Numeric Data Frames and Probabilistic Judgments in Complex Real-World Environments

Katie N. Parker

Department of Experimental Psychology

University College London

Thesis submitted for the degree of

Doctor of Philosophy (PhD)

2016

Declaration

I, Katie N. Parker, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed

Katie N. Parker

London 2016

Abstract

This thesis investigates human probabilistic judgment in complex real-world settings to identify processes underpinning biases across groups which relate to numerical frames and formats.

Experiments are conducted replicating real-world environments and data to test judgment performance based on framing and format. Regardless of background skills and experience, people in professional and consumer contexts show a strong tendency to perceive the world from a linear perspective, interpreting information in concrete, absolute terms and making judgments based on seeking and applying linear functions. Whether predicting sales, selecting between financial products, or forecasting refugee camp data, people use minimal cues and systematically apply additive methods amidst non-linear trends and percentage points to yield linear estimates in both rich and sparse informational contexts. Depending on data variability and temporality, human rationality and choice may be significantly helped or hindered by informational framing and format.

The findings deliver both theoretical and practical contributions. Across groups and individual differences, the effects of informational format and the tendency to linearly extrapolate are connected by the bias to perceive values in concrete terms and make sense of data by seeking simple referent points. People compare and combine referents using additive methods when inappropriate and adhere strongly to defaults when applied in complex numeric environments.

The practical contribution involves a framing manipulation which shows that format biases (i.e., additive processing) and optimism (i.e., associated with intertemporal effects) can be counteracted in judgments involving percentages and exponential growth rates by using absolute formats and positioning defaults in future event context information. This framing manipulation was highly effective in improving loan choice and repayment judgments compared to information in standard finance industry formats. There is a strong potential to increase rationality using this data format manipulation in other financial settings and domains such as health behaviour change in which peoples' erroneous interpretation of percentages and non-linear relations negatively impact choice and behaviours in both the short and long-term.

Acknowledgments

I would like to thank Professor Bradley C. Love for his knowledge and expertise in shaping this thesis. I would also like to thank Alan Robert Lane for his support and help throughout the course of my research.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents an overview of the thesis. Firstly, the main topic is introduced with a brief discussion of the motivation and rationale for the research. The key research interests and objectives are then explained alongside the scientific contributions and the outline of the scope of the thesis. This is followed by a description of the whole thesis structure, detailing the experiments undertaken within each chapter and an overview of the main findings.

1.1 Research Motivation

Throughout everyday life, we frequently form judgments based on numerical data which guide our behaviour in different contexts. These judgments range from simple selections with low trade-offs between choice alternatives, to critical decisions which can significantly impact future events. The effectiveness of such decisions depends on our ability to comprehend, interpret and act upon the numerical information made available to us at the point of judgment formation. Many judgments in non-expert contexts are fast and frugal with relatively low consequences. For example, deciding whether to make an impulse purchase based on a price promotion, choosing between scratch cards based on quick computation of the odds, or selecting a restaurant based on the best fish price per ounce. Biases and errors in numerical judgments such as these may not have particularly detrimental effects on future events. However, when people apply the same computational processes and numerical assumptions in more important choice situations involving future finance or health for example, the biases in computation of numeric information in the present can lead to estimates of future events which can have particularly deleterious effects.

In consumer contexts, the effectiveness of numerical judgments depends on our interpretation of numerical and statistical information in a given situation and point in time which leads to payoffs that are realized in the near future. Depending on the context, various factors may interact with decision making capabilities to shape peoples' judgments and use of numerical data which are independent of numerical skills or rationality in the mathematical sense. The underlying motives, interests, hopes, prior knowledge and beliefs of individual decision makers are thus important to how people incorporate numerical data into judgments. Factors such as these may therefore be more influential to judgment and choice than the ability to accurately interpret information in a given data environment at any single point in time, or in isolation from wider factors.

