.0005	0.026	3.955	1.266

In Model 5 each item contributes equally when the pair-wise item price comparisons are integrated to make an overall comparison between the two stores. However – as with the pairing-independent models - certain item prices may be more salient or more easily recalled, so may have a proportionately larger influence in the overall judgment. For example, larger prices may carry greater weight in the integration process, as they would have a greater impact on the total cost of a basket of items (Model 6). A weighting factor for each item was created, such that:

$$W_i = 1 + \epsilon\left(P_{A,i} + P_{B,i}\right)$$

$$f(\cdot) = \alpha \sum_{i=1}^{i=150} \left(\frac{W_i}{\sum_{i=1}^{i=150} W_i} F_i \right)$$

The same four pair-wise item price comparison processes were used, and each type of Model 6 was fit twice, to both pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.25 and the fitted parameter values are shown in Table 5.26. Three of the four model types offer an improvement over the equivalent un-weighted versions of Model 5, although the best-fitting un-weighted model (5D) has a lower AIC score than any type of Model 6. The best-fitting model for pre-checkout ratings was the Luce choice rule pair-wise comparison process (6C). The log-likelihood of all four model types was almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. For all the pre- and post-checkout rating models, all four types predict an increased probability of a high rating as the frequency of price advantages in the first store increases or as the frequency in the second store decreases ($\alpha > 0$), and for all but one of the models items with smaller prices carry more weight in the comparative judgment than items with larger prices ($\varepsilon < 0$).

TABLE 5.25
Goodness-of-fit measures for Model 6.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
6A: Stevens	-1464.2	2946.4	-1625.5	3269.0
6B: Fechner	-1465.5	2949.1	-1626.3	3270.5
6C: Luce	-1458.2	2934.4	-1626.4	3270.9
6D: Logistic	-1458.9	2935.7	-1626.6	3271.1

TABLE 5.26
Fitted parameter values for Model 6.

	Pre-Checkout			
Parameter	6A	6B	6C	6D
c_1	-1.951	-1.944	-0.485	1.035
c_2	-1.116	-1.112	0.349	1.869
c_3	-0.149	-0.149	1.324	2.844
c_4	0.580	0.577	2.062	3.581
c_5	1.539	1.532	3.031	4.548
c_6	2.682	2.670	4.169	5.686
α	1.500	519 300	2.951	5.992
β	1.613	111 360	13.60	2.393
ε	-0.038	-0.048	-0.747	0.123
	Post-Checkout			
Parameter	6A	6B	6C	6D
c_1	-1.406	-1.408	0.649	5.005
c_2	-0.672	-0.673	1.383	5.739
c_3	-0.047	-0.048	2.007	6.363
c_4	0.413	0.412	2.467	6.823
c_5	1.274	1.273	3.328	7.683
c_6	2.173	2.172	4.225	8.580
α	1.541	11 271 000	4.106	12.82
β	1.583	2 392 900	6.592	1.411
ε	-0.038	-0.048	-328 880	-0.030

As discussed earlier in this chapter, there is no strong theoretical justification for participants' integrated comparative price judgments to be biased toward small prices rather than large prices or vice versa. However, in order to make pair-wise price comparisons it is necessary to encode in memory and later retrieve prices from the first store. Items which are more typical are more easily recalled, and hence may carry more weight in the integrated comparative price judgment. Model 7 assumes that the probability of recall for a price is proportional to its typicality, using a weighting factor for each item such that:

$$W_i = 1 + \theta\left(T_{A,i} + T_{B,i}\right)$$

$$f(\cdot) = \alpha \sum_{i=1}^{i=150} \left(\frac{W_i}{\sum_{i=1}^{i=150} W_i} F_i \right)$$

The same four pair-wise item price comparison processes were used, and each type of Model 7 was fit twice, to both pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.27 and the fitted parameter values are shown in Table 5.28. All of the four model types offer an improvement over the equivalent un-weighted versions of Model 5. The best-fitting model for pre-checkout ratings was the logistic pair-wise comparison process (7D). The log-likelihood of all four model types was almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. For all the pre- and post-checkout rating models, all four types predict an increased probability of a high rating as the frequency of price advantages in the first store increases or as the frequency in the second store decreases ($\alpha > 0$). However, two of the models (7A and 7B) predict that items with more typical prices carry more weight in the comparative judgment ($\theta > 0$), while the other two models predict the opposite ($\theta < 0$). Overall, there is little evidence that price- or typicality-weighting the integration of the pair-wise item price comparisons offers a significant and consistent improvement in model fit or interpretability. Hence, the best pairing-dependent representative participant model assumes logistic comparisons between pairs of item prices, integrated in an un-weighted additive fashion (5D).

TABLE 5.27
Goodness-of-fit measures for Model 7.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
7A: Stevens	-1467.2	2954.4	-1627.4	3274.7
7B: Fechner	-1468.1	2956.2	-1627.3	3274.5
7C: Luce	-1459.4	2938.8	-1626.5	3272.9
7D: Logistic	-1456.8	2933.6	-1626.7	3273.3

TABLE 5.28
Fitted parameter values for Model 7.

	Pre-Checkout			
Parameter	7A	7B	7C	7D
c_1	-1.947	-1.945	-0.459	-0.884
c_2	-1.112	-1.112	0.375	-0.051
c_3	-0.149	-0.149	1.349	0.926
c_4	0.576	0.576	2.085	1.666
c_5	1.530	1.530	3.052	2.635
c_6	2.679	2.673	4.191	3.776
α	13.55	17.69	2.999	2.163
β	0.415	1.526	13.20	5.000
ψ	722 290	10.33	47.92	39.74
θ	173 660	211 880	-0.112	-0.209
	Post-Checkout			
Parameter	7A	7B	7C	7D
c_1	-1.404	-1.404	0.922	0.455
c_2	-0.670	-0.671	1.656	1.188
c_3	-0.048	-0.048	2.280	1.811
c_4	0.411	0.411	2.740	2.271
c_5	1.269	1.270	3.601	3.131
c_6	2.167	2.167	4.497	4.029
α	39.91	7.454	4.654	3.719
β	0.137	0.277	5.413	1.854
ψ	34 796	302 420	34.01	23.89
θ	275 730	236 750	-0.059	-0.129

5.3.3.2 Sampling Differences Models

As already discussed in relation to the pairing-independent models, it is plausible that each individual followed a slightly different judgment process because their browsing behaviour led them to sample the price information in a unique way. The same three measures of each participant's behaviour were used to characterise their personal sampling process: the basket of items purchased in each store; the time spent browsing each department in each store; and the total cost of the basket of items purchased in each store, which was shown at the checkout. As described in Chapter 1, memory traces are stronger when greater attention is paid to a stimulus at the moment of encoding and storage. It is likely that greater attention was paid to purchased items than non-purchased items, as the purchased items were selected and placed into the basket by the participant. Hence, the memory traces for the prices of purchased items should be stronger than those for the prices of un-purchased items. If participants make an overall price comparison by comparing the frequency with which each store is cheaper on pairs of matching items, the judgments should be biased towards purchased items due to the higher probability of recall. A binary variable $X_{S,i}$ was created for each item i in each store S, such that $X_{S,i} = 1$ when the item was purchased in both stores and $X_{S,i} = 0$ otherwise. A weighting factor for each price was then created, so that:

$$W_{S,i} = 1 + \pi X_{S,i}$$

A purchase-weighted version of the simple frequency count model (Model 4) was created by adjusting the frequency of price advantages such that:

$$P_{A,i} < P_{B,i} \quad \rightarrow \quad F_{A,i} = W_{A,i} \; and \; F_{B,i} = 0$$

$$P_{A,i} > P_{B,i} \quad \rightarrow \quad F_{A,i} = 0 \; and \; F_{B,i} = W_{B,i}$$

$$P_{A,i} = P_{B,i} \quad \rightarrow \quad F_{A,i} = F_{B,i} = 0$$

$$F_S = \frac{1}{\sum_{i=1}^{i=150} W_{S,i}} \sum_{i=1}^{i=150} F_{S,i}$$

If the prices of purchased items are more salient than the prices of non-purchased items, and hence bias the estimated frequency of paired-item price advantages in each store, then a positive contribution $\pi > 0$ should be observed.

Purchase-weighting (PW) was applied to each of the five types of frequency comparison (A-E), and each model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.29 and the fitted parameter values are shown in Table 5.30. Purchase-weighting all five model types improves the pre-checkout rating model fits relative to the equivalent un-weighted versions of Model 4, but offers no improvement in the post-checkout rating models. The best-fitting model for pre-checkout ratings was the power law comparison (4B+PW). The log-likelihoods of all five model types were again almost identical for the post-checkout ratings, and significantly worse than the fit for the pre-checkout ratings. For the pre-checkout rating models, all five models predict an increased probability of a high rating as the frequency of price advantages in the first store increases or as the frequency in the second store decreases ($\alpha > 0$) and frequency estimates biased toward purchased items ($\pi > 0$).

TABLE 5.29
Goodness-of-fit measures for purchase-weighted versions of Model 4.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
4A+PW: Linear	-1483.8	2983.5	-1670.2	3356.4
4B+PW: Stevens	-1479.8	2977.6	-1667.6	3353.3
4C+PW: Fechner	-1481.7	2981.3	-1669.0	3355.9
4D+PW: Luce	-1482.4	2982.8	-1668.2	3354.4
4E+PW: Logistic	-1483.0	2984.0	-1670.2	3358.4

TABLE 5.30
Fitted parameter values for purchase-weighted versions of Model 4.

	Pre-Checkout				
Parameter	4A+PW	4B+PW	4C+PW	4D+PW	4E+PW
c_1	-1.883	-1.892	-1.888	-1.237	-0.785
c_2	-1.104	-1.107	-1.101	-0.455	-0.007
c_3	-0.147	-0.147	-0.144	0.502	0.951
c_4	0.587	0.589	0.589	1.236	1.686
c_5	1.534	1.540	1.537	2.185	2.635
c_6	2.579	2.593	2.592	3.236	3.680
A	0.002	0.005	0.186	1.301	2.201
B		0.503	135 540	0.748	0.0006
Π	7.141	634.1	201 120	303.6	54.22
	Post-Checkout				
Parameter	4A+PW	4B+PW	4C+PW	4D+PW	4E+PW
c_1	-1.329	-1.331	-1.327	-0.788	44.38
c_2	-0.643	-0.641	-0.638	-0.099	45.07
c_3	-0.044	-0.041	-0.040	0.500	45.67
c_4	0.402	0.405	0.405	0.946	46.12
c_5	1.230	1.235	1.235	1.776	46.94
c_6	2.083	2.091	2.091	2.632	47.80
A	0.002	0.085	0.100	1.076	91.43
B		0.341	0.336	0.432	0.0001
Π	1.380	2.521	-2.180	-1.765	1.382

In a similar fashion, it is likely that greater attention was paid to items in product departments where the participant spent more time browsing. Hence, the memory traces for the prices of such items should be stronger than those for the prices of other items. If participants make an overall price comparison by comparing the frequency with which each store is cheaper on pairs of matching items, the judgments should be biased towards items from those departments that were browsed for the longest time, due to the higher probability of recall. Browsing time could not be measured for each individual item, but only for each of the ten product departments, each of which contained fifteen items. A variable $Y_{S,i}$ was created to represent the approximate percentage of total browsing time in store *S* dedicated to each item *i* in product department *d*:

$$Y_{S,i} = \frac{t_{S,d}(i)}{15\sum_{d=1}^{d=10} t_{S,d}(i)}$$

A weighting factor for each price was then created, so that:

$$W_{S,i} = 1 + \tau Y_{S,i}$$

A time-weighted version of the simple frequency count model (Model 4) was created by adjusting the frequency of price advantages such that:

$$P_{A,i} < P_{B,i} \quad \rightarrow \quad F_{A,i} = W_{A,i} \; and \; F_{B,i} = 0$$

$$P_{A,i} > P_{B,i} \quad \rightarrow \quad F_{A,i} = 0 \; and \; F_{B,i} = W_{B,i}$$

$$P_{A,i} = P_{B,i} \quad \rightarrow \quad F_{A,i} = F_{B,i} = 0$$

$$F_S = \frac{1}{\sum_{i=1}^{i=150} W_{S,i}} \sum_{i=1}^{i=150} F_{S,i}$$

If the prices of items from departments that were browsed for a longer time are more salient, and hence bias the estimated frequency of paired-item price advantages in each store, then a positive contribution $\tau > 0$ should be observed.

Time-weighting (TW) was applied to each of the five types of frequency comparison (A-E), and each model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.31 and the fitted parameter values are shown in Table 5.32. Time-weighting did not improve the model fits relative to the equivalent un-weighted versions of Model 4 for any of the pre- or post-checkout rating models. The best-fitting model for pre-checkout ratings was the power law comparison (4B+TW). The log-likelihoods of all five types of post-checkout rating model were again significantly worse than the fit for the pre-checkout ratings. The best-fitting models for post-checkout ratings were the power law comparison (4B+TW) and the Fechner's Law comparison (4C+TW). For the pre-checkout rating models, all five models predict an increased probability of a high rating as the frequency of price advantages in the first store increases or as the frequency in the second store decreases ($\alpha > 0$) but – unlike the pairing-independent models - frequency estimates were biased toward the items that were browsed for a longer time ($\tau > 0$) in four of the five models.

TABLE 5.31
Goodness-of-fit measures for time-weighted versions of Model 4.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
4A+TW: Linear	-1488.2	2992.4	-1670.4	3356.8
4B+TW: Stevens	-1486.5	2990.9	-1667.7	3353.3
4C+TW: Fechner	-1486.7	2991.4	-1667.9	3353.7
4D+TW: Luce	-1487.5	2993.1	-1668.7	3355.4
4E+TW: Logistic	-1488.2	2994.5	-1670.6	3359.2

TABLE 5.32
Fitted parameter values for time-weighted versions of Model 4.

	Pre-Checkout				
Parameter	4A+TW	4B+TW	4C+TW	4D+TW	4E+TW
c_1	-1.874	-1.878	-1.877	-1.226	60.70
c_2	-1.098	-1.097	-1.097	-0.446	61.48
c_3	-0.148	-0.145	-0.145	0.505	62.43
c_4	0.582	0.585	0.585	1.233	63.16
c_5	1.526	1.529	1.529	2.176	64.10
c_6	2.568	2.577	2.577	3.223	65.14
α	0.001	0.003	0.245	1.299	125.2
β		0.476	23 851	0.651	0.0001
τ	770.0	79 783	321 090	-133.1	0.018
	Post-Checkout				
Parameter	4A+TW	4B+TW	4C+TW	4D+TW	4E+TW
c_1	-1.329	-1.331	-1.330	-0.792	48.58
c_2	-0.643	-0.641	-0.640	-0.103	49.27
c_3	-0.044	-0.041	-0.041	0.496	49.87
c_4	0.402	0.405	0.405	0.942	50.32
c_5	1.230	1.235	1.235	1.771	51.14
c_6	2.082	2.090	2.090	2.626	52.00
α	0.0001	0.009	0.133	1.073	99.83
β		0.331	7 628	0.477	0.0001
τ	7 801	249 070	451 380	4 190 000	-0.011

Finally, it is likely that the total cost of the basket of items purchased in each store (C_S) will also influence judgments of relative price, especially post-checkout once the total cost has been presented to the participant. Hence, in addition to the impact of item prices already discussed, the basket totals may have an additional influence on price judgments, with participants making a comparison between the basket costs in each store and then integrating this information into their overall judgment of relative price. As previously, an additive integration process was used:

$$f(\cdot) = f(\cdot)_{item\ prices} + f(\cdot)_{basket\ costs}$$

Based upon the results from the earlier pairing-independent models, a logistic basket cost comparison was selected to be combined with each type of simple frequency count model (Model 4):

$$f(\cdot)_{basket\ costs} = \gamma \left(\frac{e^{\delta C_B}}{e^{\delta C_A} + e^{\delta C_B}} \right)$$

Each basket cost comparison (BC) model was fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.33 and the fitted parameter values are shown in Table 5.34. Adding basket cost comparisons does not improve the model fits for pre-checkout ratings, but significantly improves the fit for post-checkout ratings using any of the five types of frequency comparison. The best-fitting models for post-checkout ratings were the power law comparison (4B+BC(E)) and the Fechner's Law comparison (4C+BC(E)). The post-checkout log-likelihoods of all five BC models are similar in size to the pre-checkout log-likelihoods, suggesting that the previously observed difference was largely due to participants making a basket cost comparison after the checkout. For the post-checkout rating models, all five models predict an increased probability of a high rating as the basket cost in the first store decreases or as the basket cost in the second store increases ($\gamma > 0$).

TABLE 5.33
Goodness-of-fit measures for Model 4 with basket cost comparisons.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
4A+BC(E): Linear	-1487.3	2992.6	-1498.6	3015.1
4B+BC(E): Stevens	-1485.5	2991.0	-1496.0	3012.0
4C+BC(E): Fechner	-1485.8	2991.5	-1496.2	3012.3
4D+BC(E): Luce	-1486.6	2993.3	-1496.9	3013.8
4E+BC(E): Logistic	-1487.3	2994.6	-1498.6	3017.1

TABLE 5.34
Fitted parameter values for Model 4 with basket cost comparisons.

	Pre-Checkout				
Parameter	4A+BC(E)	4B+BC(E)	4C+BC(E)	4D+BC(E)	4E+BC(E)
c_1	-1.922	-1.926	-1.925	-1.263	60.68
c_2	-1.147	-1.146	-1.145	-0.484	61.46
c_3	-0.196	-0.194	-0.193	0.467	62.41
c_4	0.535	0.537	0.538	1.197	63.14
c_5	1.482	1.484	1.484	2.142	64.08
c_6	2.527	2.536	2.535	3.193	65.13
α	0.004	0.061	0.247	1.318	125.2
β		0.479	10.80	0.653	0.0001
γ	-0.099	-0.100	-0.100	-0.098	-0.099
δ	2.139	2.095	2.097	2.101	2.137
	Post-Checkout				
Parameter	4A+BC(E)	4B+BC(E)	4C+BC(E)	4D+BC(E)	4E+BC(E)
c_1	-0.695	-0.702	-0.702	-0.198	48.55
c_2	0.134	0.130	0.131	0.634	49.38
c_3	0.872	0.870	0.870	1.373	50.12
c_4	1.427	1.425	1.425	1.927	50.67
c_5	2.431	2.431	2.431	2.932	51.67
c_6	3.430	3.433	3.432	3.933	52.67
α	0.003	0.105	0.123	1.005	98.49
β		0.315	2.430	0.456	0.0001
γ	1.913	1.902	1.902	1.901	1.913
δ	0.317	0.323	0.323	0.324	0.317

Two of the three sampling differences – purchase-weighting and basket cost comparisons – appear to offer an improved prediction when added to an un-weighted frequency count process model. All three effects were therefore combined with the best-fitting pairing-dependent representative participant model: logistic paired-item price comparisons with an additive integration process (Model 5D). In order to combine purchase-weighting and/or time-weighting with the additive integration process, a weighting factor for each item was created, so that:

$$F_i = \frac{e^{\beta \mathrm{P}_{B,i}}}{e^{\beta \mathrm{P}_{A,i}} + e^{\beta \mathrm{P}_{B,i}}}$$

$$f(\cdot) = \alpha \sum_{i=1}^{i=150} \left(\frac{W_i}{\sum_{i=1}^{i=150} W_i} F_i \right)$$

For purchase-weighting, the weighting factor used was:

$$W_i = 1 + \pi X_i$$

Where $X_i = 1$ when the item was purchased in both stores and $X_i = 0$ otherwise. For time-weighting, the weighting factor used was:

$$W_i = 1 + \frac{\tau}{2}(Y_{A,i} + Y_{B,i})$$

To combine purchase- and time-weighting, the weighting factor used was:

$$W_i = 1 + \pi X_i + \frac{\tau}{2}(Y_{A,i} + Y_{B,i})$$

As before, the basket cost comparisons were included using additive integration:

$$f(\cdot) = f_{item\ prices}(\cdot) + \gamma \left(\frac{e^{\delta \mathrm{C}_B}}{e^{\delta \mathrm{C}_A} + e^{\delta \mathrm{C}_B}} \right)$$

All seven possible permutations of sampling differences models were fit to pre- and post-checkout ratings. The goodness-of-fit measures for the models are shown in Table 5.35 and the fitted parameter values are shown in Table 5.36. All the seven sampling differences models have improved AIC scores relative to the representative participant model (5D), for both pre- and post-checkout ratings. The best-fitting model for both pre- and post-checkout ratings was the logistic pair-wise comparison model, with purchase-weighting and a basket cost comparison (5D+PW+BC(E)). Adding time-weighting did not improve the AIC of either model. The AIC of this model is slightly lower than the AIC of the best-fitting pairing-independent model for pre-checkout ratings (2916.0 vs. 2916.7) and much lower for post-checkout ratings (2941.3 vs. 2948.8). Thus, it appears that a pairing-dependent process model involving pair-wise item price comparisons is a marginally better explanation of the observed experimental data than a pairing-independent process model. Accounting for differences in the way price information is sampled significantly improves the fit of the models, as does accounting for additional information provided such as the total basket cost in each store.