For example, the choice to use a high interest rate credit card to buy a holiday is likely to be driven by factors such as impulsivity and desirability which override rational financial judgment in the consumer context. Commercial and marketing sectors are effective in applying informational frames and manipulations which utilise human emotional responses such as trust and confidence. When activated in the decision making context, these factors are frequently shown to offset rational choice and behaviours. Decisions regarding personal finances and health rely upon peoples' unaided judgment of the information made available to consumers by professional organisations and governing bodies. Therefore, depending on the motives of organisations and how individuals interpret and act upon the data made available, the favourability of future outcomes for individuals and groups can significantly vary. For example, signing a long-term financial agreement without fully understanding interest rates, or interpreting risk statistics as indicative of no need to pay into insurance schemes or visit a doctor are decisions which can significantly impact future health and wealth.

Long-term probabilistic judgments such as these involve accurately interpreting initial values and comparatively evaluating data in the context of choice alternatives to 'cognitively model' the likelihoods of different outcomes. In this sense, people are required to accurately extrapolate values into the future. The ability to accurately model temporal effects is therefore important to understanding how values change over time depending on the functional relations between different variables or the nature of cause and effect in a particular context. Long-term decisions involve the actualization of payoffs far into the future, or possibly never. Thus, the feedback which is more readily available on frequent, short term judgments and choices is less likely to be available for long-term decision making. Combined with the problems of interpreting and applying numerical information (in particular percentage and rate data), the absence of feedback on long-term probabilistic judgment is likely to contribute to the difficulty experienced in deciphering the most optimal choice and course of action. This is particularly relevant in contexts such as financial planning and health related judgment and behaviour.

In this sense, the lack of available feedback compounds the inability to 'cognitively model' probable outcomes by weighting and extrapolating variables into the future. This effect may also interact with how values are discounted over time. For example, individuals who are less able to accurately extrapolate current states into future states may also be more inclined to discount at a higher rate, leading to more 'impulsive' decision making. Conversely, individuals more capable or attuned to extrapolating current values into the future with greater accuracy may express lower discounting rates which could translate into more adaptive financial and health related choices and behaviours.

The robustness of particular numerical biases in human probabilistic judgment is evident in the parallels between the judgmental tendencies of novice and expert decision makers. Trained professionals frequently make fast, heuristic judgments based on varied, complex numerical information which can have critical consequences and far reaching effects. Moreover, expert judgments are often formed under difficult conditions involving limited time and informational constraints. In contexts such as humanitarian aid, emergency services, military, finance, or natural disaster forecasting for example, professionals are depended upon to make highly accurate assessments of current events and forecasts of future outcomes. The findings in this thesis show however, that despite specialist domain knowledge and experience, there are judgmental biases and erroneous numerical processing strategies which characterise peoples' judgments. The findings indicate the potentiality of hazardous outcomes in professional judgment domains, making it important to further examine the characteristics and strengths of expert decision processes to identify how best to apply specialist knowledge, and combine statistical model data with human intuition and insight.

1.2 Research Objectives

In sum, this thesis investigates the effects of numerical data on peoples' ability to formulate effective judgments and choice when data is presented in different frames and formats. The key objectives are to investigate expert and non-expert judgment and choice in real-world decision domains to examine the robustness of biases and understand how levels of numerical skill and domain knowledge may impact judgment performance in different environments. Performance is tested under conditions of differing complexity and informational formats to identify parameters which effect rationality and facilitate or hinder human judgment.

The experiments conducted in this thesis seek to identify biases and limitations in probabilistic inference based on numerical information which effect human judgments across contexts and informational environments. To delineate the scope of this thesis, the objective is to examine the differential effects of numerical information on probabilistic estimates across data contexts and expertise levels. The professional contexts investigated in this thesis, namely the fields of retail and humanitarian aid, are domains also commonly analysed in traditional forecasting research. However, unlike traditional forecasting studies which often focus on the fitting of statistical models to data and human estimates, the experiments undertaken in this thesis are aimed at furthering the understanding of probabilistic judgment based on how people interact with data frames and formats in different contexts. The overall goal is thus one of practical application, involving ultimately how insights may be applied to systems and data formats to aid judgment and choice in both consumer and professional domains.

Rather than comparing experts with novices per se, that aim is to examine performance among both groups to assess the robustness of biases involved in numerical probabilistic inference. As expected, the judgmental biases identified are so strong that experts are shown to exhibit them too. The commonalities in the judgmental processes observed among novice and professional populations is thus illustrative of the overarching biases and propensities which characterize human judgment processes and shape rationality where numeric computations are involved. The parallels in biases are indicative of inherent features in the formation of inferences which hold despite numerical skills, specialist training and domain knowledge. The findings contribute to our understanding of the underlying nature of human rationality and how cognitive processes interact with informational frames and formats in present-day digital and physical decision environments. The evidence provides implications for how we may optimise data environments for effective judgment and choice by manipulating frames and formats to promote effective cognitive adaptation to data for decision making in professional and novice contexts.

Across many judgment environments involving statistical data, human decision processes are characterised by biases and cognitive limitations which yield judgments incongruent with probabilistic models of rationality. Although people are often shown to be ‘irrational’ in the mathematical sense, the ecological perspective suggests that peoples’ judgmental processes are in fact adapted to environmental demands, based on the development of effective and cognitively efficient heuristic strategies and ‘intuitive’ judgment processes (e.g., Gigerenzer, 2008). However, an array of findings from various fields suggest that judgmental processes may not be adapted to particular tasks or environments. There are situations in which human judgments are shown to be particularly ineffective, even when specialist training and domain knowledge are involved. For example, in the context of financial forecasting (Önkal & Muradoğlu, 1994; Wilkie-Thomson, Onkal-Atay, & Pollock, 1997) and investment decision making (Newall & Love, 2015), experts are shown to make poor decisions based on inaccurate numerical estimates. In such cases, many experts' judgments are shown to be no more accurate than those of inexperienced novices (Armstrong, 1980, 1991; Lawrence & O'Connor, 1993).

This suggests that there is a necessity to develop methods and formats of communicating information which are suited to the propensities and characteristics of human decision processes and the way in which numeric data is cognitive encoded of numerical data. Such an approach could be more beneficial than training people to form judgments in accordance with probability theory and statistical forecasting models. Throughout the following chapters, the findings from the domains of professional retail forecasting and humanitarian aid are applied to consumer judgment in an online financial choice environment. The results of the application to financial decision making successfully demonstrate the potential for using data frames and formats to improve judgment and choice based on enhancing the fit between data displays and peoples' cognitive propensities in complex numeric judgment situations.

1.3 Thesis Structure and Summary

Chapter 2 delivers a review of the relevant background literature followed by the first of five experiments in chapter 3 in which forecasting performance is examined among professional retailers employed by a major UK Supermarket. Conducted in the forecasters' every-day environment at the Supermarket's headquarters, employees are assessed in their ability to forecast product sales following linear versus exponential trends when observed in absolute values versus percentage points. Findings delivered insights into peoples' interpretation of non-linear growth functions and the formation of inferences based on applying additive versus multiplicative methods to different number formats. Employees showed a robust propensity to predict future sales by linearly extrapolating the trends based on the last two data points observed per trial. This resulted in systematic under-forecasting (trend-damping) of increasing sales trends and over-forecasting (anti-damping) of decreasing trends.

In chapter 4 the analysis of probabilistic judgment in complex real-world environments is further explored. Experiment 2 assesses the forecasting performance of novices and professional humanitarian aid workers when forecasting refugee camp data. Despite domain knowledge and familiarity with the informational cues, aid workers' predictions were no more accurate than those of novices, and both groups formed judgments by extracting and projecting trends linearly. Noise was shown to increase the tendency to linearly extrapolate and was predictive of forecasting error, particularly so among the aid workers. This suggests that professionals were more likely than novices to seek meaning in complexity, leading them to erroneously extract linear trends. Analysis of the effects of noise showed that when all observable cues trended in the same direction, both groups were more inclined to linearly project the target data congruently with the 'common trend'. The tendency to seek and linearly extend the direction of congruous trends was associated with increased forecasting error, suggesting that regardless of causality, context information was incorporated to help guide judgment formation, thus acting to amplify the linear bias.