TABLE 5.35
Goodness-of-fit measures for Model 5D with sampling differences.

	Pre-Checkout		Post-Checkout	
Model	Log-Likelihood	AIC	Log-Likelihood	AIC
5D+PW	-1451.8	2921.5	-1610.3	3238.6
5D+TW	-1459.4	2936.7	-1626.0	3270.1
5D+BC(E)	-1456.6	2933.3	-1465.1	2950.1
5D+PW+TW	-1451.6	2923.3	-1610.0	3239.9
5D+PW+BC(E)	-1447.0	2916.0	-1459.6	2941.3
5D+TW+BC(E)	-1456.4	2934.8	-1463.7	2949.3
5D+PW+TW+BC(E)	-1446.9	2917.7	-1458.7	2941.3

TABLE 5.36
Fitted parameter values for Model 5D with sampling differences.

	Pre-Checkout						
Parameter	5D+PW	5D+TW	5D +BC(E)	5D+PW +TW	5D+PW +BC(E)	5D+TW +BC(E)	5D+PW +TW +BC(E)
c_1	2.143	1.217	1.360	2.107	2.489	1.430	2.435
c_2	2.987	2.052	2.194	2.952	3.334	2.266	3.281
c_3	3.973	3.027	3.170	3.938	4.322	3.241	4.269
c_4	4.713	3.762	3.908	4.678	5.067	3.979	5.014
c_5	5.686	4.729	4.880	5.651	6.048	4.951	5.995
c_6	6.830	5.869	6.031	6.797	7.210	6.105	7.158
α	8.267	6.355	0.046	8.195	9.203	6.961	9.096
β	2.505	3.259	2.966	2.590	2.330	3.000	2.413
π	9.682			18.42	10.94		20.08
τ		105.8		142.1		91.17	131.2
γ			-0.180		-0.239	-0.181	-0.239
δ			1.528		1.101	1.537	1.109
	Post-Checkout						
Parameter	5D+PW	5D+TW	5D +BC(E)	5D+PW +TW	5D+PW +BC(E)	5D+TW +BC(E)	5D+PW +TW +BC(E)
c_1	7.694	3.969	4.657	7.083	5.555	4.167	5.043
c_2	8.446	4.703	5.531	7.835	6.436	5.043	5.926
c_3	9.083	5.327	6.294	8.472	7.205	5.806	6.694
c_4	9.551	5.787	6.860	8.940	7.775	6.373	7.265
c_5	10.42	6.649	7.894	9.810	8.811	7.409	8.303
c_6	11.33	7.546	8.930	10.72	9.849	8.447	9.343
α	18.29	10.74	0.073	17.06	12.82	9.945	11.79
β	0.962	1.564	1.281	1.074	1.227	1.605	1.439
π	17.47			97.06	7.583		66 003
τ		294.3		7.083		2 596	1270600
γ			1.831		1.775	1.831	1.775
δ			0.342		0.349	0.345	0.351

5.3.3.3 Interpreting Model Parameters

To understand and interpret the selected pairing-dependent model (5D+PW+BC(E)) and the fitted parameter values, I have again explored the predictions of the model at typical values of the input data as well as the sensitivity of those predictions to changes in the input variables. As before, the model was decomposed into its constituent parts (pair-wise price comparisons, purchase-weighting, and a basket cost comparison) and each part was explored independently, by holding the rest constant at the mean values observed in the experimental data.

The impact of the input prices was explored for a representative participant, with mean repeat-purchasing probability for each item, and with the mean basket costs C_A = £50.25 and C_B = £50.88. The prices in the first store were fixed as the control store, Jones. The prices in the second store were varied to test how the model predictions changed in response to those variations. Figure 5.18 shows how the probability distribution for the pre-checkout ratings varies as a function of the discount in the second store. The probability of rating the second store as cheaper (R_{AB} = 1, 2, or 3) increases from 45% when the two stores have the same mean price to 95% when the second store is 30% cheaper on all items. Figure 5.19 shows the probability distributions for the post-checkout ratings. The probability of rating the second store as cheaper increases from 42% when the two stores have the same mean price to 92% when the second store is 30% cheaper on all items. Figures 5.20 and 5.21 show the observed mean pre- and post-checkout comparative price judgment ratings from each discount condition in Experiment 3, and the expected values from the model. The post-checkout ratings are strongly influenced by the basket costs in each store and hence they are noisier and less well-explained by the price variations than the pre-checkout ratings. The shapes of the pre- and post-checkout response curves are similar, although the pre-checkout ratings exhibit a slightly diminishing response to the discount level in the large discount conditions. The post-checkout model appears to under-predict the impact of discount level on comparative price judgments. This is likely due to the impact of the discount level on the basket cost in the second store, whereas the basket costs were held constant in these figures.

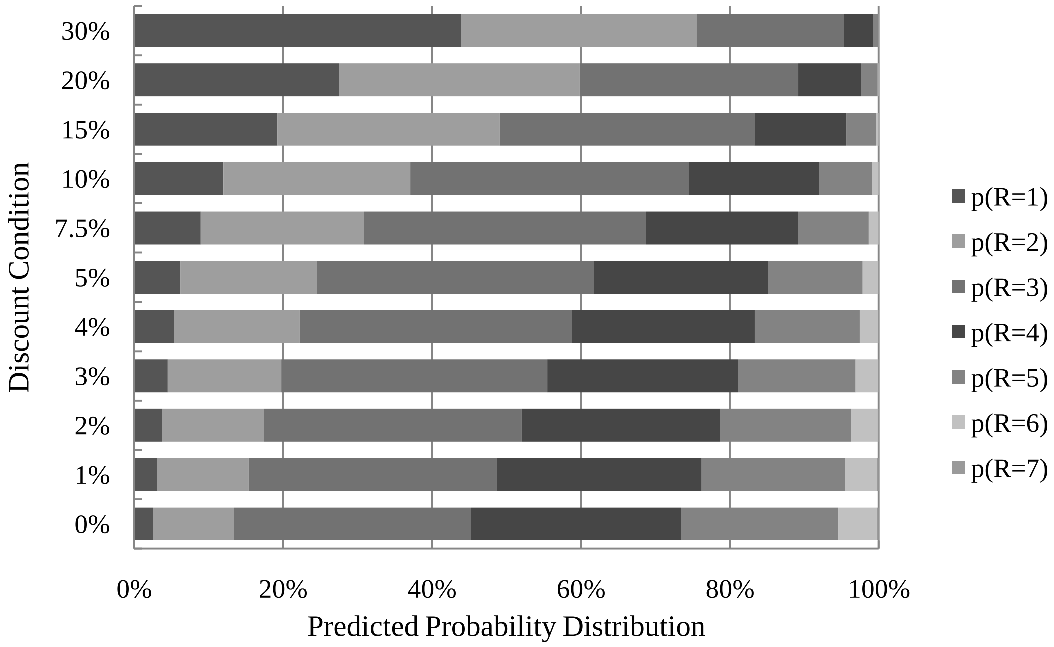

Figure 5.18: Predicted probability distribution for pre-checkout ratings as a function of the mean price difference between the two stores (Model 5D+PW+BC(E)).

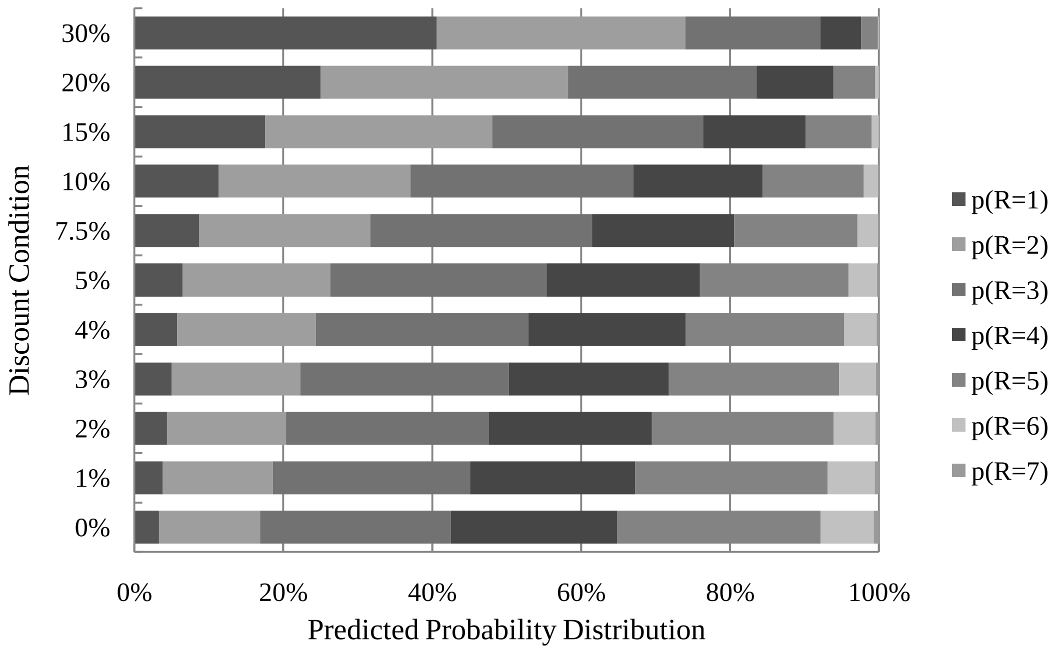

Figure 5.19: Predicted probability distribution for post-checkout ratings as a function of the mean price difference between the two stores (Model 5D+PW+BC(E)).

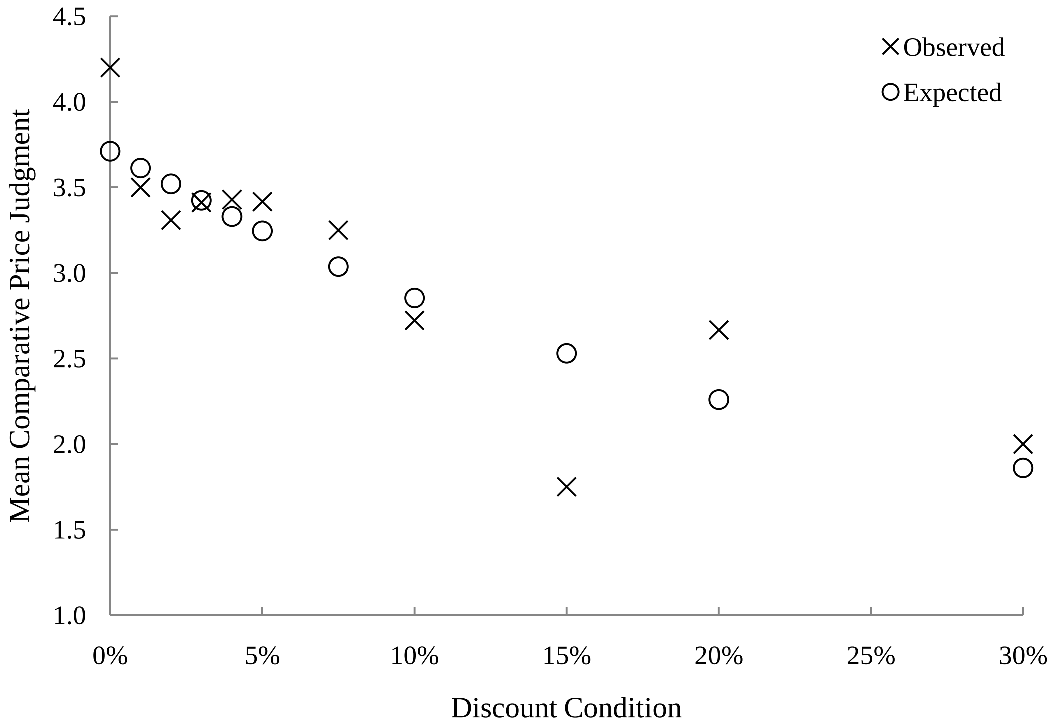

Figure 5.20: Observed and expected values of pre-checkout ratings as a function of the mean price difference between the two stores (Experiment 3 and Model 5D+PW+BC(E)).

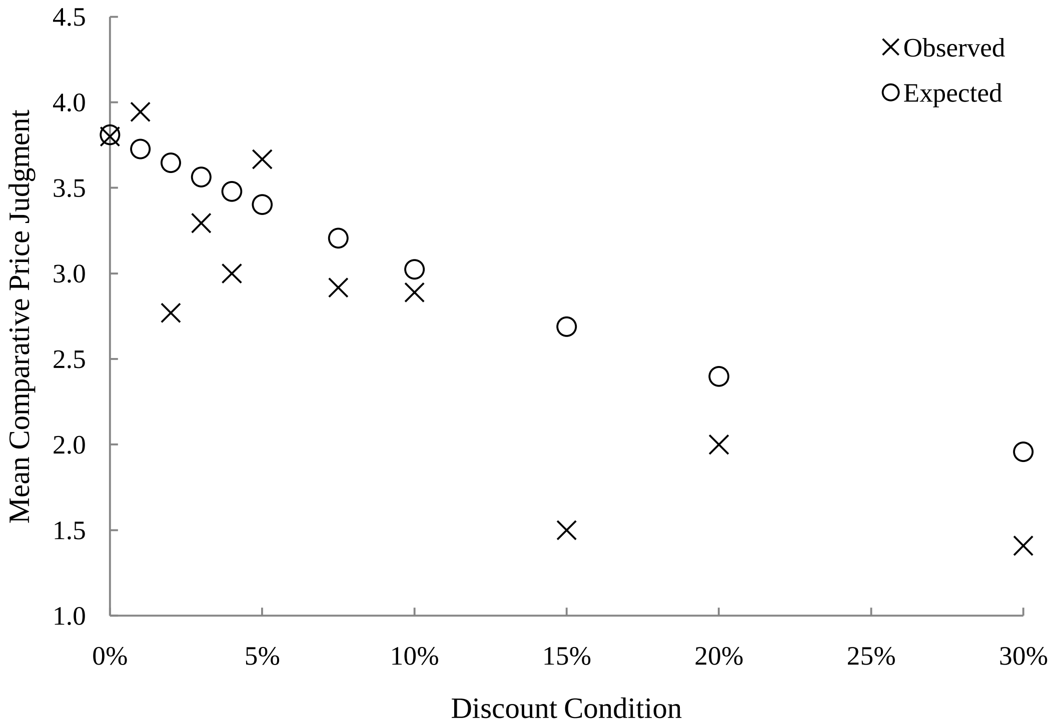

Figure 5.21: Observed and expected values of post-checkout ratings as a function of the mean price difference between the two stores (Experiment 3 and Model 5D+PW+BC(E)).

The model also predicts the pattern of comparative price judgments observed in the frequency and magnitude conditions of Experiment 4. Figure 5.22 shows how the predicted probability distributions for the pre-checkout ratings vary as a function of the frequency and magnitude of price advantages in the second store. When the magnitude of the price advantages in the second store is small (5%), the probability of rating the second store as cheaper (R_{AB} = 1, 2, or 3) increases from 46% when the second store is cheaper on 20% of the items to 49% when the second store is cheaper on 80% items. When the magnitude of the price advantages in the second store is large (20%), the probability of rating the second store as cheaper increases from 42% when the second store is cheaper on 20% of the items to 66% when the second store is cheaper on 80% items. Figure 5.23 shows the predicted probability distributions for the post-checkout ratings. The post-checkout ratings are less sensitive than the pre-checkout ratings to frequency increases: when the magnitude of the price advantages is small, the probability of rating the second store as cheaper increases from 43% to 44%; when the magnitude of the price advantages is large, the probability of rating the second store as cheaper increases from 42% to 57%. This is reflected in the pair-wise item price comparison parameter values, with the price difference sensitivity in the pre-checkout model (β = 2.33) being much greater than the price difference sensitivity in the post-checkout model (β = 1.23). Figures 5.24 and 5.25 show the observed mean pre- and post-checkout comparative price judgment ratings from each price condition in Experiment 4 and the expected values from the model. As with the earlier pairing-independent model, the post-checkout model appears to somewhat under-predict the impact of frequency on comparative price judgments.

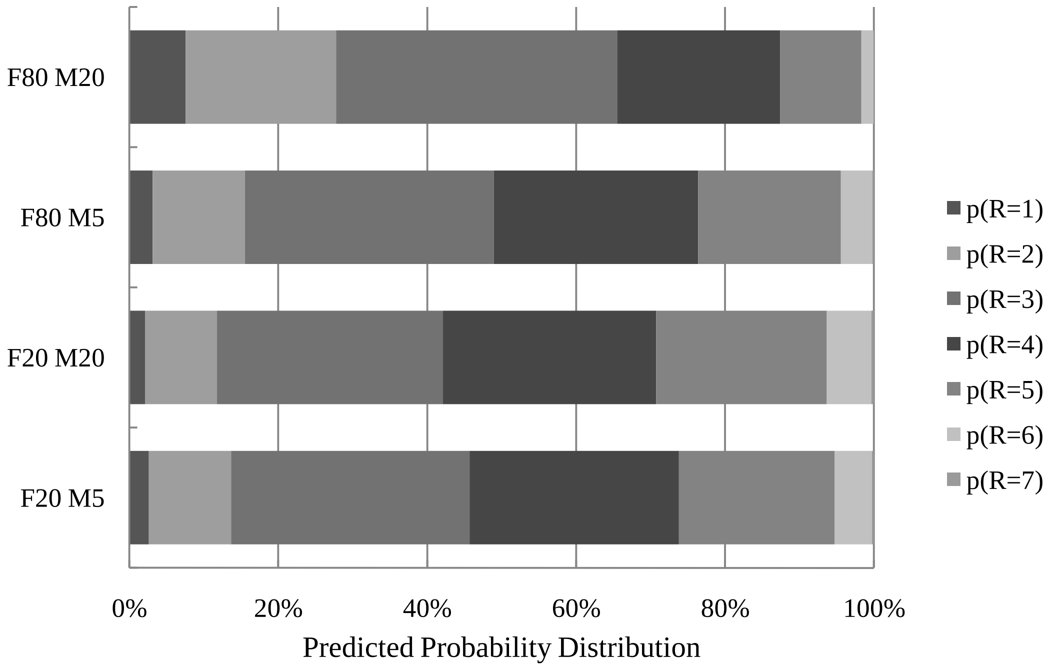

Figure 5.22: Predicted probability distribution for pre-checkout ratings as a function of the frequency and magnitude of price advantages in Store B (Model 5D+PW+BC(E)).