In chapter 5, the effects of financial information frames and formats are examined using a randomized controlled trial. Experiment 3a compares traditional industry mortgage price comparison data formats with total mortgage costs presented in absolute values and framed over current versus future interest rates. The simultaneous presentation of the rate alternatives with a future rates default setting is shown to significantly increase the optimality of mortgage choices compared to traditional formats and the sequential disclosure of rata data with a current rates default.

Experiment 3b furthers the examination of the framing effect identified in experiment 3a with the addition of a behavioural disclaimer designed to increase judgment effectiveness by raising the saliency of rate data and the financial implications of future rate variability. The disclaimer was shown have no additive impact on overall judgment effectiveness in either the control or the experimental conditions which suggested judgment performance could not be improved by simply prompting people to apply greater cognitive effort and attentional resources in the context of rate and percentage information. However, there was evidence that the disclaimer increased the propensity to make comparisons between rate frames based on changes in choice scores and proportions of choices made in the current and future rate frames. Behavioural prompts may therefore be useful in judgment contexts in which cues or attributes are weighted more subjectively. For example, evaluative judgments involving choices between houses or cars could be improved by increasing the comparative analysis of certain attributes.

In chapter 6, experiment 4 further explores the current versus future rate framing effect identified in experiment 3a and 3b in the context of monthly mortgage repayment judgments. The components of the effect are examined in two individual data disclosures to test the impact of a default manipulation on loan repayment decisions separately from the effect of a default with the addition of future rate context data. In condition 1 and 2, loan costs for repayment over the loan full term in current rates are disclosed alongside a default figure for the cost at a reduced loan term. In condition 2, the same information is presented, except with additional future rate context data. The added future rate information acts to heighten the saliency of the default figure and greatly increase the range between the minimum and maximum suggested repayment amounts.

The manipulation in condition 1 was shown to significantly increase the monthly amount people chose to repay above the minimum suggested figure to clear the balance over the full term in current rates. In condition 2 however, the addition of the future rate data was shown to be more effective in generating higher repayment judgments. Combined with the tendency to arithmetically combine the minimum and maximum suggested amounts, the positioning of the default in highly salient future rate context information acted to increase the anchoring effect which resulted in more optimal (higher) monthly repayment judgments. Higher temporal preference and lower financial literacy scores were associated with the tendency to make smaller (less optimal) monthly repayment decisions and educational level interacted with the framing manipulations in both conditions. This suggests that utilising defaults and future event context disclosures may be useful in communicating intertemporal effects to reduce the harmful impact of format biases and optimistic tendencies in financial judgments among lower educational groups.

The final chapter 7 provides a summary of the main results and contributions followed by an overall review and evaluation of the findings, including alternative perspectives and study limitations. This is proceeded by further research requirements and suggestions for practical applications and design, followed by final concluding comments.

Chapter 2

Conceptual Background

This chapter summarises findings relating to the theoretical underpinnings of human probabilistic judgment relevant to the biases identified within this thesis. Firstly, various themes and possible explanations are discussed for the biases and propensities in human probabilistic inference. This is followed by the main review which addresses the themes by discussing theory and findings relating to human biases associated with the framing and formatting of numerical information. In the first section of the review, perspectives on the tendency to assume linear relations and the importance of data format to probabilistic judgment are discussed. Bounded rationality vs. mathematical (optimization) models of human judgment are then explained, assessing the frequency hypothesis as an alternative rationale for linear prediction biases. Lastly, perspectives on reasoning in accordance with probability theory are evaluated, citing effects of noise as the source of irrationality.

When forming decisions based on numerical information, the format and presentation of the data across settings and contexts can be very important to how it is interpreted and used in decision making. The framing and format of numerical data is shown to have a strong impact on how people evaluate options, judge risk and make choices across an array of professional and consumer judgment domains. The differences in interpretation can depend on peoples' individual motives and expectations from a given judgment situation, combined with the computational methods employed to process numerical data. Thus, depending on the setting, the snap-shot judgments and heuristic strategies we apply to on-the-spot choices can significantly impact our predictions, shaping future outcomes for better or worse.

The area in which the quality of peoples' choices and probabilistic decisions particularly suffer is where percentage and rate information is involved. This phenomenon is likely to stem from the tendency to assume that percentages can be treated in the same way as absolute values and thus added and subtracted arithmetically. This intuitive, 'additive' method of data processing is a plausible bias in human cognition because it short-cuts the time and effort necessary for more complex geometric operations. The tendency to process percentages and rates as absolute numbers also results in linear probabilistic estimates when on-the-spot numerical computations are extrapolated into future values. Therefore, the propensity to interpret and predict events in the world around us in a linear fashion can shape both perceptions of the environment as well as determine actual outcomes. It also follows that a linear interpretation and prediction bias will increase peoples' sensitivity to linear relations, heightening the inclination to seek and extract linear trends in noisier, more complex environments. For example, in the context of financial information, assuming linearity can lead to overly optimistic cost estimates by underestimating the exponential growth of compound interest.

There are various possible explanations for why people are predisposed to assume liner relations in the world around them and to quantify event occurrences in absolute values. For example, applying a linear line of best fit in a complex, noisy environment provides an effective means of generating an estimate or choice when time, information or knowledge is limited or absent. Based on ecological perspectives, it is plausible to assume that the human mind is better equipped to rationalise in concrete terms using whole numbers rather than single point percentages with normalized base rates. This is because, pre-probability theory, people would have monitored and predicted events by observing them as and when they occurred, accruing data about the environment in frequency format through natural sampling.

From an evolutionary perspective, people are considered to have evolved to rationalise using information in absolute terms. This suggests that the ability to physically compare and contrast information or choice alternatives is likely to be important to gaining the perspective necessary to derive meaning and value from data when making an estimate or selection. Hence, people find it easier to evaluate a potential option or data point when the context or surrounding information is richer, but also simplified and framed in absolute terms using concrete examples which correspond directly to the real world.

The tendency to think and predict linearly is also likely to impact the way in which people interpret changes and variations in quantities of groups, objects and events. For example, people may be primed to seek and detect changes in event frequencies or populations by counting sample sizes rather than inferring proportional differences between subsets in the environment. This again is likely to stem from the propensity to monitor, collate and interpret data in absolute terms, based on counting and comparing frequencies using additive processing methods.

The downside of this bias means that people may be overly sensitive to random fluctuations in quantities and occurrences of events and interpret temporary fluctuations as being true changes in the environment. This possibility suggests that judgments based on concrete frequency counts and comparison of observable samples may restrict people' ability to rationalise beyond the data in the immediate environment, or limit judgment abilities to low variance data, or non-complex domains. The ability to determine actual changes by detecting proportional differences would thus involve understanding and processing base rate information which is often shown to be a barrier to judgment performance. From this perspective, the tendency to think and predict linearly may also act to hinder judgmental processes by limiting the ability to integrate and interpret broader, more complex data. The linear bias could therefore contribute to problems in processing base rate information which is necessary for distinguishing between minor frequency fluctuations and actual large scale proportional shifts in the environment.