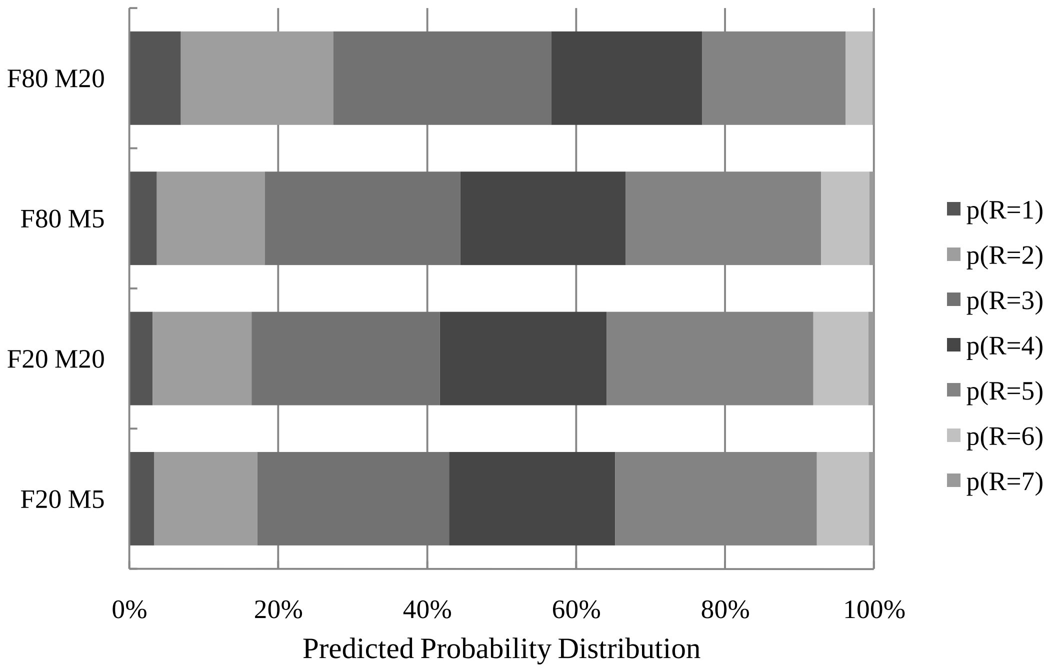

Figure 5.23: Predicted probability distribution for post-checkout ratings as a function of the frequency and magnitude of price advantages in Store B (Model 5D+PW+BC(E)).

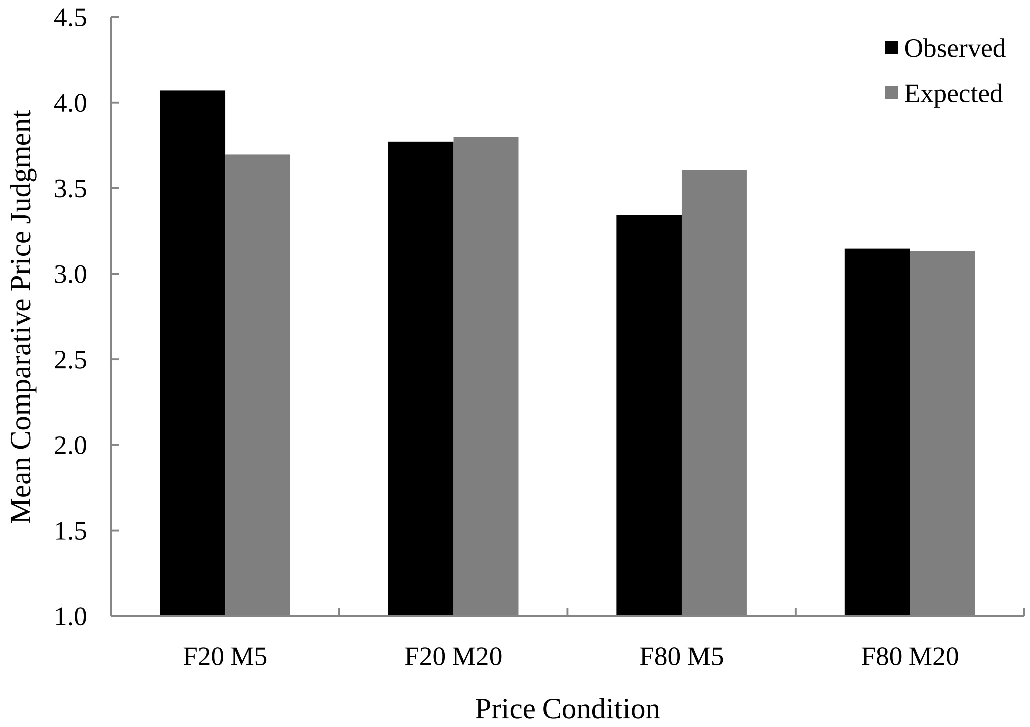

Figure 5.24: Observed and expected values of pre-checkout ratings as a function of the price distribution in each store (Experiment 4 and Model 5D+PW+BC(E)).

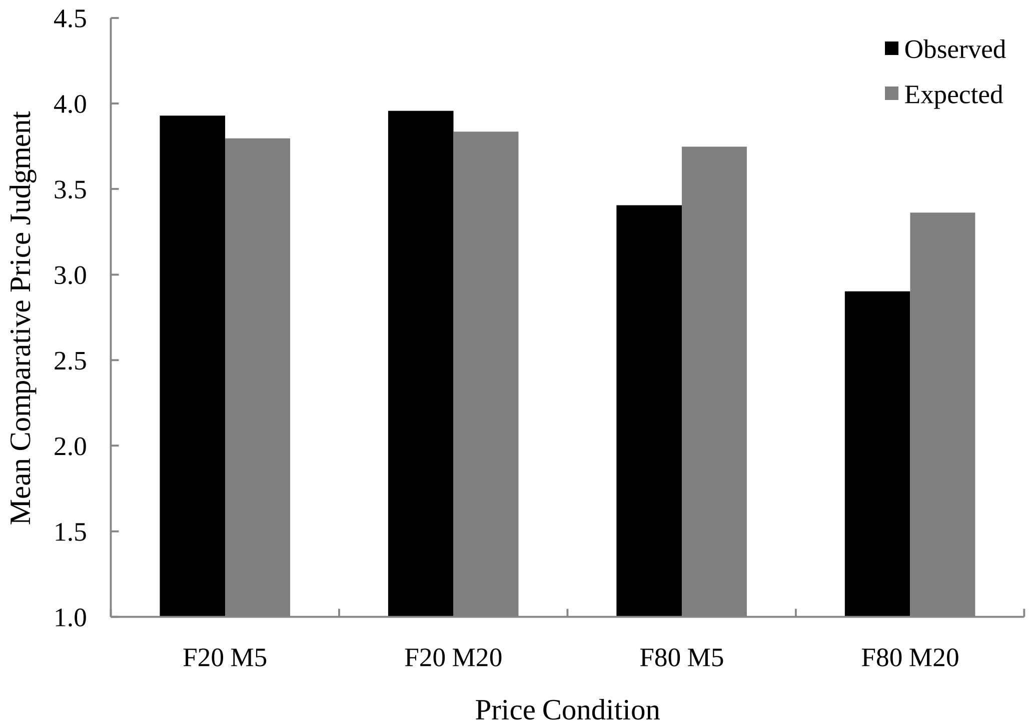

Figure 5.25: Observed and expected values of post-checkout ratings as a function of the price distribution in each store (Experiment 4 and Model 5D+PW+BC(E)).

Because the purchase weighting is applied to pair-wise item price comparisons, the effect is not separable from the impact of the price distributions. The magnitude of the impact of purchase-weighting is illustrated here by setting the item prices in the test store to 10% lower than the control store, and by holding the total basket costs C_A and C_B at the mean observed values. Figure 5.26 shows the predicted probability distributions for the pre-checkout ratings for five different patterns of purchasing behaviour in the two stores: (i) the mean observed repeat-purchase probability for each item; (ii) no items purchased in either store; (iii) all items purchased in both stores; (iv) only the 31 smallest price items purchased in both stores; and (v) only the 32 largest price items purchased in both stores. In the first three cases, the probability of rating the second store as cheaper (R_{AB} = 1, 2, or 3) is very similar, at 74%, 78% and 78% respectively. When only the 31 smallest price items are purchased in the second store, the probability of rating the second store as cheaper falls to 61%. When only the 32 largest price items are purchased, the probability rises to 97%. Figure 5.27 shows the predicted probability distributions for the post-checkout ratings, for the same five patterns of purchasing behaviour. Once again, in the first three cases, the probability of rating the second store as cheaper is very similar, at 67%, 70% and 70% respectively. The impact of purchasing only the smallest or largest price items is less extreme for the post-checkout ratings, with the probability of rating the second store as cheaper falling to 57% when only the 31 smallest price items are purchased, and rising to 92% when only the 32 largest price items are purchased. This is reflected in the parameter values, with purchased items having 12 times more impact than un-purchased items in the pre-checkout model (π = 10.9), and 9 times more impact than un-purchased items in the post-checkout model (π = 7.6).

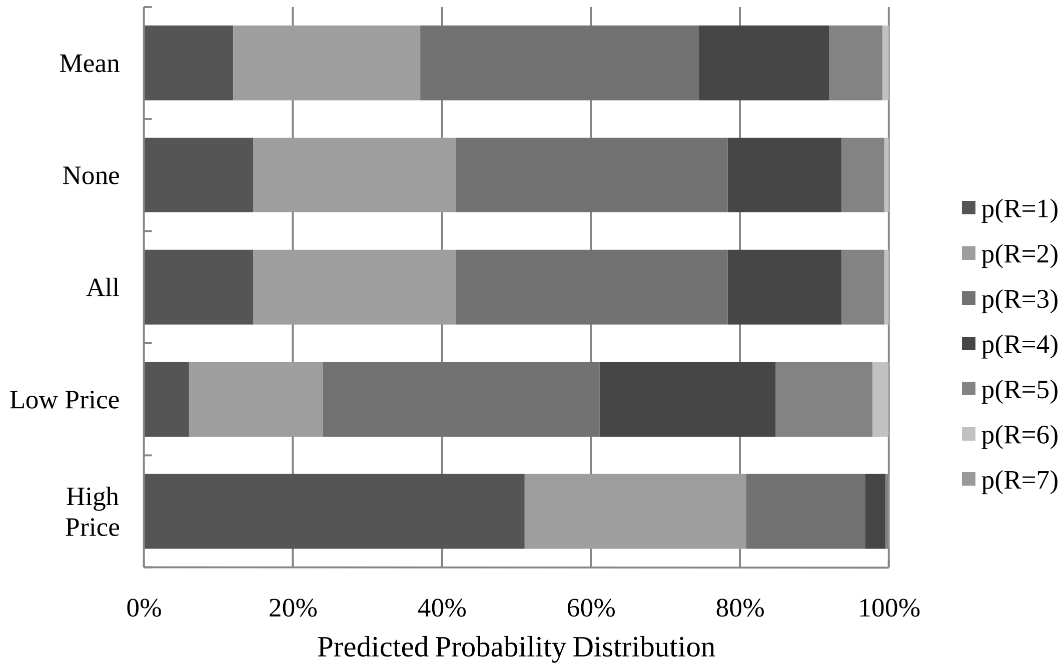

Figure 5.26: Predicted probability distribution for pre-checkout ratings as a function of the purchasing behaviour in Store B (Model 5D+PW+BC(E)).

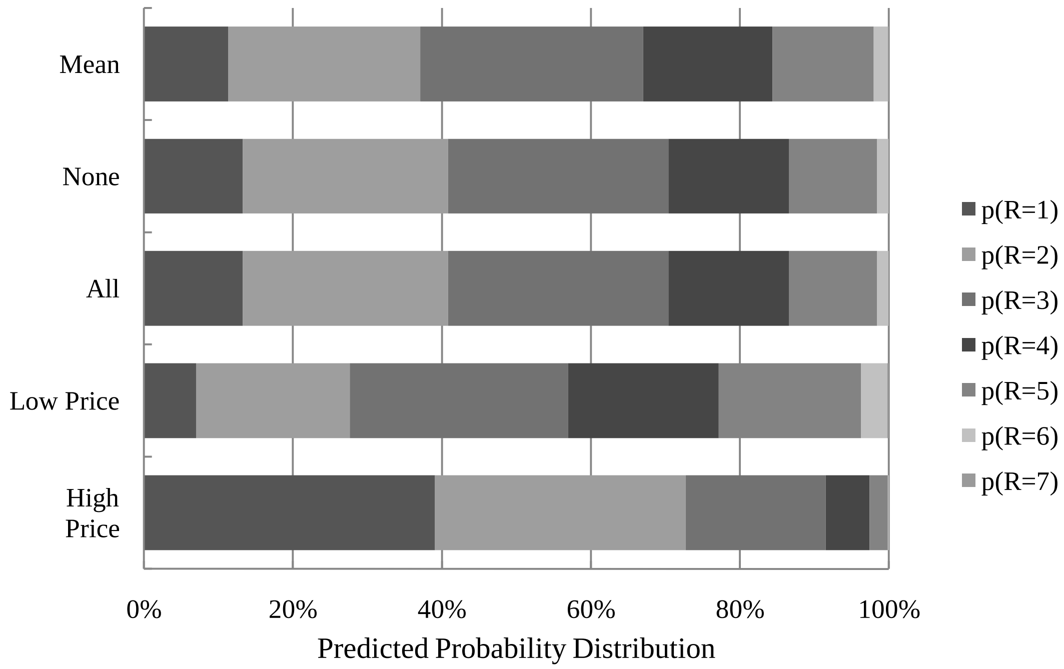

Figure 5.27: Predicted probability distribution for post-checkout ratings as a function of the purchasing behaviour in Store B (Model 5D+PW+BC(E)).

Finally, by setting the item prices in the test store to match the control store, and by holding the repeat-purchasing behaviour for each item at the mean observed values, the impact of variations in basket cost was independently tested. The total basket cost in the first store was also held constant at the mean observed value of £50.25. Figure 5.28 shows the predicted probability distributions for the pre-checkout ratings for five different levels of basket cost in the second store: (i) £30; (ii) £40; (iii) £50; (iv) £60; and (v) £70. The impact of the basket cost on pre-checkout ratings is small and in a counter-intuitive direction, with the probability of rating the second store as cheaper (R_{AB} = 1, 2, or 3) increasing from 39% when C_B = £30 to 48% when C_B = £70. One possible explanation is that participants purchased slightly more items when they perceived the second store to be cheaper, increasing the basket cost at the same time as increasing the probability of rating the second store as cheaper. Figure 5.29 shows the predicted probability distributions for the post-checkout ratings, for the same five values of basket cost C_B. The impact of the basket cost on post-checkout ratings is much greater and in the intuitive direction, with the probability of rating the second store as cheaper falling from 79% when C_B = £30 to 16% when C_B = £70. This difference is reflected in the parameter values, with a much larger scaling parameter for the basket cost comparison term in the post-checkout rating model (γ = 1.78) compared to the pre-checkout rating model (γ = -0.24). Before the checkout, comparative price judgments are dominated by the impact of the price distributions in each store ($|\alpha| / |\gamma| = 38.5$), whilst after the checkout judgments are also strongly influenced by the basket cost comparison ($|\alpha| / |\gamma| = 7.2$). When the basket costs are correlated with price differences, as in Experiment 3, the basket cost comparison will intensify the impact of price differences on comparative judgments.

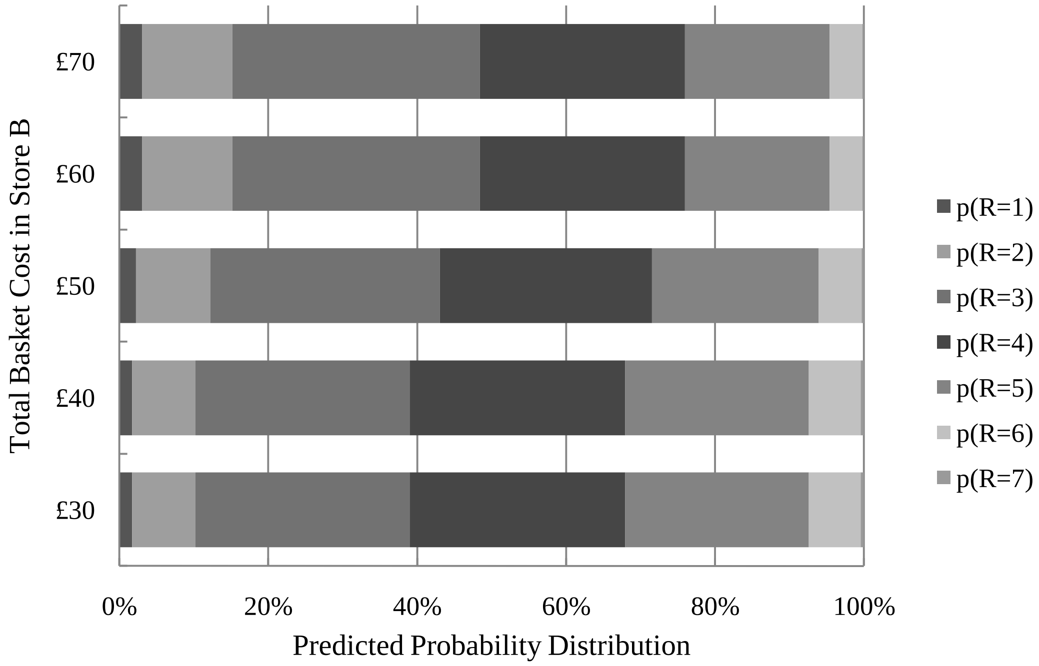

Figure 5.28: Predicted probability distribution for pre-checkout ratings as a function of the total basket cost in Store B (Model 5D+PW+BC(E)).

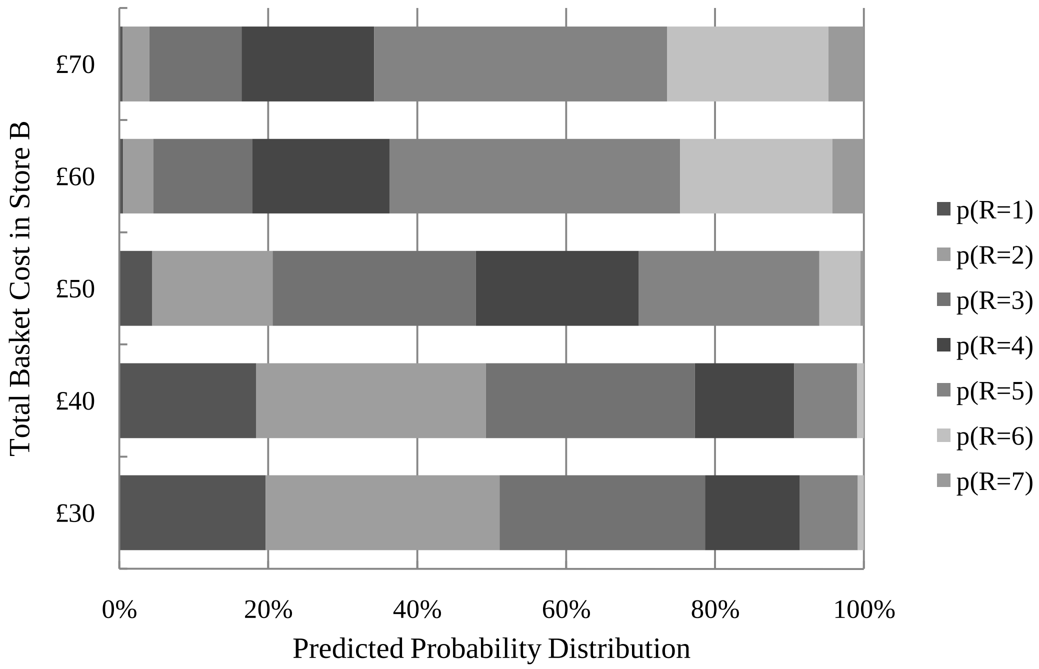

Figure 5.29: Predicted probability distribution for post-checkout ratings as a function of the total basket cost in Store B (Model 5D+PW+BC(E)).