The strength of the tendency for linearity and arithmetic operations suggests people may be inherently incapable of using percentage information or estimating probability in accordance with mathematical models. However, it is possible that peoples' cognitive capacity for frequency data and concrete values does not preclude probabilistic rationality in the mathematical sense. Rather than assuming that statistical rationality can only be achieved through accurately interpreting data in probability formats, it is possible that judgments based on frequency data can translate into mathematically accurate probability estimates when judgment variance is taken into consideration. In this view, the natural variation in events as they are encoded in memory create noise which can present as prediction inaccuracy when judgments are examined in isolation. However, when considering a set of estimates for a category of events, it is reasonable to assume that averaging the estimates will cancel out the noise associated with natural sampling to yield inferences which correspond with the predictions made by probability models.

This suggests that people may not be inherently 'irrational' in the mathematical sense, rather that the measurement of human probabilistic inference and the format in which information is communicated is key to ascertaining peoples' judgment rationality. As the wealth of evidence relating to the effectiveness of frequency formats on statistical inference indicate, peoples' judgmental processes and cognitive propensities may not be inherent blockers of statistical rationality. Instead, they represent an alternative methodology for collating, interpreting and applying information. When the data format is fit to the cognitive methods, people are consistently capable of yielding mathematically accurate inferences in Bayesian probability tasks. It is therefore plausible that by taking the mean of peoples' on-the-spot probability estimates for real-world events, the variance associated with peoples' sampling methods will cancel out and judgments will correspond with statistical models.

In answer to these possibilities, the following review discusses the anomalies and biases in human probabilistic judgment relating to the framing and formatting of numerical information. Each subsection provides an overview of findings and theories which relate to the ideas and themes outlined above.

2.1 Simultaneous Versus Sequential Information Processing

It is well recognised in the field of finance and economics that investment returns must be processed *sequentially* using geometric operations to determine values, rather than *simultaneously*, using arithmetic operations, which can lead to gross overestimation of return values (Bodie, Kane & Marcus, 2001). For example, if a stock were to increase by 20% one year, and then decrease by 20% in the next year, the intuitive reaction to this gain and loss is to assume that the stock's final value is equal to its initial value by adding and subtracting, (i.e., simultaneously processing) the two values; +20% (-20%) = 0. However, the final value of the stock would in fact be less than its initial value, computed by multiplying the values geometrically; 20% x (-20%) = -0.4%. The reason for this becomes apparent when drawing attention to the effect of the percentage changes on the base value of the stock in each year. For example, if an initial investment of £100 were to rise by 20% in one year, its value would be £120. If it then fell by 20% the next year, the final value is correctly calculated by applying the second percentage change to the initial value and the results of the first percentage change, thus yielding a final value of £96 (0.4% less than the stock's initial value). Changes in percentages therefore require increases and decreases to be processed multiplicatively in sequence, taking into account the effects of previous percentage changes on base values at each time point.

The misunderstanding of geometric mathematical operations, combined with the added cognitive load of sequentially processing (instead of just additively summing points), can lead people to treat percentages as absolute values in many contexts. This results in the tendency to add and subtract percentages in the same way absolute values are additively combined using the arithmetic mean. Unsurprisingly, this bias is considered responsible for an array of judgmental errors in different settings (Juslin, Karlsson, & Olsson, 2008; Juslin, 2015; Juslin, Lindskog, & Mayerhofer, 2015).

2.2 Processing Percentages and Judging Investment Returns

The effect of percentage information on real-world decision performance is clearly demonstrated throughout the domain of financial decision making. When choosing between financial product options, or making investment decisions based on fees and returns, the tendency to treat percentages as absolutes can lead people to significantly misjudge future gains and losses. This is derived from the assumption of linear relations between input and output variables occurring from simultaneously processing percentages. Evidence of this exists in investment fields, where reframing percentage fees in absolute value formats is shown to increase peoples' attention to the costs (Choi, Laibson & Madrian, 2010; Hastings & Tejeda-Ashton, 2008).

The bias is also observed in experts, for example Newall and Love (2015) found that among 1,973 online investors, 45.7% were unable to determine that the yield from a stock which increased by 10% one year and decreased by 10% the next year, and did not pay any dividends, would be less than its initial value; with 33.9% judging it to be equal to its initial value. These results provide strong evidence of the propensity to interpret returns based on arithmetically summing the gains and losses to yield a final return of zero. Worryingly, error magnitude in this context is therefore dependent on the size of the return. For small changes such as 10%, underestimating losses may not be too detrimental. However, for larger returns, the underestimation of losses can be catastrophic.