5.4 Cognitive Modelling Conclusions

5.4.1 Overview of Models

The models compared in this chapter were categorised into families of similar models, whose relative performances can be used to draw conclusions about the cognitive process followed by participants in Experiments 3 and 4. Comparing pairing-dependent models (that assume item-level price comparisons across the two stores) with pairing-independent models (that assume the mean item price of each store is estimated independently) tells us whether participants were able to store and recall specific item prices from the first store when judging the second store. Within each family of models, comparing different functional forms of comparison process tells us what the underlying cognitive mechanism might be. Finally, comparing representative participant models (that assume all item prices are sampled and available) with sampling differences models (that assume individual participants sample a limited number of prices, determined by their browsing behaviour) tells us whether participants made an automatic judgment after viewing all the information or a deliberate but difficult judgment based on sampled and recalled information.

5.4.2 Pairing Dependence

The hypothesized explanation for the frequency effect observed in Experiments 2 and 4 was that participants' comparative price judgments are strongly influenced by the frequency with which one store is cheaper than the other on identical items. The magnitude of those price differences has a diminishing impact: initially increasing a price difference raises the likelihood of that difference being noticed, but for larger differences this effect begins to saturate and the frequency cue dominates comparative judgments. This *pairing-dependent* explanation assumes that participants are able to recall the prices of at least some items in the first store, and compare those recalled prices to the appropriate prices in the second store. The models tested in this chapter pitted this explanation against an alternative hypothesis: that participants estimated the mean price in each store independently – in other words, without matching the prices of individual items across the two stores – and then compared those estimates to make a comparative price judgment. Such *pairing-independent* explanation impose fewer demands upon participants' memory for items and prices, and the cognitive resources required for such judgments are much lower due to the smaller number of calculations involved.

The results of the model-fitting exercise in this chapter support the pairing-dependent process explanation, with the best-fitting pre- and post-checkout models both incorporating a series of pair-wise price comparisons across matching items in each store. However, the AIC advantage of this model over pairing-independent models (most notably the typicality-weighted mean price estimate model) was small, so this result is indicative rather than conclusive. Nonetheless, the frequency cue does appear to play a role in participants' comparative judgments, even if the magnitude of the price differences moderates their impact in the case of small differences. The best explanation of the observed data is one that assumes pairs of prices for matching items are compared, and the store with the highest frequency of (noticeable) advantages is judged to be cheaper.

5.4.3 Functional Form of Comparison Process

Of the five types of comparison judgment compared – linear, Steven's Power Law, Fechner's Law, Luce's Choice Rule and a logistic choice rule – no single process was the best-fitting in every case. In general, the two parameter models were better than the linear model, despite the AIC penalty for the extra parameter. In the majority of cases, Steven's Power Law is a better fit than a Fechner's Law comparison, although the results are averaged across individuals, which Steven's own work was criticized for (see Chapter 1). In general, the Luce and logistic comparisons fit better than the other three models, with the logistic model having a slight edge, but there were cases in which each type of model had the lowest AIC. The best-fitting pairing-independent and pairing-dependent models both utilized a logistic comparison process, as suggested by Buyukkurt, lending support to an S-shaped valuation function. The S-shape of the Luce and logistic models may make them inherently more flexible than the other functional forms, with the ability to describe a wider range of data given the same space of input parameters. If this were the case, then a more complex model selection criterion such as the Fisher Information Matrix (FIM) would be required to determine whether the slight advantage of the S-shaped models is due to their inherent flexibility or whether they are truly a better explanation of the observed data.

The fitted values of β in the pairing-dependent logistic comparison model (5D+PW+BC(E)) suggest a moderate degree of curvature with respect to the magnitude of paired-item price differences. Figure 5.30 shows the valuation function for a single item comparison, for an item priced at £1.95 (the mean item price) in the first store, for a range of prices in the second store. Within a limited range of price differences (about +/-20%) the item valuation function is approximately linear, but differences outside of this range have a diminishing marginal impact. The curvature of the pre-checkout comparisons is greater than that of the post-checkout comparisons, reflecting the fact that item price differences play a reduced role in comparative judgments once the total basket cost in each store is revealed.

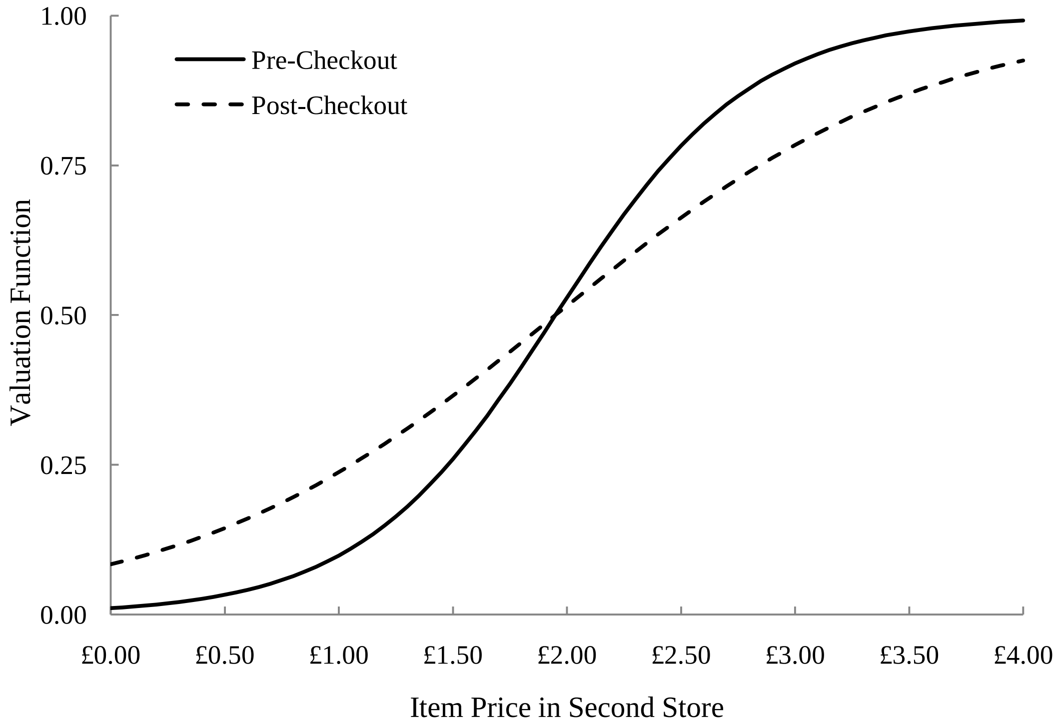

Figure 5.30: Item valuation function for an item priced £1.95 in the first store, for pre- and post-checkout logistic comparisons (Model 5D+PW+BC(E)).

5.4.4 Sampling Differences

As hypothesized, the modelling results strongly support a role for the information sampling process followed by each individual participant in determining how the comparative price judgment is made. Three types of information sampling differences were tested: the basket of items purchased in each store; the proportion of time spent browsing each department in the two stores; and the total cost of each basket revealed at the checkout. The browsing time did not prove to significantly impact upon participants' judgments, and often worked in a counter-intuitive fashion, with less attention being paid to items that had been browsed for longer. However, the other two sampling differences proved to be extremely important, and incorporating them into the process models yielded large improvements in AIC scores. Thus, it appears that the way information was sampled by each participant (i.e. the items they purchased), and the way in which that information was communicated to each participant (i.e. the total cost of their own personal basket of items) determine which prices carried most weight in the integrative comparative price judgments.

The prices of items that were purchased in both stores carry eleven times more weight than the prices of un-purchased items in the integration of pre-checkout comparative judgments. This supports the notion that participants' judgments were disproportionately influenced by prices which had received more attention, and hence were easier to recall. In turn, this also supports pairing-dependence as a plausible process explanation, as it suggests that participants stored and recalled individual item prices from the first store, rather than making a pairing-independent estimate of the average item price and only storing that estimate in memory. After the checkout, once the total basket cost in the second store was revealed, participants' judgments were strongly influenced by the total cost of the basket of purchased items in each store. In general, the tendency to use basket cost as a proxy for the average item price in a store is likely to work well, as basket cost and store prices should be positively correlated in the majority of cases. This simplifying heuristic ("rate the store with the lower basket cost as cheaper") further increases the relative importance of purchased items, as these constitute the basket of items whose cost is given at the checkout. Thus, although participants were asked to judge the overall price level in each store, their judgment was strongly biased towards the prices of purchased items, both because greater attention was paid to those prices and also because those items determined the basket costs presented.

5.4.5 Implications

If people make complex comparative judgments in the manner described by the best-fitting process model, then there are a number of implications for our ability to make intuitive statistical judgments. Firstly, these judgments will often exhibit systematic biases, such as the frequency effect described in this thesis. In this case, one set of prices is consistently judged to be lower than another, despite having the same average magnitude (Experiment 4). Systematic biases can potentially be exploited by other agents, or lead to sub-optimal decisions being made on the basis of biased judgments. Secondly, the information sampling process has a strong influence on the outcome of the judgment process, with attentional effects and memory processes both introducing potential sources of error into the judgment process. Thirdly, when simpler diagnostic information is available – in this case the total basket costs – then people exhibit a tendency to fall back on decision-making heuristics that may be fast and demand few cognitive resources, but may also be inaccurate or lead to further systematic biases. I explore these implications in more detail in the final chapter, as well as discussing possible applications of these results and potential future extensions to this research.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSIONS

6.1 Summary of Empirical Findings

The research presented in this thesis investigated a type of intuitive statistical judgment which has previously received little attention in the psychology literature: comparative judgments of the average magnitude of collections of paired stimuli, such as item prices in two stores. As described in the first chapter, one strand of the psychology literature has concentrated upon people's ability to discriminate between stimuli of different magnitudes, from the pioneering work of Weber (e.g. Gigerenzer & Murray, 1987), through the psychophysical research of Fechner and Stevens (e.g. Stevens, 1957), to Signal Detection Theory (e.g. Tanner & Swets, 1954). A related strand of research has considered people's ability to make intuitive statistical judgments (summarized by Petersen & Beach, 1967), including inference of population averages (e.g. Irwin et al., 1956) which – for numerical stimuli - appear to be calculated from a sample of observed values held in memory (Malmi & Samson, 1983). However, a series of results from the consumer research literature concerning comparative price judgments (e.g. Alba et al., 1994) suggesting a systematic bias in comparative judgments of the means of two paired distributions of prices, appears to have no direct analogue in the psychology literature.

The key distinction between this type of judgment and those previously mentioned is the coupling of individual values (in this case item prices) between the two distributions. This appears to bias judgments concerning which distribution has the lower average magnitude toward the distribution with the highest frequency of paired advantages, even when the true mean is identical or lower in the other distribution. Whilst this is consistent with a body of research suggesting that frequency information is automatically encoded and readily processed (e.g. Hasher & Zacks, 1984), subsequent consumer research literature obtained conflicting results when the distributions were temporal and item prices paired at points in time (Alba et al., 1999) or when the relative salience of frequency and magnitude information was manipulated (Lalwani & Monroe, 2005). The experiments presented in this thesis were designed to systematically explore whether the frequency effect described in Chapter 1 occurs when frequency information is not readily available (Experiments 1 and 2) and whether it persists in a realistic situation where price information is sampled incidentally during a browsing and shopping task (Experiments 3 and 4). Furthermore, by fitting and comparing cognitive process models to the experimental data, this thesis aims to elucidate how such paired-item comparative judgments are made, and how the information sampling process impacts upon those judgments. The empirical findings from each chapter are summarized below.

The experiments in Chapter 2 first replicated the original frequency effect in paired presentation of item price information (Experiment 1) and then determined the impact of switching to a pooled presentation of price information (Experiment 2). Paired presentation resembles comparative price advertising, where pairs of prices for a single item in two different stores are presented side-by-side. Pooled presentation resembles the (much more common) experience of sampling prices from two stores sequentially, meaning that prices from the first store would have to be encoded, stored and retrieved from memory in order to make paired comparisons with the second store. Experiment 1 replicated the experimental method of Alba et al (1994), using the difference between two basket cost estimates as the dependent measure of comparative price judgments. Because the experimental design pitted prior beliefs against the frequency and magnitude of paired-item price differences, it was also important to conduct a replication using familiar UK brands and a realistic set of products and prices. Experiment 1 confirmed the previous result: although basket cost estimates are influenced by prior beliefs, the dominant factor is the frequency with which one store is cheaper than the other (a so-called frequency effect). Furthermore, a subjective confidence judgment of which store was cheaper also exhibited a frequency effect, and participants were sensitive to frequency differences between the experimental conditions. Experiment 2 repeated the method of Experiment 1, but changed the presentation of the price information from *paired* to *pooled* as described above. Both the basket cost estimates and subjective confidence judgments continued to show a frequency effect under pooled presentation, despite the fact that participants were no longer sensitive to frequency differences between the experimental conditions. A meta-analysis of Experiments 1 and 2 showed that the frequency effect in basket cost estimates was significantly reduced by the switch in information presentation format (after accounting for the effect of prior beliefs), as was the frequency effect in subjective confidence judgments. Nonetheless, a weaker frequency effect persisted despite the frequency cue no longer being readily available or automatically processed.

The replication experiments suffered from a number of limitations, in particular employing a design with low ecological validity. As outlined in Chapter 1, this not only limits the external validity of any findings but also constrains our ability to conclude that observed judgments reflect genuine intuitive judgments rather than behaviour induced by the artificial demands of the experimental task. Firstly, basket cost estimates were employed as the measure of comparative price perceptions, which is inappropriate when price perceptions causally influence basket costs (through purchasing choices) as well as being influenced by them. Secondly, participants were encouraged to pay attention to every price rather than sampling price information in a more naturalistic browsing mode. Thirdly, the number of item prices used (30) was far smaller than the number of items and prices found in a typical store. Additionally, only outcome measures were collected so it was impossible to determine how the information sampling process followed during the experiment might have influenced comparative price judgments. Hence, Chapter 3 presented a novel experimental design, developed for this thesis, intended to address these flaws and to elicit comparative price judgments made in an ecologically-valid environment and task. Participants conducted a realistic shopping task in two different fictional stores, before making a comparative judgment as to which of the stores was the cheapest. The comparative judgment was made before the checkout (i.e. immediately after browsing the item prices) and repeated after the checkout once the total cost of the basket of purchased items had been presented. The experiment was completed online, allowing a large number of participants to complete the experiment in a cost-effective and efficient manner, as well as providing access to a wider and more demographically-representative subject pool than traditional convenience samples (e.g. Birnbaum, 1999).

Experiment 3 tested participants' sensitivity to price differences in this experimental shopping paradigm by varying the mean item price in the two stores. The prices in a control store were fixed across all experimental conditions, while the mean item price in a test store was varied in a between-subjects manipulation by discounting every item by the same percentage. As expected, participants' comparative price judgments were sensitive to changes in discount level, with a just-noticeable difference in mean item price of about 3%. Unlike Experiments 1 and 2, the comparative price judgments were decoupled from basket cost estimates in the second store, which appear to be have been made by anchoring and adjusting from the basket cost in the first store. The results of Experiment 3 were used in Chapter 4 to calibrate a robust test of the effects of the frequency and magnitude of inter-store price differences and to select an appropriate sample size using *a priori* power analysis. Experiment 4 used the same procedure and task as Experiment 3, but the prices in the test store were varied in a 2x2 between-subjects design in which the frequency and magnitude of price advantages were independently manipulated, whilst holding the mean item price fixed.

The previously-observed frequency effect was found in the ecologically-valid setting of Experiment 4: a store with a high number of small price advantages is perceived to be cheaper than a store with a low number of large price advantages, when the two stores have the same mean item price. Perceived price differences between the control and test store were also larger when the magnitude of price advantages (and disadvantages) was greater, but there was no interaction between the frequency and magnitude effects. Again, price judgments were decoupled from basket cost estimates. Furthermore, analysis of purchasing behaviour suggested that participants in a high-frequency store were more likely to notice when an item they had purchased in the previous store was more expensive, and exclude it from their basket. Despite noticing the more expensive prices, the strategic non-purchasers were also more likely to judge the high-frequency store to be cheaper. This supports the hypothesis that the frequency effect is driven by paired-item price differences, as those participants who demonstrated memory for individual item prices across the two stores were also more prone to a frequency bias in their comparative price judgments. Finally, in order to test this hypothesis more rigorously Chapter 5 presented the results of a model-fitting exercise in which different families of cognitive process models were fitted to the observed judgments from Experiments 3 and 4, and then compared using an information-theoretic model selection criterion, the Akaike Information Criterion (AIC).

The models tested in Chapter 5 fell into four broad categories, depending upon whether they were pairing-dependent or pairing-independent, and upon whether they assumed representative behaviour by all participants or allowed for differences between individuals in the way price information was sampled. Although the AIC differences were small, pairing-dependent models (that assume participants compare the prices of individual items matched across the two stores) explained the observed data slightly better than pairing-independent models (that assume participants estimate the mean item price in each store independently). Accounting for the way in which price information was sampled by each individual significantly improved the explanatory power of the cognitive models, especially allowing for the price of purchased items to carry more weight in the comparative judgment and assuming that post-checkout judgments were also influenced by the total basket cost seen in each store. Within each of the four categories of models, different functional forms for comparison judgments were compared. In almost every case, non-linear functions out-performed models that assumed a linear relationship between price differences and judgment ratings. Furthermore, S-shaped functions such as the logistic function or the Luce choice rule tended to have a slight advantage over the Fechner ratio and Steven's power law, although this was not a consistent finding.

The best explanation of the observed data was a cognitive judgment process in which participants make logistic pair-wise comparisons between the prices of each item in the two stores, and then integrate those comparisons in an additive fashion, placing more weight on items that they have purchased in both stores. After the checkout, participants also make a logistic comparison between the total basket costs in each store, and integrate that judgment additively with their prior judgment based on item prices. In this explanatory model, the frequency effect arises because of the non-linearity of the pair-wise item price comparisons. Although the impact (or, alternatively, the probability of being noticed) of each pair-wise price difference increases with the magnitude of the price difference, doubling the size of a price difference produces less than twice the impact. Consistent with the S-shaped valuation function described by Buyukkurt (1986) and reproduced in Chapter 1, this leads to the conclusion that a store with many small price advantages is perceived as cheaper than a store with a few large price advantages. Increased attention paid to certain item prices - caused by purchasing those items - makes it easier for participants to recall those particular prices from the first store and make the pair-wise comparisons, hence inter-store price differences on purchased items have many times more impact than price differences on un-purchased items. However, when other salient diagnostic information is also given, such as the total basket cost in each store, participants show a strong tendency to use a simple heuristic ("the store with the smaller basket cost is cheaper") in order to make the comparative judgment rather than the more cognitively-demanding approach of considering all item prices.