Newall (2016a) demonstrated that even in the face of increasing return magnitudes, the tendency to take the arithmetic average of the past returns remained robust. Again, posing the question of final stock value following equal returns of either +/-10% or +/-50%, Newall found that 50.8% of 981 online respondents incorrectly predicted the final return value of the stock for both the 10% and 50% return magnitudes. The modal response of "equal to its initial value" again indicated simultaneous processing in action. Moreover, the ability to correctly judge downside financial risk was shown to positively correlate with financial literacy (Fernandes, Lynch, & Netemeyer, 2014) and numeracy (Cokely, Galesic, Schulz, Ghazal, & Garcia-

Retamero, 2012), the levels of which were moderate and low for literacy (8.6 out of 13) and numeracy (1.6 out of 4), respectively.

Increasing the +/-50% return to +/-100% (to create a final value of zero) showed that, although more people made correct judgments in the +/-100% conditions (53.2%) compared to the +/-10% condition (34.0%), the proportions of incorrect "equal to its initial value" responses reflecting simultaneous processing were virtually the same per condition (43.1% and 40.6% in the +/-10% and +/-100% conditions, respectively). Simultaneous processing was also found to be resistant to financial incentives, with a $0.10 bonus for a correct answer making no difference to rates of error (40.6% without incentives and 48.5% with incentives) or simultaneous processing (35.7% without incentives, and 37.3% with incentives).

The only manipulation found to increase the rate of correct answers based on *sequential* processing of the returns, was the following behavioural prompt: "When answering, try to imagine what would happen to a $100 initial investment over the two years. Think about the investment's value after year one, and then its value after year two". The prompt shifted the modal response from "equal to its initial value" (in the non-prompt condition) to "less than its initial value" in the prompt condition, and this was found to be equally effective regardless of return sizes.

In sum, these findings suggest that biases associated with percentage data are highly robust and difficult to counteract. However, there might be advantages to delivering financial information relating to interest rates in multiple framings or modes. For example, providing rate data in absolute values or frequency formats (e.g., Gigerenzer & Hoffrage, 1995) in combination with a text description of the data and/or task, may be effective in 'nudging' decision makers towards applying the correct methods of processing to improve judgment effectiveness.

2.3 The Linear Prediction Heuristic in Financial Judgment

Increases in judgment error with sizes of investment returns is critical not only for investors, but for the financial well-being of the general population. Consumer financial decision making is a clear example of how quick, cognitively easy computations are performed using the data currently available, which govern our interpretation of likelihood. In this respect, the tendency to simultaneously process percentage and rate information translates into the tendency to linearly extrapolate values into the future.

This *linear prediction heurist* is highly robust and shown to characterise human probabilistic inference across many judgment domains.

For example, taking out a credit agreement based on one's predicted ability to make the repayments is dependent on how compound interest rates are interpreted. Akin to computing investment returns, failure to process interest payments sequentially can result in dramatic underestimation of repayment costs. Imagine for example, a £100 payment made on a credit card at a typical APR of 34.9%, repaid over 12 years. Figure 2.1 shows how a quick summation of the interest costs prior to making a purchase can lead to a drastic misinterpretation of the final credit fee. Exacerbated by opaque and misleading financial information, people frequently make decisions such as these which can have highly adverse long term effects.