6.2 Conclusions and Contributions

Through the empirical results and computational modelling summarized in the previous section, this thesis has provided robust evidence for a systematic and significant bias in comparative judgments of paired distributions of stimuli. Such judgments have so far received little attention in the psychology literature, being a special case of the more general (and widely-studied) domain of multi-attribute decision-making. However, unlike the choice between two items with a range of attributes – such as when choosing a car, where one might compare the quality, speed, safety and fuel-efficiency of different options – there is no need to consider issues such as the proper relative weighting of each attribute when comparing two distributions of a single attribute. Furthermore, the judgment studied in this thesis is based on objective numerical values, not subjectively determined preferences or other abstract qualities, putting it in the domain of intuitive statistical judgments. Hence, the normatively correct response is unequivocal and any deviation from the normative response represents a genuine bias. As described in Chapter 1, prior research into intuitive statistical judgments have focused upon a single distribution of values, or comparisons between two un-paired distributions. The unique feature of the judgment task studied in this thesis is the pairing of items between the two distributions, which gives rise to frequency information that appears to be the source of the observed bias: when two paired distributions with the same mean are compared, the distribution with the higher frequency of values that are smaller than the equivalent values in the other distribution tends to be judged the smaller of the two distributions.

Such a bias would be of little interest if analogues of such judgments were not common in everyday life. However, the two tasks chosen for the experimental tests in this thesis – assessments of comparative price lists and comparison shopping in two supermarkets – are ubiquitous in today’s society. Other examples were suggested in Chapter 1, such as comparing the price of two similar items (or the same item in two stores) over a period of time in order to determine which has the lowest average price. Doubtless there are many more instances, for example comparing two people's abilities or personality by comparing their performance or behaviour on a range of specific occasions; comparing the quality of two TV channels, restaurants or holiday destinations from a sample of experiences; or choosing the best investment by studying the history of past returns on different stocks (although this judgment may also depend on temporal structure). The prevalence of complex numerical information and of comparative choice judgments in our modern environment make it important to understand under what circumstances people might make systematic and predictable errors of judgment. Numerical information is often an intrinsic feature of economic decisions – such as where and what to purchase, or how and when to invest – and so any errors of judgment could have negative welfare implications. Furthermore, predictable errors (as opposed to random errors or noise) could potentially be exploited by other more sophisticated economic agents. For example, comparative advertising of item prices or the returns from financial investments could be presented in a way that deliberately favours an inferior option.

In the case of the comparison shopping task studied in this thesis, there is clearly a lot of interest from practitioners such as store managers or government regulators in understanding how consumers make price judgments. The review in Chapter 1 described the large volume of literature in the marketing and consumer research fields that is devoted to price judgments and price perceptions. Indeed, it was this body of research that originally described the frequency effect, and began to explore the boundary conditions under which it may or may not be observed. However, these prior experiments suffered from a number of methodological shortcomings, such as the confounding of price judgments with basket cost estimates; failing to reduce the salience of the frequency cue by adopting pooled presentation of the price information; and inadequately separating the effects of the frequency and magnitude cues in the experimental designs. The research described in this thesis has for the first time demonstrated that the frequency bias occurs in a realistic setting, with a large number of prices and incidental sampling of those prices. I have also demonstrated that the frequency effect influences price perceptions even when basket cost estimates are no longer an appropriate measure, because of the problem of reverse causality (price perceptions determine purchasing behaviour, which in turn drives the basket cost). By demonstrating these effects in an ecologically-valid experimental setting, the domain in which such effects are immediately and practically relevant has widened from the narrow case of comparative advertising to the much broader case of comparison shopping.

Beyond the identification of a decision-making bias and elucidating the practical implications for consumers and retailers, the other contribution this thesis makes to the cognitive psychology literature is methodological. The methodological contribution has three aspects: a deliberate focus upon the ecological and external validity of the experimental task; the collection of process measures to supplement the outcome measures usually collected; and the use of a novel online format with large sample sizes and a representative sample of participants. The first of these methodological contributions relates to the Brunswik-ian notion of *representative design*, described in Chapter 1. Whilst controlled experimental designs were employed for this research and not observations taken from a real environment, a lot of effort was put into re-creating the key elements of the judgment context: the type of information; the format in which it was presented; the task given to participants;

the process by which information was sampled; and a naturalistic judgment task rather than an abstract and unfamiliar statistical estimate. By employing a representative design and avoiding the use of abstract tasks and judgments often found in the cognitive psychology literature, the external validity of the research is greatly increased. There is a much smaller risk that the observed bias was caused by features of the experimental design and hence we can be much more confident that the frequency bias would also occur in real-world judgments.

The second methodological contribution of this thesis is to demonstrate the importance and value of collecting process measures in experimental judgment tasks, and then employing those measures in cognitive models to determine how the judgments were made. The traditional *structural* approach in cognitive psychology is to vary aspects of the judgment task and context (the inputs) and measure the impact on the judgments made (the output). Different hypothesized judgment processes are then compared, either with traditional linear statistics such as ANOVA or through more sophisticated models and simulations. The more direct approach employed in this thesis is to measure rather than infer key aspects of the process followed by participants in sampling information and making a judgment. Incorporating these process measures into subsequent cognitive models has two advantages. Firstly, cognitive models that assume quite different underlying sampling and judgment processes can often make quite similar predictions for the final judgment. Discriminating between such models with a structural approach requires extremely clever experimental design - to test special cases in which competing models make different predictions - or very large sample sizes to enable model selection criterion to distinguish the better-performing hypothesis. In such cases, it is clearly better to test the underlying assumptions more directly with process measures. Secondly, incorporating individual-level process measures into cognitive models removes one potentially large source of variation between participants. By factoring out "noise" introduced by individual-level differences in the process followed, any remaining differences in predictive power found between competing cognitive models are more likely to be genuine. As demonstrated in Chapter 5, the resulting models allow one to separately predict the impact of changes in the process followed by an individual and the impact of structural changes to the context or task. This thesis is, of course, not the first time process measures have been collected and employed in experimental judgment tasks, but it does form part of the growing challenge to the dominant exclusively structural approach.

The final methodological contribution of this thesis is in further illustrating the benefits of using the internet as a data collection tool. As outlined in Chapter 1, the use of the internet in published studies is still limited – although growing – despite some significant advantages (e.g. Birnbaum, 1999). Most importantly for this thesis, web-based experiments allow large samples to be collected in a cost- and time-effective manner. Because data collection occurs in parallel, data from thousands of respondents can be collected in a matter of days rather than weeks or months and by a single experimenter. Furthermore, the available pool of participants is much wider than the traditional convenience samples used in many published studies, which usually consists of psychology undergraduate students completing experiments in return for course credits. The participants in Experiments 3 and 4 were likely to have far more background knowledge and experience of grocery prices to bring into the experiment than undergraduate students. Their motivation to participate and to engage with the experimental task seriously was also likely to be greater, as they willingly volunteered to participate rather than being coerced. In my opinion, these benefits alone outweigh the small loss of experimental control caused by not being able to oversee the experimental procedure in person (which in practice probably rarely occurs even in lab-based experiments) and not being able to verify the identity of participants as effectively. The former problem can be compensated for by increasing sample sizes and experimental power or by using process measures - such as time taken and behaviour during the experiment – to filter out poor quality data. The latter problem can be mitigated by the use of checks such as IP addresses and e-mail addresses, as well as recruiting participants from well-managed panels of volunteers. By following the best-practice guidelines laid out in Chapter 1, this thesis has demonstrated that web-based research allows engaging and immersive judgment tasks to be employed with sample sizes that would be impractical or impossible in a traditional lab-based setting.

6.3 Limitations and Future Directions

In addition to the theoretical and methodological contributions described above, it is also important to highlight limitations in the research, and to outline how future investigations might address these limitations and extend the findings of this thesis. The first and most obvious limitation lies in the experimental design: the choice to restrict the experimental conditions in Experiments 3 and 4 to the special case where the two stores contain an identical range of items. While this was done quite deliberately to maximize the strength of pairing between items, the cognitive process models employed in Chapter 5 struggled to differentiate between the competing hypotheses of pairing-dependence and pairing-independence. Varying the degree of pairing between the two stores, for example by substituting different items into the test store or by varying the number of items in the test store, would provide an alternative method to verify or reject the proposed pairing-dependent explanation for the frequency effect. It should be noted, however, that this would be a significant experimental undertaking because of the fact that pairing between items is unlikely to be a binary variable, but rather items are likely to be judged to be more or less similar to each other on a continuous spectrum.

Based on our understanding of memory as consisting of traces laid down by past experiences (e.g. Wickelgren & Norman, 1966), when a participant wishes to judge whether the price of an item is high or low, they will recall prices observed for items that are (i) more similar to the test item, (ii) more typical, (iii) more recent, and (iv) more salient. In Chapter 5 I made the simplifying assumption that the only price recalled when an item was judged in the second store was for the corresponding item in the first store. In reality, the participant was likely to have also recalled the prices of other similar items, both from the first store and from past experiences. The degree of similarity between any two items might depend upon the number of shared attributes, the closeness of the two items on those attributes, and the importance placed on each attribute in judging similarity or dissimilarity. As a result, once the range of items presented in each store differs, it becomes necessary to measure and control the degree of similarity between the two ranges. That in turn would necessitate either collecting information on product attributes in order to predict the perceived similarity and degree of pairing between items, or collecting direct similarity judgments for pairs of items. Without this it would be impossible to determine to what degree participants were attempting to compare item prices between the two stores or were making independent estimates of the mean item price in each store. Whilst this raises significant practical issues for the experimenter, this would be a useful future extension to the present research. In addition to testing the conclusion of this thesis - that the frequency effect arises because of item price comparisons between the two stores – it would be of particular interest to determine whether participants switch from a pairing-dependent to a pairing-independent judgment process when the similarity of the ranges in each store drops below a certain threshold.

As mentioned above, consideration of the memory recall processes inherent in the comparative judgment also highlights further limitations in the current experimental design and analytical strategy. Firstly, the typicality of each item in the store (based on all attributes and not just the price as modelled in Chapter 5) could potentially influence the relative contribution of an item to the overall comparative judgment. This item typicality could in future research be determined in a separate series of judgments, and then added to cognitive process models. Secondly, the recency with which item prices in the first store are observed could determine the likelihood that they are recalled by participants when making item-level price judgments in the second store. This recency could be manipulated experimentally or be controlled for by collecting process measures such as the order in which each product department was browsed. Thirdly, the salience of particular item prices in the first store could be influenced by factors other than whether or not the item was purchased. Again, this could be varied experimentally, for example by drawing attention to particular items through the use of different colour or size fonts, or by adding promotions to the store. Finally, the background knowledge of participants concerning grocery item prices and their prior experiences of shopping in high or low price stores could also determine how they judge the prices shown in the second store. In future studies this could be determined through appropriate questions and controlled for in the subsequent analysis.

Another significant limitation of the current research is in the type and detail of process measures collected. The web-based data collection method limited the degree to which participant behaviour could be directly observed or the degree to which participants could be questioned for insights into the judgment process they had followed. Future studies might employ lab-based techniques such as eye-tracking or videoing of participants to obtain a more detailed understanding of their behaviour, and hence the process by which information was sampled. Additionally, techniques such as think-aloud protocols (in which participants are asked to provide a running commentary of their thought processes as they complete the task) could shed further light on the cognitive judgment process. Even in future web-based studies, more detailed process measures could be collected such as tracking mouse movements during the experiment.

Finally, whilst the cognitive modelling in Chapter 5 was able to shed more light on the frequency effect than the traditional linear statistics of the prior chapters, there were limitations in the modelling approach. I have already described the simplifying assumptions made concerning the way in which items were paired between the two stores through recall from memory. This and other limitations relate primarily to a lack of sufficient data to be able to accurately model the cognitive judgment process, especially the content of the traces which make up the memory of the first store. However, there are other limitations in the modelling approach that could be overcome in future through more sophisticated analysis of existing data. For example, it was assumed that all participants followed the same judgment process and the models were compared against each other on that basis. It might in fact be the case that the sample of participants was heterogeneous and contained people making the judgment using two or more different processes. A model that incorporated a mixture of different processes might well perform better than a single process. If that were the case, then it would be the task of future research to determine what factors influence the judgment process adopted by any specific individual. For example, do more experienced shoppers use a different judgment process from naïve shoppers?

The experimental tasks used in this thesis were chosen to match the domain of comparative price judgments in which the frequency effect was first identified (e.g. Alba et al., 1994). One interesting avenue for future research would be to test for the existence of a frequency effect in analogous judgments in other tasks and domains. As already mentioned, there are analogous price tasks in which prices are paired in alternative ways, such as at points in time. Previous studies have considered the task of comparing the average price of two items, sampled over a period of time (e.g. Alba et al., 1999). In future, other numerical tasks could be used to look for frequency effects, such as comparing the returns of different investments; judging the larger of two companies from staff numbers in different regions of the world; or any other task in which numerical values are (superfluously) paired in order to create a frequency cue. If the frequency effect is found to bias judgments across a range of intuitive statistical judgments, then an important avenue of future research would be to test for a frequency effect in other judgment domains. This work could have important implications for a wide range of learning and judgment tasks, ranging from forecasting to person perception.

Prior research described in Chapter 1 appears to show that the visual field represents statistical properties of sets of items, rather than the detail of each individual member (e.g. Ariely, 2001). If this is the case, then a frequency effect would not occur in the visual domain. However, the tasks employed in the prior research parallel the un-paired distribution tasks employed in prior experiments on intuitive statistical judgments. It would be interesting to recreate an analogous task to Experiment 3 in the visual domain by presenting two paired distributions of items on a screen. One possibility would be to pair the items in the two distributions through one physical property unique to each item (e.g. shape or colour) and ask participants to make comparative judgments of the average value of a second physical property that varies across the two distributions (e.g. area or brightness). Further manipulations might parallel other parts of this thesis, such as comparing paired and pooled presentation formats as was done in Experiments 1 and 2. Similarly, one could devise tests of auditory judgments of two sets of sounds, being received by each ear. The sounds could be paired by playing them at the same point in time, or by using an auditory property unique to each item such as pitch. The two sets of sounds could then be compared on their perceived average value of another auditory property such as loudness. Testing for the existence of a frequency effect in sensory judgments through psychophysical research is perhaps the most exciting avenue for future research. Comparing intuitive judgments of numerical information with those of sensory stimuli could help us understand whether all judgment processes have a common underlying basis or whether humans have specialized and domain-specific judgment processes that operate independently. This would have far-reaching implications, extending the importance of the frequency effect far beyond the limited context explored in this thesis, so we may reasonably hope that the challenge is soon taken up.

REFERENCES

Adaval, R., & Monroe, K. B. (2002). Automatic Construction and Use of Contextual Information for Product and Price Evaluations. *Journal of Consumer Research*, *28*(4), 572-588.

Alba, J. W., Broniarczyk, S. M., Shimp, T. A., & Urbany, J. E. (1994). The Influence of Prior Beliefs, Frequency Cues, and Magnitude Cues on Consumers' Perceptions of Comparative Price Data. *Journal of Consumer Research, 21*(2), 219-235.

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge Calibration: What Consumers Know and What They Think They Know. *Journal of Consumer Research, 27*(2), 123-156.

Alba, J. W., Mela, C. F., Shimp, T. A., & Urbany, J. E. (1999). The Effect of Discount Frequency and Depth on Consumer Price Judgments. *Journal of Consumer Research*, *26*(2), 99-114.

Alford, B. L., & Biswas, A. (2002). The Effects of Discount Level, Price Consciousness and Sale Proneness on Consumers' Price Perception and Behavioral Intention. *Journal of Business Research, 55*(9), 775-783.

Alford, B. L., & Engelland, B. T. (2000). Advertised Reference Price Effects on Consumer Price Estimates, Value Perception, and Search Intention. *Journal of Business Research*, *48*(2), 93-100.

Anderson, J. R., & Milson, R. (1989). Human Memory: An Adaptive Perspective. *Psychological Review, 96*(4), 703-719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the Environment in Memory. *Psychological Science, 2*(6), 396-408.

Anderson, N. H. (1965). Averaging Versus Adding as a Stimulus Combination Rule in Impression Formation. *Journal of Experimental Psychology, 70*(4), 394-400.

Anderson, N. H. (1970). Functional Measurement and Psychophysical Judgment. *Psychological Review, 77*(3), 153-170.

Anderson, N. H. (1971). Integration Theory and Attitude Change. *Psychological Review, 78*(3), 171-206.

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science, 12*(2), 157-162.

Arnold, S. J., Oum, T. H., & Tigert, D. J. (1983). Determinant Attributes in Retail Patronage: Seasonal, Temporal, Regional, and International Comparisons. [Article]. *Journal of Marketing Research, 20*(2), 149-157.

Ashcraft, M. H. (1992). Cognitive Arithmetic - A Review of Data and Theory. *Cognition, 44*(1-2), 75-106.

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Nat Rev Neurosci, 4*(10), 829-839.

Barron, G., & Erev, I. (2003). Small Feedback-Based Decisions and Their Limited Correspondence to Description-Based Decisions. *Journal of Behavioral Decision Making, 16*(3), 215-233.

Bell, D. R., & Lattin, J. M. (1998). Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket" Shoppers Prefer EDLP. [Article]. *Marketing Science, 17*(1), 66-88.

Birnbaum, M. H. (1999). Testing Critical Properties of Decision Making on the Internet. *Psychological Science, 10*(5), 399-407.

Birnbaum, M. H. (2004). Human Research and Data Collection via the Internet. *Annual Review of Psychology, 55*(1), 803-832.

Birnbaum, M. H., & Wakcher, S. V. (2002). Web-Based Experiments Controlled by JavaScript: An Example from Probability Learning. *Behavior Research Methods, Instruments, & Computers, 34*(2), 189-199.

Biswas, A., & Blair, E. A. (1991). Contextual Effects of Reference Prices in Retail Advertisements. *Journal of Marketing, 55*(3), 1-12.

Biswas, A., Pullig, C., Krishnan, B. C., & Burton, S. (1999). Consumer Evaluation of Reference Price Advertisements: Effects of Other Brands' Prices and Semantic Cues. *Journal of Public Policy & Marketing, 18*(1), 52-65.

Biswas, A., Wilson, E. J., & Licata, J. W. (1993). Reference Pricing Studies in Marketing - A Synthesis of Research Results. *Journal of Business Research, 27*(3), 239-256.

Blair, E. A., Harris, J., & Monroe, K. B. (2002). Effects of Shopping Information on Consumers' Responses to Comparative Price Claims. *Journal of Retailing, 78*(3), 175-181.

Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, Counting, and Statistical Inference: The Role of Frequency and Whole-Object Representations in Judgment Under Uncertainty. [Article]. *Journal of Experimental Psychology: General, 127*(1), 3-21.

Brezsnyak, M. (1999). Presenting Psychology Research on the Internet: Obstacles and Opportunities. *Behaviour Research Methods, Instruments & Computers, 31*(2), 281-286.

Briesch, R. A., Krishnamurthi, L., Mazumdar, T., & Raj, S. P. (1997). A Comparative Analysis of Reference Price Models. *Journal of Consumer Research, 24*(2), 202-214.

Broniarczyk, S. M., & Alba, J. W. (1994). Theory versus Data in Prediction and Correlation Tasks. *Organizational Behavior and Human Decision Processes, 57*(1), 117-139.

Brown, F. E. (1969). Price Image versus Price Reality. *Journal of Marketing Research, 6*(2), 185-191.

Brown, F. E. (1971). Who Perceives Supermarket Prices Most Validly? *Journal of Marketing Research, 8*(1), 110-113.

Brunswik, E. (1937). Psychology as a Science of Objective Relations. *Philosophy of Science, 4*(2), 227-260.

Brunswik, E. (1955). Representative Design and Probabilistic Theory in a Functional Psychology. *Psychological Review, 62*(3), 193-217.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for Psychological Research: Personality Testing on the World Wide Web. *British Journal of Psychology, 90*(1), 125-144.

Burke, R. R., Harlam, B. A., Kahn, B. E., & Lodish, L. M. (1992). Comparing Dynamic Consumer Choice in Real and Computer-simulated Environments. *Journal of Consumer Research, 19*(1), 71-82.

Buyukkurt, B. K. (1986). Integration of Serially Sampled Price Information - Modeling and Some Findings. *Journal of Consumer Research, 13*(3), 357-373.

Buyukkurt, B. K., & Buyukkurt, M. D. (1986). Perceived Correlates Of Store Price Image: An Application Of The Bootstrap, *Advances in Consumer Research* (Vol. 13, pp. 42): Association for Consumer Research.

Chong, S. C., & Treisman, A. (2003). Representation of Statistical Properties. *Vision Research, 43*(4), 393-404.

Chong, S. C., & Treisman, A. (2005). Statistical Processing: Computing the Average Size in Perceptual Groups. *Vision Research, 45*(7), 891-900.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ.: Lawrence Erlbaum Associates.

Compeau, L. D., & Grewal, D. (1998). Comparative Price Advertising: An Integrative Review. *Journal of Public Policy & Marketing, 17*(2), 257-273.

Compeau, L. D., Grewal, D., & Chandrashekaran, R. (2002). Comparative Price Advertising: Believe it or Not. *Journal of Consumer Affairs, 36*(2), 284-294.

Connover, J. N. (1986). The Accuracy of Price Knowledge: Issues in Research Methodology. [Article]. *Advances in Consumer Research, 13*(1), 589-593.

Cosmides, L., & Tooby, J. (1996). Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment under Uncertainty. *Cognition, 58*(1), 1-73.

Cowan, N. (1988). Evolving Conceptions of Memory Storage, Selective Attention, and their Mutual Constraints within the Human Information-Processing System. *Psychological Bulletin, 104*(2), 163-191.

Cox, A. D., & Cox, D. (1990). Competing on Price - the Role of Retail Price Advertisements in Shaping Store-Price Image. *Journal of Retailing, 66*(4), 428-445.

Craig, C. S., Ghosh, A., & McLafferty, S. (1984). Models of the Retail Location Process: A Review. *Journal of Retailing, 60*(1), 5-36.

Dawes, R. (1979). The Robust Beauty of Improper Linear Models in Decision Making. *American Psychologist, 34*, 571-582.

Dellabitta, A. J., Monroe, K. B., & McGinnis, J. M. (1981). Consumer Perceptions of Comparative Price Advertisements. *Journal of Marketing Research, 18*(4), 416-427.

Desai, K. K., & Talukdar, D. (2003). Relationship Between Product Groups' Price Perceptions, Shopper's Basket Size and Grocery Store's Overall Store Price Image. *Psychology & Marketing, 20*(10), 903-933.

Dickson, P. R., & Sawyer, A. G. (1986). Methods to Research Shoppers' Knowledge of Supermarket Prices. *Advances in Consumer Research, 13*(1), 584-588.

Dickson, P. R., & Sawyer, A. G. (1990). The Price Knowledge and Search of Supermarket Shoppers. *Journal of Marketing, 54*(3), 42-53.

Ellickson, P. B., & Misra, S. (2008). Supermarket Pricing Strategies. *Marketing Science, 27*(5), 811-828.

Epstein, S. (1994). Integration of the Cognitive and the Psychodynamic Unconscious. *American Psychologist, 49*(8), 709-724.

Estelami, H., & Lehmann, D. R. (2001). The Impact of Research Design on Consumer Price Recall Accuracy: An Integrative Review. *Journal of the Academy of Marketing Science, 29*(1), 36-49.

Estes, W. K. (1976). The Cognitive Side of Probability Learning. *Psychological Review, 83*(1), 37-64.

Estes, W. K., Campbell, J. A., Hatsopoulos, N., & Hurwitz, J. B. (1989). Base-Rate Effects in Category Learning: A Comparison of Parallel Network and Memory Storage-Retrieval Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 556-571.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A Flexible Statistical Power Analysis for the Social, Behavioral, and Biomedical Sciences. *Behaviour Research Methods, 39*(2), 175-191.

Fiser, J., & Aslin, R. N. (2002). Statistical Learning of Higher-Order Temporal Structure From Visual Shape Sequences. *Journal of Experimental Psychology-Learning Memory and Cognition, 28*(3), 458-467.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual Learning: Differentiation or Enrichment? *Psychological Review*, *62*(1), 32-41.

Gigerenzer, G., & Hoffrage, U. (1995). How to Improve Bayesian reasoning Without Instruction: Frequency Formats. *Psychological Review, 102*(4), 684-704.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as Intuitive Statistics*. Hillsdale, New Jersey: Laurence Erlbaum Associates.

Gillund, G., & Shiffrin, R. M. (1984). A Retrieval Model for Both Recognition and Recall. *Psychological Review, 91*(1), 1-67.

Grewal, D., & Compeau, L. D. (1992). Comparative Price Advertising: Informative or Deceptive? *Journal of Public Policy & Marketing, 11*(1), 52-62.

Grewal, D., Monroe, K. B., & Krishnan, R. (1998). The Effects of Price-Comparison Advertising on Buyers' Perceptions of Acquisition Value, Transaction Value and Behavioral Intentions. *Journal of Marketing*, *62*(2), 46-59.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science, 17*(9), 767-773.

Hasher, L., & Chromiak, W. (1977). The Processing of Frequency Information: An Automatic Mechanism? *Journal of Verbal Learning and Verbal Behavior, 16*, 173-184.

Hasher, L., & Zacks, R. T. (1979). Automatic and Effortful Processes in Memory. *Journal of Experimental Psychology-General, 108*(3), 356-388.

Hasher, L., & Zacks, R. T. (1984). Automatic Processing of Fundamental Information: The Case of Frequency of Occurrence. *American Psychologist, 39*(12), 1372-1388.

Heath, T. B., Chatterjee, S., & France, K. R. (1995). Mental Accounting and Changes in Price - the Frame Dependence of Reference Dependence. *Journal of Consumer Research, 22*(1), 90-97.

Herr, P. M. (1989). Priming Price - Prior Knowledge and Context Effects. *Journal of Consumer Research, 16*(1), 67-75.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions From Experience and the Effect of Rare Events in Risky Choice. *Psychological Science, 15*(8), 534-539.

Hertwig, R., Pachur, T., & Kurzenhäuser, S. (2005). Judgments of Risk Frequencies: Tests of Possible Cognitive Mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 621-642.

Hewson, C. M., Laurent, D., & Vogel, C. M. (1996). Proper Methodologies for Psychological and Sociological Studies Conducted via the Internet. *Behaviour Research Methods, Instruments & Computers, 28*(2), 186-191.

Hintzman, D. L. (1988). Judgments of Frequency and Recognition Memory in a Multiple-Trace Memory Model. *Psychological Review, 95*(4), 528-551.

Hong, W. Y., Thong, J. Y. L., & Tam, K. Y. (2004). The Effects of Information Format and Shopping Task on Consumers' Online Shopping Behavior: A Cognitive Fit Perspective. *Journal of Management Information Systems, 21*(3), 149-184.

Howell, W. C. (1973). Representation of Frequency in Memory. *Psychological Bulletin, 80*(1), 44-53.

Hutchinson, J. W., & Alba, J. W. (1997). Heuristics and Biases in the `Eyeballing' of Data: The Effects of Context on Intuitive Correlation Assessment. *Journal of Experimental Psychology: Learning, Memory & Cognition, 23*(3), 591.

Irwin, F. W., Smith, W. A. S., & Mayfield, J. F. (1956). Tests of Two Theories of Decision in an "Expanded Judgment" Situation. *Journal of Experimental Psychology, 51*(4), 261-268.

Jacoby, L. L. (1991). A Process Dissociation Framework: Separating Automatic from Intentional Uses of Memory. *Journal of Memory and Language, 30*(5), 513-541.

Jacoby, L. L. (1996). Dissociating Automatic and Consciously Controlled Effects of Study/Test Compatibility. *Journal of Memory and Language, 35*(1), 32-52.

Janiszewski, C., & Lichtenstein, D. R. (1999). A Range Theory Account of Price Perception. *Journal of Consumer Research, 25*(4), 353-368.

Josephs, R. A., Giesler, R. B., & Silvera, D. H. (1994). Judgment by Quantity. *Journal of Experimental Psychology - General, 123*(1), 21-32.

Juliusson, A., Gamble, A., & Garling, T. (2005). Learning the Value of a New Currency from Prices. *Journal of Experimental Psychology - Applied, 11*(1), 45-52.

Juslin, P., Winman, A., & Hansson, P. (2007). The Naïve Intuitive Statistician: A Naïve Sampling Model of Intuitive Confidence Intervals. *Psychological Review, 114*(3), 678-703.

Kahneman, D. (2003). A Perspective on Judgment and Choice - Mapping Bounded Rationality. *American Psychologist*, *58*(9), 697-720.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a Constraint on Profit-Seeking: Entitlements in the Market. *The American Economic Review, 76*(4), 728-741.

Kahneman, D., & Miller, D. T. (1986). Norm Theory: Comparing Reality to its Alternatives. *Psychological Review*, *93*(2), 136-153.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291.

Kalyanaram, G., & Winer, R. S. (1995). Empirical Generalizations from Reference Price Research. *Marketing Science, 14*(3), G161-G169.

Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the Misperception of Variability. *Journal of Experimental Psychology: General*, *131*(2), 287-297.

Kelley, H. H. (1973). The Processes of Causal Attribution. *American Psychologist, 28*(2), 107-128.

Kelley, H. H., & Michela, J. L. (1980). Attribution Theory and Research. *Annual Review of Psychology, 31*(1), 457-501.

Kemp, S., & Willetts, K. (1996). Remembering the Price of Wool. *Journal of Economic Psychology, 17*(1), 115-125.

Klein, N. M., & Oglethorpe, J. E. (1987). Cognitive Reference Points in Consumer Decision-Making. *Advances in Consumer Research, 14*, 183-187.

Kopalle, P. K., & Lindsey-Mullikin, J. (2003). The Impact of External Reference Price on Consumer Price Expectations. *Journal of Retailing, 79*(4), 225-236.

Lal, R., & Rao, R. (1997). Supermarket Competition: The Case of Every Day Low Pricing. [Article]. *Marketing Science, 16*(1), 60-80.

Lalwani, Ashok K., & Monroe, K. B. (2005). A Reexamination of Frequency-Depth Effects in Consumer Price Judgments. *Journal of Consumer Research, 32*(3), 480-485.

Lamberts, K. (1994). Flexible Tuning of Similarity in Exemplar-Based Categorization. *Journal of Experimental Psychology-Learning Memory and Cognition, 20*(5), 1003-1021.

Lattin, J. M., & Bucklin, R. E. (1989). Reference Effects of Price and Promotion on Brand Choice Behavior. *Journal of Marketing Research, 26*(3), 299-310.

Lendanowsky, S., Griffiths, T. L., & Kalish, M. L. (in press). The Wisdom of Individuals: Exploring People's Knowledge about Everyday Events using Iterated Learning. *Cognitive Science*.

Lichtenstein, S., Slovic, P., Fishhoff, B., Layman, M., & Combs, B. (1978). Judged Frequency of Lethal Events. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 551-578.

Longo, M. R., & Lourenco, S. F. (2007). Spatial Attention and the Mental Number Line: Evidence for Characteristic Biases and Compression. *Neuropsychologia, 45*(7), 1400-1407.

Luce, R. D. (2002). A Psychophysical Theory of Intensity Proportions, Joint Presentations, and Matches. *Psychological Review, 109*(3), 520-532.

Magi, A. W., & Julander, C. R. (2005). Consumers' Store-Level Price Knowledge: Why Are Some Consumers More Knowledgeable Than Others? *Journal of Retailing, 81*(4), 319-329.

Malmi, R. A., & Samson, D. J. (1983). Intuitive Averaging of Categorized Numerical Stimuli. *Journal of Verbal Learning and Verbal Behavior, 22*(5), 547-559.

March, J. G. (1996). Learning to be Risk Averse. *Psychological Review, 103*(2), 309-319.

Mazumdar, T., & Monroe, K. B. (1990). The Effects of Buyers' Intentions to Learn Price Information on Price Encoding. *Journal of Retailing, 66*(1), 15-32.

Mazumdar, T., & Monroe, K. B. (1992). Effects of Inter-Store and In-Store Price Comparisons on Price Recall Accuracy and Confidence. *Journal of Retailing, 68*(1), 66-89.

Mazumdar, T., Raj, S. P., & Sinha, I. (2005). Reference Price Research: Review and Propositions. *Journal of Marketing, 69*(4), 84-102.

McCloskey, M., & Macaruso, P. (1995). Representing and Using Numerical Information. *American Psychologist, 50*(5), 351-363.

McGoldrick, P. J., Betts, E. J., & Wilson, A. F. (1999). Modelling Consumer Price Cognition: Evidence from Discount and Superstore Sectors. *Service Industries Journal, 19*(1), 171-193.

McGoldrick, P. J., & Marks, H. J. (1987). Shoppers Awareness of Retail Grocery Prices. *European Journal of Marketing, 21*(3), 63-76.

Monroe, K. B. (1971a). Measuring Price Thresholds by Psychophysics and Latitudes of Acceptance. *Journal of Marketing Research, 8*(4), 460-464.

Monroe, K. B. (1971b). Psychophysics of Prices: A Reappraisal. *Journal of Marketing Research, 8*(2), 248-250.

Monroe, K. B. (1973). Buyers' Subjective Perceptions of Price. *Journal of Marketing Research, 10*(1), 70-80.

Monroe, K. B., Dellabitta, A. J., & Downey, S. L. (1977). Contextual Influences on Subjective Price Perceptions. *Journal of Business Research, 5*(4), 277-291.

Monroe, K. B., & Lee, A. Y. (1999). Remembering Versus Knowing: Issues in Buyers' Processing of Price Information. *Journal of the Academy of Marketing Science, 27*(2), 207-225.

Monroe, K. B., Powell, C. P., & Choudhury, P. K. (1986). Recall Versus Recognition as a Measure of Price Awareness. *Advances in Consumer Research, 13,* 594-599.

Murdock, B. B. (1982). A Theory for the Storage and Retrieval of Item and Associative Information. *Psychological Review, 89*(6), 609-626.

Neisser, U. (1991). A Case of Misplaced Nostalgia. *American Psychologist, 46*(1), 34-36.

Niedrich, R. W., Sharma, S., & Wedell, D. H. (2001). Reference Price and Price Perceptions: A Comparison of Alternative Models. *Journal of Consumer Research, 28*(3), 339-354.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-Based Accounts of "Multiple-System" Phenomena in Perceptual Categorization. *Psychonomic Bulletin & Review, 7*(3), 375-402.

Ofir, C., Raghubir, P., Brosh, G., Monroe, K. B., & Heiman, A. (2008). Memory-Based Store Price Judgments: The Role of Knowledge and Shopping Experience. *Journal of Retailing, 84*(4), 414-423.

Oliveira-Castro, J. M. (2003). Effects of Base Price upon Search Behavior of Consumers in a Supermarket: An Operant Analysis. *Journal of Economic Psychology, 24*(5), 637-652.

Pansky, A., & Algom, D. (1999). Stroop and Garner Effects in Comparative Judgment of Numerals: The Role of Attention. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 39-58.

Pansky, A., & Algom, D. (2002). Comparative Judgment of Numerosity and Numerical Magnitude: Attention Preempts Automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(2), 259-274.

Parducci, A. (1965). Category Judgment: A Range-Frequency Model. *Psychological Review, 72*(6), 407-418.

Pechmann, C. (1996). Do Consumers Overgeneralize One-Sided Comparative Price Claims and Are More Stringent Regulations Needed? *Journal of Marketing Research, 33*(2), 150-162.

Peterson, C., & Beach, L. R. (1967). Man as an Intuitive Statistician. *Psychological Bulletin, 68*(1), 29-46.

Peterson, C., & Miller, A. (1964). Mode, Median and Mean as Optimal Strategies. *Journal of Experimental Psychology, 68*(4), 363-367.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of Associative Memory. *Psychological Review, 88*(2), 93-134.

Rajendran, K. N., & Tellis, G. J. (1994). Contextual and Temporal Components of Reference Price. *Journal of Marketing, 58*(1), 22-34.

Rao, A. R., & Monroe, K. B. (1988). The Moderating Effect of Prior Knowledge on Cue Utilization in Product Evaluations. *The Journal of Consumer Research, 15*(2), 253-264.

Ratner, R. K., Kahn, B. E., & Kahneman, D. (1999). Choosing Less-Preferred Experiences for the Sake of Variety. *Journal of Consumer Research, 26*(1), 1-15.

Real, L. A. (1991). Animal Choice Behavior and the Evolution of Cognitive Architecture. *Science, 253*(5023), 980-986.

Real, L. A., & Caraco, T. (1986). Risk and Foraging in Stochastic Environments. *Annual Review of Ecology and Systematics, 17*, 371-390.

Reimers, S., & Maylor, E. A. (2005). Task Switching Across the Life Span: Effects of Age on General and Specific Switch Costs. *Developmental Psychology, 41*(4), 661-671.

Reimers, S., & Stewart, N. (2007). Adobe Flash as a Medium for Online Experimentation: A Test of Reaction Time Measurement Capabilities. *Behavior Research Methods, 39*(3), 365-370.

Reips, U.-D. (2001). The Web Experimental Psychology Lab: Five Years of Data Collection on the Internet. *Behavior Research Methods, Instruments, & Computers, 33*(2), 201-211.

Schmidt, W. C. (1997). World-Wide Web Survey Research: Benefits, Potential Problems and Solutions. *Behaviour Research Methods, Instruments & Computers, 29*(2), 274-279.

Schmidt, W. C. (2001). Presentation Accuracy of Web Animation Methods. *Behavior Research Methods, Instruments, & Computers, 33*(2), 187-200.

Schooler, L. J., & Anderson, J. R. (1997). The Role of Process in the Rational Analysis of Memory. *Cognitive Psychology, 32*(3), 219-250.

Schuler, H. J. (1979). A Disaggregate Store-Choice Model of Spatial Decision-Making. *The Professional Geographer, 31*(2), 146-156.

Seiders, K., & Costley, C. L. (1994). Price Awareness of Consumers Exposed to Intense Retail Rivalry - a Field-Study. *Advances in Consumer Research, 21*, 79-85.

Shanks, D. R. (1990). Connectionism and the Learning of Probabilistic Concepts. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 42*(2), 209 - 237.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A Re-Examination of Probability Matching and Rational Choice. *Journal of Behavioral Decision Making, 15*(3), 233-250.

Simonson, I. (1990). The Effect of Purchase Quantity and Timing on Variety-Seeking Behaviour. *Journal of Marketing Research, XXVII*, 150-162.

Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society, Series B* (13), 238-241.

Skitka, L. J., & Sargis, E. G. (2006). The Internet as Psychological Laboratory. *Annual Review of Psychology, 57*(1), 529-555.

Smith, M. A., & Leigh, B. (1997). Virtual Subjects: Using the Internet as an Alternative Source of Subjects and Research Environment. *Behaviour Research Methods, Instruments & Computers, 29*(4), 496-505.

Stanovich, K. E., & West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? *Behavioral and Brain Sciences, 23*(5), 645-665.

Stevens, S. S. (1957). On the Psychophysical Law. *Psychological Review, 64*(3), 153-181.

Stevens, S. S. (1966). A Metric for the Social Consensus. *Science, 151*(3710), 530-541.

Suri, R., & Monroe, K. B. (2003). The Effects of Time Constraints on Consumers' Judgments of Prices and Products. *Journal of Consumer Research, 30*(1), 92-104.

Swets, J. A. (1986). Indices of Discrimination or Diagnostic Accuracy: Their ROCs and Implied Models. *Psychological Bulletin, 99*(1), 100-117.

Tanner, W. P., Jr., & Swets, J. A. (1954). A Decision-Making Theory of Visual Detection. *Psychological Review, 61*(6), 401-409.

Thomas, M., & Morwitz, V. (2005). Penny Wise and Pound Foolish: The Left-Digit Effect in Price Cognition. *Journal of Consumer Research, 32*(1), 54-64.

Thurstone, L. L. (1927). A Law of Comparative Judgment. *Psychological Review, 34*(4), 273-286.

Tulving, E., & Thomson, D. M. (1973). Encoding Specificity and Retrieval Processes in Episodic Memory. *Psychological Review, 80*(5), 352-373.

Tversky, A., & Kahneman, D. (1973). Availability - Heuristic for Judging Frequency and Probability. *Cognitive Psychology, 5*(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment Under Uncertainty - Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and Intentional Processing of Numerical Information. *Journal of Experimental Psychology - Learning Memory and Cognition, 18*(1), 166-179.

Urbany, J. E., Bearden, W. O., & Weilbaker, D. C. (1988). The Effect of Plausible and Exaggerated Reference Prices on Consumer Perceptions and Price Search. *Journal of Consumer Research, 15*(1), 95-110.

Urbany, J. E., & Dickson, P. R. (1991). Consumer Normal Price Estimation - Market Versus Personal Standards. *Journal of Consumer Research, 18*(1), 45-52.

Van Osselaer, S. M. J., Alba, J. W., & Manchanda, P. (2004). Irrelevant Information and Mediated Intertemporal Choice. *Journal of Consumer Psychology, 14*(3), 257-270.

Vanhuele, M., & Dreze, X. (2002). Measuring the Price Knowledge Shoppers Bring to the Store. *Journal of Marketing, 66*(4), 72-85.

Vermeir, I., & Van Kenhove, P. (2005). The Influence of Need for Closure and Perceived Time Pressure on Search Effort for Price and Promotional Information in a Grocery Shopping Context. *Psychology & Marketing, 22*(1), 71-95.

Wakefield, K. L., & Inman, J. J. (1993). Who Are the Price Vigilantes - An Investigation of Differentiating Characteristics Influencing Price Information-Processing. *Journal of Retailing, 69*(2), 216-233.

Waters, E. A., Weinstein, N. D., Colditz, G. A., & Emmons, K. (2006). Formats for Improving Risk Communication in Medical Tradeoff Decisions. *Journal of Health Communication: International Perspectives, 11*(2), 167-182.

Weber, E. U., Shafir, S., & Blais, A.-R. (2004). Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation. *Psychological Review*, *111*(2), 430-445.

Wickelgren, W. A., & Norman, D. A. (1966). Strength Models and Serial Position in Short-Term Recognition Memory. *Journal of Mathematical Psychology, 3*(2), 316-347.

Winer, R. S. (1986). A Reference Price Model of Brand Choice for Frequently Purchased Products. *Journal of Consumer Research*, *13*(2), 250-256.

Winman, A., Hansson, P., & Juslin, P. (2004). Subjective Probability Intervals: How to Reduce Overconfidence by Interval Evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1167-1175.

Yadav, M. S., & Seiders, K. (1998). Is the Price Right? Understanding Contingent Processing in Reference Price Formation. *Journal of Retailing*, *74*(3), 311-329.

APPENDIX A1: Item Descriptions for Experiments 3 and 4

Item	Department	Description	Price
1	Fruit & Veg	White Seedless Grapes (1 Kg)	2.18
2	Fruit & Veg	Royal Gala Apples (1 Kg)	1.49
3	Fruit & Veg	Conference Pears (1 Kg)	1.18
4	Fruit & Veg	Bananas (1 Kg)	0.74
5	Fruit & Veg	Medium Sized Tomatoes (1 Kg)	1.29
6	Fruit & Veg	Mixed Peppers x3	1.25
7	Fruit & Veg	White Potatoes (2.5 Kg)	0.88
8	Fruit & Veg	Baking Potatoes (1 Kg)	0.85
9	Fruit & Veg	Cauliflower	0.78
10	Fruit & Veg	Whole Cucumber	0.68
11	Fruit & Veg	Iceberg Lettuce	0.57
12	Fruit & Veg	Onions (1 Kg)	0.55
13	Fruit & Veg	Carrots (1 Kg)	0.49
14	Fruit & Veg	Bunch of Spring Onions	0.48
15	Fruit & Veg	Half Cucumber	0.34
16	Meat & Poultry	Fresh Chicken Breast Fillets (1 Kg)	7.73
17	Meat & Poultry	Whole Fresh Chicken (2.5 Kg)	4.45
18	Meat & Poultry	Whole Fresh Chicken (1.5 Kg)	2.96
19	Meat & Poultry	Frozen Chicken (1.8Kg)	2.81
20	Meat & Poultry	Whole Fresh Chicken (1 Kg)	1.93
21	Meat & Poultry	Fresh Beef Sirloin Steak (1 Kg)	11.31
22	Meat & Poultry	Fresh Beef Rump Steak (1 Kg)	7.56
23	Meat & Poultry	Fresh Beef Topside (1 Kg)	6.55
24	Meat & Poultry	Extra Lean Minced Beef (1 Kg)	3.07
25	Meat & Poultry	British Fresh Lamb Loin Chops (1 Kg)	9.95
26	Meat & Poultry	British Fresh Leg of Lamb (1 Kg)	6.69
27	Meat & Poultry	New Zealand Leg of Lamb (1 Kg)	5.88
28	Meat & Poultry	Fresh Boneless Leg of Pork (1 Kg)	3.50
29	Meat & Poultry	Smith's Frozen Pork Chops (1 kg)	3.12
30	Meat & Poultry	Fresh Pork Boneless Rolled Shoulder (1 Kg)	2.58
31	Grocery	Smith's Fusilli Pasta Twists 1Kg	0.64
32	Grocery	Chicken Tonight (500g)	1.16
33	Grocery	Dolmio Sauce (500g)	1.11
34	Grocery	Smith's Pasta Sauce (500g)	0.88
35	Grocery	Walkers Assorted Crisps x6	0.92
36	Grocery	KP Hula Hoops Assorted x7	0.81
37	Grocery	Smith's Multi-Pack Crisps x6	0.79

Item	Department	Description	Price
38	Grocery	Smith's Tortilla Chips (200g)	0.28
39	Grocery	Batchelors Cup-a-Soup Tomato x4	0.82
40	Grocery	Oxo Cubes Red x12	0.67
41	Grocery	Knorr Packet Tomato Soup (makes 1.5 pints)	0.53
42	Grocery	Cadburys Dairy Milk (200g)	1.04
43	Grocery	Kit Kat Two Finger x8	0.83
44	Grocery	Smith's White Granulated Sugar (1 Kg)	0.68
45	Grocery	Weetabix x24	1.17
46	Canned Goods	Smith's Red Salmon (212g)	1.73
47	Canned Goods	Smith's Tuna Chunks (215g)	0.32
48	Canned Goods	Smith's Sardines (120g)	0.30
49	Canned Goods	Fray Bentos Corned Beef (340g)	0.95
50	Canned Goods	Smith's Ham (200g)	0.69
51	Canned Goods	Heinz Baked Beans in Tomato Sauce 4 x 415g	1.48
52	Canned Goods	Heinz Baked Beans (415g)	0.38
53	Canned Goods	Smith's Baked Beans (420g)	0.11
54	Canned Goods	Smith's Canned Garden Peas (300g)	0.23
55	Canned Goods	Smith's Canned Sweetcorn (300g)	0.21
56	Canned Goods	Smith's Plum Tomatoes (420g)	0.19
57	Canned Goods	Heinz Ready to Serve Tomato Soup (400g)	0.49
58	Canned Goods	Ambrosia Creamed Rice (425g)	0.43
59	Canned Goods	Smith's Fruit Cocktail (411g)	0.24
60	Canned Goods	Smith's Peach Slices (420g)	0.18
61	Beverages	Tetley Tea Bags x160	2.73
62	Beverages	Smith's Premium Tea Bags x160	2.54
63	Beverages	PG Tips Pyramid Tea Bags x80	1.44
64	Beverages	Smith's Tea Bags x80	0.34
65	Beverages	Nescafe Gold Blend Instant Coffee (100g)	2.14
66	Beverages	Nescafe Instant Coffee (100g)	1.63
67	Beverages	Smith's UHT Pure Orange Juice 4 x 1 Litre	2.18
68	Beverages	Smith's Fresh Pure Orange Juice (1 litre)	0.70
69	Beverages	Smith's UHT Pure Orange Juice (1 litre)	0.32
70	Beverages	Ribena Blackcurrant (1 litre)	2.41
71	Beverages	Pepsi 6 x 330ml	1.94
72	Beverages	Coca Cola (2 litres)	1.28
73	Beverages	Pepsi Cola (2 litres)	1.28
74	Beverages	Schweppes Tonic (1 litre)	0.76
75	Beverages	Smith's Lemonade (2 litres)	0.14
76	Household & Pet Food	Andrex Bathroom Tissue x9	3.58

Item	Department	Description	Price
77	Household & Pet Food	Andrex Bathroom Tissue x4	1.65
78	Household & Pet Food	Smith's Luxury Soft Toilet Tissue x4	1.55
79	Household & Pet Food	Smith's Toilet Tissue x4	0.42
80	Household & Pet Food	Smith's Kitchen Towel x3	0.98
81	Household & Pet Food	Wash 'n' Go Shampoo & Conditioner 200ml	1.49
82	Household & Pet Food	Sure Roll-On Anti-Perspirant (50ml)	1.05
83	Household & Pet Food	Bic Razors Economy x10	0.74
84	Household & Pet Food	Domestos (750ml)	0.90
85	Household & Pet Food	Fairy Liquid (500ml)	0.88
86	Household & Pet Food	Lenor Care (1 litre)	0.88
87	Household & Pet Food	Persil Washing Up Liquid (500ml)	0.72
88	Household & Pet Food	Whiskas Supermeat (390g)	0.47
89	Household & Pet Food	Pedigree Chum (400g)	0.42
90	Household & Pet Food	Kit-E-Kat (400g)	0.37
91	Bakery	Smith's Crusty White Bloomer (800g)	0.74
92	Bakery	Kingsmill Medium White Sliced Loaf (800g)	0.73
93	Bakery	Hovis White Sliced Loaf (800g)	0.68
94	Bakery	Nimble White Sliced Loaf (400g)	0.64
95	Bakery	Smith's Crusty White Split Tin Loaf (800g)	0.63
96	Bakery	Smith's White Seeded Burger Buns x6	0.49
97	Bakery	Smith's Crusty White Bloomer (400g)	0.48
98	Bakery	Smith's White Finger Rolls x6	0.36
99	Bakery	Smith's Sliced Danish Loaf (400g)	0.30
100	Bakery	Smith's White Sliced Bread (800g)	0.23
101	Bakery	Mr Kipling French Fancies x8	1.35
102	Bakery	Mr Kipling Manor House Cake	1.06
103	Bakery	Smith's Jam Doughnuts x10	0.99
104	Bakery	Cadburys Mini Rolls x6	0.98
105	Bakery	Smith's Apple Pies x6	0.40
106	Dairy	Fresh Semi Skimmed Milk (6 pints)	1.51
107	Dairy	Pasteurised Milk (4 pints)	1.03
108	Dairy	Fresh Semi-Skimmed Milk (1 pint)	0.29
109	Dairy	Smith's Fresh Double Cream (284ml)	0.62
110	Dairy	Smith's Fresh Single Cream (284ml)	0.51
111	Dairy	Flora (1Kg)	1.57
112	Dairy	Lurpak Slightly Salted Butter (250g)	0.97
113	Dairy	St Ivel Gold Light (500g)	0.87
114	Dairy	Anchor Butter (250g)	0.82
115	Dairy	Smith's Salted Butter (250g)	0.53

Item	Department	Description	Price
116	Dairy	Smith's Cheddar (1 Kg)	3.08
117	Dairy	Dairylea Cheese Spread Portions x6	0.49
118	Dairy	Medium Eggs x6 (Size 3)	0.72
119	Dairy	Smith's Low Fat Fruit Yoghurt 4 x 125g	0.83
120	Dairy	Muller Fruit Corner (175g)	0.38
121	Frozen Foods	Birds Eye Garden Peas (1.8kg)	2.54
122	Frozen Foods	Birds Eye Garden Peas (907g)	1.17
123	Frozen Foods	Smith's Petit Pois (907g)	0.92
124	Frozen Foods	Smith's Frozen Peas (1Kg)	0.78
125	Frozen Foods	McCain Oven Chips (1.8kg)	1.55
126	Frozen Foods	Birds Eye Potato Waffles x12	1.38
127	Frozen Foods	Smith's Straight Cut Oven Chips (1kg)	0.78
128	Frozen Foods	Smith's Fish Fingers x20 (500g)	2.21
129	Frozen Foods	Birds Eye Cod Fillet Fish Fingers x10	1.34
130	Frozen Foods	Smith's Cod Fillet Fish Fingers x10	1.14
131	Frozen Foods	Birds Eye Roast Beef Platter (340g)	2.10
132	Frozen Foods	Mr Brain's Faggots x4 (378g)	1.00
133	Frozen Foods	Walls Soft Scoop Blue Ribbon Vanilla Ice Cream (2 l)	1.78
134	Frozen Foods	Smith's Soft Scoop Vanilla Ice Cream (4 litres)	1.70
135	Frozen Foods	Mars Chocolate Ice Cream x4	1.64
136	Off-Licence	Stella Artois 24 x 330ml	14.98
137	Off-Licence	Guiness Draught Bitter 4 x 440ml	3.78
138	Off-Licence	Strongbow Dry Cider (2 litres)	2.98
139	Off-Licence	Smith's Bitter 4 x 440ml	0.86
140	Off-Licence	Smith's French Medium Dry White 3 ltr box	9.89
141	Off-Licence	Kumala Reserve Cabernet Sauvignon (75cl)	5.49
142	Off-Licence	Jacobs Creek Shiraz-Cabernet (75cl)	4.98
143	Off-Licence	Smith's Australian White (75cl)	2.73
144	Off-Licence	Smith's Claret (75cl)	2.54
145	Off-Licence	Smith's Liebfraumilch Medium White (75cl)	1.92
146	Off-Licence	Teachers Scotch Whisky (70cl)	11.67
147	Off-Licence	Baileys Original Irish Cream (70cl)	10.47
148	Off-Licence	Smirnoff Vodka Red (70cl)	9.67
149	Off-Licence	Smith's Vodka (70cl)	7.16
150	Off-Licence	Bacardi Breezer 4 x 275ml	4.47

APPENDIX A2: Item Prices for Experiments 3 and 4

Item	0%	1%	2%	3%	4%	5%	7.5%	10%	15%	20%	30%	F20 M5	F20 M20	F80 M5	F80 M20
1	2.18	2.16	2.14	2.11	2.09	2.07	2.02	1.96	1.85	1.74	1.53	2.07	1.74	2.07	1.74
2	1.49	1.48	1.46	1.45	1.43	1.42	1.38	1.34	1.27	1.19	1.04	1.53	1.63	1.42	1.19
3	1.18	1.17	1.16	1.14	1.13	1.12	1.09	1.06	1.00	0.94	0.83	1.21	1.29	1.12	0.94
4	0.74	0.73	0.73	0.72	0.71	0.70	0.68	0.67	0.63	0.59	0.52	0.76	0.81	0.70	0.59
5	1.29	1.28	1.26	1.25	1.24	1.23	1.19	1.16	1.10	1.03	0.90	1.32	1.41	1.59	2.52
6	1.25	1.24	1.23	1.21	1.20	1.19	1.16	1.13	1.06	1.00	0.88	1.19	1.00	1.19	1.00
7	0.88	0.87	0.86	0.85	0.84	0.84	0.81	0.79	0.75	0.70	0.62	0.90	0.96	0.84	0.70
8	0.85	0.84	0.83	0.82	0.82	0.81	0.79	0.77	0.72	0.68	0.60	0.87	0.93	0.81	0.68
9	0.78	0.77	0.76	0.76	0.75	0.74	0.72	0.70	0.66	0.62	0.55	0.80	0.85	0.74	0.62
10	0.68	0.67	0.67	0.66	0.65	0.65	0.63	0.61	0.58	0.54	0.48	0.70	0.75	0.84	1.33
11	0.57	0.56	0.56	0.55	0.55	0.54	0.53	0.51	0.48	0.46	0.40	0.54	0.46	0.54	0.46
12	0.55	0.54	0.54	0.53	0.53	0.52	0.51	0.50	0.47	0.44	0.39	0.56	0.60	0.52	0.44
13	0.49	0.49	0.48	0.48	0.47	0.47	0.45	0.44	0.42	0.39	0.34	0.50	0.54	0.47	0.39
14	0.48	0.48	0.47	0.47	0.46	0.46	0.44	0.43	0.41	0.38	0.34	0.49	0.53	0.46	0.38
15	0.34	0.34	0.33	0.33	0.33	0.32	0.31	0.31	0.29	0.27	0.24	0.35	0.37	0.42	0.66
16	7.73	7.65	7.58	7.50	7.42	7.34	7.15	6.96	6.57	6.18	5.41	7.34	6.18	7.34	6.18
17	4.45	4.41	4.36	4.32	4.27	4.23	4.12	4.01	3.78	3.56	3.12	4.56	4.88	4.23	3.56
18	2.96	2.93	2.90	2.87	2.84	2.81	2.74	2.66	2.52	2.37	2.07	3.03	3.24	2.81	2.37
19	2.81	2.78	2.75	2.73	2.70	2.67	2.60	2.53	2.39	2.25	1.97	2.88	3.08	2.67	2.25
20	1.93	1.91	1.89	1.87	1.85	1.83	1.79	1.74	1.64	1.54	1.35	1.98	2.11	2.39	3.76
21	11.31	11.20	11.08	10.97	10.86	10.74	10.46	10.18	9.61	9.05	7.92	10.74	9.05	10.74	9.05
22	7.56	7.48	7.41	7.33	7.26	7.18	6.99	6.80	6.43	6.05	5.29	7.74	8.28	7.18	6.05
23	6.55	6.48	6.42	6.35	6.29	6.22	6.06	5.90	5.57	5.24	4.59	6.70	7.18	6.22	5.24
24	3.07	3.04	3.01	2.98	2.95	2.92	2.84	2.76	2.61	2.46	2.15	3.14	3.36	2.92	2.46
25	9.95	9.85	9.75	9.65	9.55	9.45	9.20	8.96	8.46	7.96	6.97	10.18	10.90	12.30	19.40
26	6.69	6.62	6.56	6.49	6.42	6.36	6.19	6.02	5.69	5.35	4.68	6.36	5.35	6.36	5.35
27	5.88	5.82	5.76	5.70	5.64	5.59	5.44	5.29	5.00	4.70	4.12	6.02	6.44	5.59	4.70
28	3.50	3.47	3.43	3.40	3.36	3.33	3.24	3.15	2.98	2.80	2.45	3.58	3.83	3.33	2.80
29	3.12	3.09	3.06	3.03	3.00	2.96	2.89	2.81	2.65	2.50	2.18	3.19	3.42	2.96	2.50
30	2.58	2.55	2.53	2.50	2.48	2.45	2.39	2.32	2.19	2.06	1.81	2.64	2.83	3.19	5.03
31	0.64	0.63	0.63	0.62	0.61	0.61	0.59	0.58	0.54	0.51	0.45	0.61	0.51	0.61	0.51
32	1.16	1.15	1.14	1.13	1.11	1.10	1.07	1.04	0.99	0.93	0.81	1.19	1.27	1.10	0.93
33	1.11	1.10	1.09	1.08	1.07	1.05	1.03	1.00	0.94	0.89	0.78	1.14	1.22	1.05	0.89
34	0.88	0.87	0.86	0.85	0.84	0.84	0.81	0.79	0.75	0.70	0.62	0.90	0.96	0.84	0.70
35	0.92	0.91	0.90	0.89	0.88	0.87	0.85	0.83	0.78	0.74	0.64	0.94	1.01	1.14	1.79
36	0.81	0.80	0.79	0.79	0.78	0.77	0.75	0.73	0.69	0.65	0.57	0.77	0.65	0.77	0.65
37	0.79	0.78	0.77	0.77	0.76	0.75	0.73	0.71	0.67	0.63	0.55	0.81	0.87	0.75	0.63
38	0.28	0.28	0.27	0.27	0.27	0.27	0.26	0.25	0.24	0.22	0.20	0.29	0.31	0.27	0.22
39	0.82	0.81	0.80	0.80	0.79	0.78	0.76	0.74	0.70	0.66	0.57	0.84	0.90	0.78	0.66
40	0.67	0.66	0.66	0.65	0.64	0.64	0.62	0.60	0.57	0.54	0.47	0.69	0.73	0.83	1.31
41	0.53	0.52	0.52	0.51	0.51	0.50	0.49	0.48	0.45	0.42	0.37	0.50	0.42	0.50	0.42
42	1.04	1.03	1.02	1.01	1.00	0.99	0.96	0.94	0.88	0.83	0.73	1.06	1.14	0.99	0.83
43	0.83	0.82	0.81	0.81	0.80	0.79	0.77	0.75	0.71	0.66	0.58	0.85	0.91	0.79	0.66
44	0.68	0.67	0.67	0.66	0.65	0.65	0.63	0.61	0.58	0.54	0.48	0.70	0.75	0.65	0.54
45	1.17	1.16	1.15	1.13	1.12	1.11	1.08	1.05	0.99	0.94	0.82	1.20	1.28	1.45	2.28
46	1.73	1.71	1.69	1.67	1.66	1.64	1.60	1.55	1.47	1.38	1.21	1.64	1.38	1.64	1.38

Item	0%	1%	2%	3%	4%	5%	7.5%	10%	15%	20%	30%	F20 M5	F20 M20	F80 M5	F80 M20
47	0.32	0.32	0.31	0.31	0.31	0.30	0.30	0.29	0.27	0.26	0.22	0.33	0.35	0.30	0.26
48	0.30	0.30	0.29	0.29	0.29	0.29	0.28	0.27	0.26	0.24	0.21	0.31	0.33	0.29	0.24
49	0.95	0.94	0.93	0.92	0.91	0.90	0.88	0.86	0.81	0.76	0.67	0.97	1.04	0.90	0.76
50	0.69	0.68	0.68	0.67	0.66	0.66	0.64	0.62	0.59	0.55	0.48	0.71	0.76	0.85	1.35
51	1.48	1.47	1.45	1.44	1.42	1.41	1.37	1.33	1.26	1.18	1.04	1.41	1.18	1.41	1.18
52	0.38	0.38	0.37	0.37	0.36	0.36	0.35	0.34	0.32	0.30	0.27	0.39	0.42	0.36	0.30
53	0.11	0.11	0.11	0.11	0.11	0.10	0.10	0.10	0.09	0.09	0.08	0.11	0.12	0.10	0.09
54	0.23	0.23	0.23	0.22	0.22	0.22	0.21	0.21	0.20	0.18	0.16	0.24	0.25	0.22	0.18
55	0.21	0.21	0.21	0.20	0.20	0.20	0.19	0.19	0.18	0.17	0.15	0.21	0.23	0.26	0.41
56	0.19	0.19	0.19	0.18	0.18	0.18	0.18	0.17	0.16	0.15	0.13	0.18	0.15	0.18	0.15
57	0.49	0.49	0.48	0.48	0.47	0.47	0.45	0.44	0.42	0.39	0.34	0.50	0.54	0.47	0.39
58	0.43	0.43	0.42	0.42	0.41	0.41	0.40	0.39	0.37	0.34	0.30	0.44	0.47	0.41	0.34
59	0.24	0.24	0.24	0.23	0.23	0.23	0.22	0.22	0.20	0.19	0.17	0.25	0.26	0.23	0.19
60	0.18	0.18	0.18	0.17	0.17	0.17	0.17	0.16	0.15	0.14	0.13	0.18	0.20	0.22	0.35
61	2.73	2.70	2.68	2.65	2.62	2.59	2.53	2.46	2.32	2.18	1.91	2.59	2.18	2.59	2.18
62	2.54	2.51	2.49	2.46	2.44	2.41	2.35	2.29	2.16	2.03	1.78	2.60	2.78	2.41	2.03
63	1.44	1.43	1.41	1.40	1.38	1.37	1.33	1.30	1.22	1.15	1.01	1.47	1.58	1.37	1.15
64	0.34	0.34	0.33	0.33	0.33	0.32	0.31	0.31	0.29	0.27	0.24	0.35	0.37	0.32	0.27
65	2.14	2.12	2.10	2.08	2.05	2.03	1.98	1.93	1.82	1.71	1.50	2.19	2.34	2.65	4.17
66	1.63	1.61	1.60	1.58	1.56	1.55	1.51	1.47	1.39	1.30	1.14	1.55	1.30	1.55	1.30
67	2.18	2.16	2.14	2.11	2.09	2.07	2.02	1.96	1.85	1.74	1.53	2.23	2.39	2.07	1.74
68	0.70	0.69	0.69	0.68	0.67	0.67	0.65	0.63	0.60	0.56	0.49	0.72	0.77	0.67	0.56
69	0.32	0.32	0.31	0.31	0.31	0.30	0.30	0.29	0.27	0.26	0.22	0.33	0.35	0.30	0.26
70	2.41	2.39	2.36	2.34	2.31	2.29	2.23	2.17	2.05	1.93	1.69	2.47	2.64	2.98	4.70
71	1.94	1.92	1.90	1.88	1.86	1.84	1.79	1.75	1.65	1.55	1.36	1.84	1.55	1.84	1.55
72	1.28	1.27	1.25	1.24	1.23	1.22	1.18	1.15	1.09	1.02	0.90	1.31	1.40	1.22	1.02
73	1.28	1.27	1.25	1.24	1.23	1.22	1.18	1.15	1.09	1.02	0.90	1.31	1.40	1.22	1.02
74	0.76	0.75	0.74	0.74	0.73	0.72	0.70	0.68	0.65	0.61	0.53	0.78	0.83	0.72	0.61
75	0.14	0.14	0.14	0.14	0.13	0.13	0.13	0.13	0.12	0.11	0.10	0.14	0.15	0.17	0.27
76	3.58	3.54	3.51	3.47	3.44	3.40	3.31	3.22	3.04	2.86	2.51	3.40	2.86	3.40	2.86
77	1.65	1.63	1.62	1.60	1.58	1.57	1.53	1.49	1.40	1.32	1.16	1.69	1.81	1.57	1.32
78	1.55	1.53	1.52	1.50	1.49	1.47	1.43	1.40	1.32	1.24	1.09	1.59	1.70	1.47	1.24
79	0.42	0.42	0.41	0.41	0.40	0.40	0.39	0.38	0.36	0.34	0.29	0.43	0.46	0.40	0.34
80	0.98	0.97	0.96	0.95	0.94	0.93	0.91	0.88	0.83	0.78	0.69	1.00	1.07	1.21	1.91
81	1.49	1.48	1.46	1.45	1.43	1.42	1.38	1.34	1.27	1.19	1.04	1.42	1.19	1.42	1.19
82	1.05	1.04	1.03	1.02	1.01	1.00	0.97	0.95	0.89	0.84	0.74	1.07	1.15	1.00	0.84
83	0.74	0.73	0.73	0.72	0.71	0.70	0.68	0.67	0.63	0.59	0.52	0.76	0.81	0.70	0.59
84	0.90	0.89	0.88	0.87	0.86	0.86	0.83	0.81	0.77	0.72	0.63	0.92	0.99	0.86	0.72
85	0.88	0.87	0.86	0.85	0.84	0.84	0.81	0.79	0.75	0.70	0.62	0.90	0.96	1.09	1.72
86	0.88	0.87	0.86	0.85	0.84	0.84	0.81	0.79	0.75	0.70	0.62	0.84	0.70	0.84	0.70
87	0.72	0.71	0.71	0.70	0.69	0.68	0.67	0.65	0.61	0.58	0.50	0.74	0.79	0.68	0.58
88	0.47	0.47	0.46	0.46	0.45	0.45	0.43	0.42	0.40	0.38	0.33	0.48	0.51	0.45	0.38
89	0.42	0.42	0.41	0.41	0.40	0.40	0.39	0.38	0.36	0.34	0.29	0.43	0.46	0.40	0.34
90	0.37	0.37	0.36	0.36	0.36	0.35	0.34	0.33	0.31	0.30	0.26	0.38	0.41	0.46	0.72
91	0.74	0.73	0.73	0.72	0.71	0.70	0.68	0.67	0.63	0.59	0.52	0.70	0.59	0.70	0.59
92	0.73	0.72	0.72	0.71	0.70	0.69	0.68	0.66	0.62	0.58	0.51	0.75	0.80	0.69	0.58
93	0.68	0.67	0.67	0.66	0.65	0.65	0.63	0.61	0.58	0.54	0.48	0.70	0.75	0.65	0.54
94	0.64	0.63	0.63	0.62	0.61	0.61	0.59	0.58	0.54	0.51	0.45	0.66	0.70	0.61	0.51

Item	0%	1%	2%	3%	4%	5%	7.5%	10%	15%	20%	30%	F20 M5	F20 M20	F80 M5	F80 M20
95	0.63	0.62	0.62	0.61	0.60	0.60	0.58	0.57	0.54	0.50	0.44	0.64	0.69	0.78	1.23
96	0.49	0.49	0.48	0.48	0.47	0.47	0.45	0.44	0.42	0.39	0.34	0.47	0.39	0.47	0.39
97	0.48	0.48	0.47	0.47	0.46	0.46	0.44	0.43	0.41	0.38	0.34	0.49	0.53	0.46	0.38
98	0.36	0.36	0.35	0.35	0.35	0.34	0.33	0.32	0.31	0.29	0.25	0.37	0.39	0.34	0.29
99	0.30	0.30	0.29	0.29	0.29	0.29	0.28	0.27	0.26	0.24	0.21	0.31	0.33	0.29	0.24
100	0.23	0.23	0.23	0.22	0.22	0.22	0.21	0.21	0.20	0.18	0.16	0.24	0.25	0.28	0.45
101	1.35	1.34	1.32	1.31	1.30	1.28	1.25	1.22	1.15	1.08	0.95	1.28	1.08	1.28	1.08
102	1.06	1.05	1.04	1.03	1.02	1.01	0.98	0.95	0.90	0.85	0.74	1.09	1.16	1.01	0.85
103	0.99	0.98	0.97	0.96	0.95	0.94	0.92	0.89	0.84	0.79	0.69	1.01	1.08	0.94	0.79
104	0.98	0.97	0.96	0.95	0.94	0.93	0.91	0.88	0.83	0.78	0.69	1.00	1.07	0.93	0.78
105	0.40	0.40	0.39	0.39	0.38	0.38	0.37	0.36	0.34	0.32	0.28	0.41	0.44	0.49	0.78
106	1.51	1.49	1.48	1.46	1.45	1.43	1.40	1.36	1.28	1.21	1.06	1.43	1.21	1.43	1.21
107	1.03	1.02	1.01	1.00	0.99	0.98	0.95	0.93	0.88	0.82	0.72	1.05	1.13	0.98	0.82
108	0.29	0.29	0.28	0.28	0.28	0.28	0.27	0.26	0.25	0.23	0.20	0.30	0.32	0.28	0.23
109	0.62	0.61	0.61	0.60	0.60	0.59	0.57	0.56	0.53	0.50	0.43	0.63	0.68	0.59	0.50
110	0.51	0.50	0.50	0.49	0.49	0.48	0.47	0.46	0.43	0.41	0.36	0.52	0.56	0.63	0.99
111	1.57	1.55	1.54	1.52	1.51	1.49	1.45	1.41	1.33	1.26	1.10	1.49	1.26	1.49	1.26
112	0.97	0.96	0.95	0.94	0.93	0.92	0.90	0.87	0.82	0.78	0.68	0.99	1.06	0.92	0.78
113	0.87	0.86	0.85	0.84	0.84	0.83	0.80	0.78	0.74	0.70	0.61	0.89	0.95	0.83	0.70
114	0.82	0.81	0.80	0.80	0.79	0.78	0.76	0.74	0.70	0.66	0.57	0.84	0.90	0.78	0.66
115	0.53	0.52	0.52	0.51	0.51	0.50	0.49	0.48	0.45	0.42	0.37	0.54	0.58	0.66	1.03
116	3.08	3.05	3.02	2.99	2.96	2.93	2.85	2.77	2.62	2.46	2.16	2.93	2.46	2.93	2.46
117	0.49	0.49	0.48	0.48	0.47	0.47	0.45	0.44	0.42	0.39	0.34	0.50	0.54	0.47	0.39
118	0.72	0.71	0.71	0.70	0.69	0.68	0.67	0.65	0.61	0.58	0.50	0.74	0.79	0.68	0.58
119	0.83	0.82	0.81	0.81	0.80	0.79	0.77	0.75	0.71	0.66	0.58	0.85	0.91	0.79	0.66
120	0.38	0.38	0.37	0.37	0.36	0.36	0.35	0.34	0.32	0.30	0.27	0.39	0.42	0.47	0.74
121	2.54	2.51	2.49	2.46	2.44	2.41	2.35	2.29	2.16	2.03	1.78	2.41	2.03	2.41	2.03
122	1.17	1.16	1.15	1.13	1.12	1.11	1.08	1.05	0.99	0.94	0.82	1.20	1.28	1.11	0.94
123	0.92	0.91	0.90	0.89	0.88	0.87	0.85	0.83	0.78	0.74	0.64	0.94	1.01	0.87	0.74
124	0.78	0.77	0.76	0.76	0.75	0.74	0.72	0.70	0.66	0.62	0.55	0.80	0.85	0.74	0.62
125	1.55	1.53	1.52	1.50	1.49	1.47	1.43	1.40	1.32	1.24	1.09	1.59	1.70	1.92	3.02
126	1.38	1.37	1.35	1.34	1.32	1.31	1.28	1.24	1.17	1.10	0.97	1.31	1.10	1.31	1.10
127	0.78	0.77	0.76	0.76	0.75	0.74	0.72	0.70	0.66	0.62	0.55	0.80	0.85	0.74	0.62
128	2.21	2.19	2.17	2.14	2.12	2.10	2.04	1.99	1.88	1.77	1.55	2.26	2.42	2.10	1.77
129	1.34	1.33	1.31	1.30	1.29	1.27	1.24	1.21	1.14	1.07	0.94	1.37	1.47	1.27	1.07
130	1.14	1.13	1.12	1.11	1.09	1.08	1.05	1.03	0.97	0.91	0.80	1.17	1.25	1.41	2.22
131	2.10	2.08	2.06	2.04	2.02	2.00	1.94	1.89	1.79	1.68	1.47	2.00	1.68	2.00	1.68
132	1.00	0.99	0.98	0.97	0.96	0.95	0.93	0.90	0.85	0.80	0.70	1.02	1.10	0.95	0.80
133	1.78	1.76	1.74	1.73	1.71	1.69	1.65	1.60	1.51	1.42	1.25	1.82	1.95	1.69	1.42
134	1.70	1.68	1.67	1.65	1.63	1.62	1.57	1.53	1.45	1.36	1.19	1.74	1.86	1.62	1.36
135	1.64	1.62	1.61	1.59	1.57	1.56	1.52	1.48	1.39	1.31	1.15	1.68	1.80	2.03	3.20
136	14.98	14.83	14.68	14.53	14.38	14.23	13.86	13.48	12.73	11.98	10.49	14.23	11.98	14.23	11.98
137	3.78	3.74	3.70	3.67	3.63	3.59	3.50	3.40	3.21	3.02	2.65	3.87	4.14	3.59	3.02
138	2.98	2.95	2.92	2.89	2.86	2.83	2.76	2.68	2.53	2.38	2.09	3.05	3.27	2.83	2.38
139	0.86	0.85	0.84	0.83	0.83	0.82	0.80	0.77	0.73	0.69	0.60	0.88	0.94	0.82	0.69
140	9.89	9.79	9.69	9.59	9.49	9.40	9.15	8.90	8.41	7.91	6.92	10.12	10.84	12.22	19.29
141	5.49	5.44	5.38	5.33	5.27	5.22	5.08	4.94	4.67	4.39	3.84	5.22	4.39	5.22	4.39
142	4.98	4.93	4.88	4.83	4.78	4.73	4.61	4.48	4.23	3.98	3.49	5.10	5.46	4.73	3.98

Item	0%	1%	2%	3%	4%	5%	7.5%	10%	15%	20%	30%	F20 M5	F20 M20	F80 M5	F80 M20
143	2.73	2.70	2.68	2.65	2.62	2.59	2.53	2.46	2.32	2.18	1.91	2.79	2.99	2.59	2.18
144	2.54	2.51	2.49	2.46	2.44	2.41	2.35	2.29	2.16	2.03	1.78	2.60	2.78	2.41	2.03
145	1.92	1.90	1.88	1.86	1.84	1.82	1.78	1.73	1.63	1.54	1.34	1.97	2.10	2.37	3.74
146	11.67	11.55	11.44	11.32	11.20	11.09	10.79	10.50	9.92	9.34	8.17	11.09	9.34	11.09	9.34
147	10.47	10.37	10.26	10.16	10.05	9.95	9.68	9.42	8.90	8.38	7.33	10.72	11.47	9.95	8.38
148	9.67	9.57	9.48	9.38	9.28	9.19	8.94	8.70	8.22	7.74	6.77	9.90	10.60	9.19	7.74
149	7.16	7.09	7.02	6.95	6.87	6.80	6.62	6.44	6.09	5.73	5.01	7.33	7.85	6.80	5.73
150	4.47	4.43	4.38	4.34	4.29	4.25	4.13	4.02	3.80	3.58	3.13	4.58	4.90	5.52	8.72