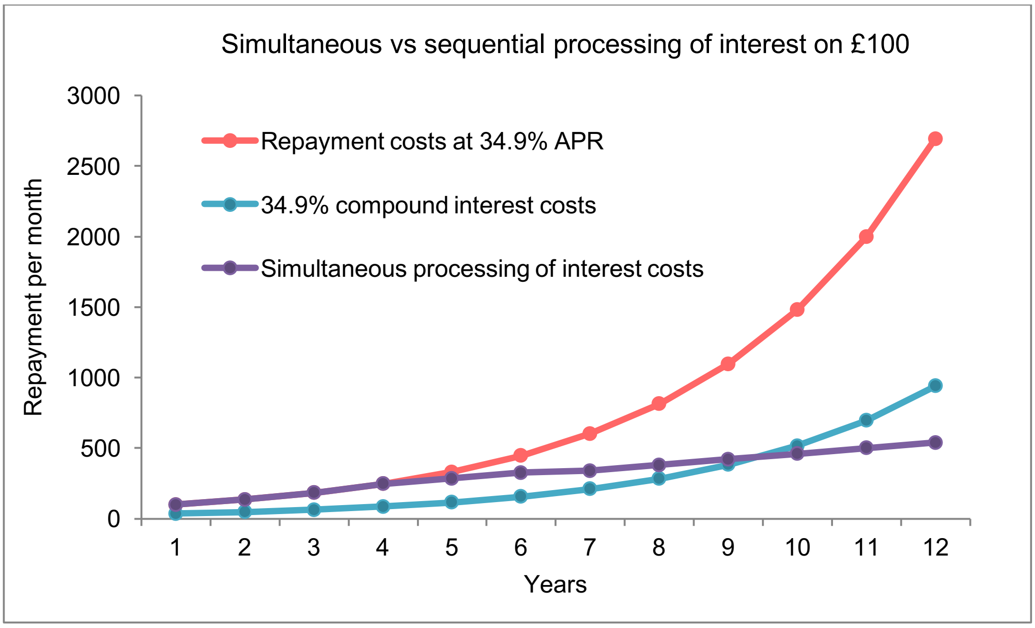


Figure 2.1 Simultaneous Vs Sequential Processing of Interest Rates

The effects of applying arithmetic (simultaneous) processing to interest costs on a £100 loan at an APR of 34.9% over a 12-year period. The purple line shows how the total repayment cost is judged as £540 when the compound interest payments are simultaneously added to the new balance each month. Calculating the interest costs in this way results in a linear prediction of the total interest costs over the term of the agreement. However, the true cost of the credit (based on sequentially processing the interest per month) follows a non-linear curve. The red line shows the true loan costs (£2,692), illustrating the size of the potential judgment error.

Stango and Zinman (2009) formalized the notion of the linear prediction heuristic in an analysis of biases related to decisions to borrow and save. People showed a future value bias associated with saving, and a payment/interest bias associated with borrowing, causing borrowing costs and saving rewards to be underestimated. In each case, biases stemmed from what the authors refer to as the general propensity to linearize functions containing exponential terms, leading to the systematic underestimation of exponential growth, or the *exponential growth bias*.

McKenzie and Liersch (2011) demonstrated similar effects when assessing peoples' ability to compute savings over 10, 20, 30, or 40 years if $200 or $400 were paid in monthly at a fixed annual interest rate of 5% or 10%. Participants systematically underestimated the amount saved at each point in time, based on linear extrapolation of the interest. These findings highlight the widespread misconception of compound interest, and how simultaneously processing percentages and rates can lead to gross underestimations of costs.

2.4 Function Learning Theories

The question of why people exhibit such a strong propensity for linear judgments is addressed by the study of function learning which delivers empirical evidence relating to how people learn predictive relationships between continuous stimulus and response variables.

The theories and models of function learning are useful in explaining how people develop awareness of causal relations between known factors in the environment, and are able to make predictions about future events in new situations (Busemeyer, McDaniel & Byun, 1997). Such theories posit that learners do not explicitly compute the functional forms of causal relationships, but instead learn functional relations through cognitive mechanisms that give rise to a formal system of function awareness (Kalish, 2013). Researchers consider there to be one of two formal systems which perform different kinds of computations that generate the learning of functions. The first system is considered to hypothesize one of a small number of explicit functions in new scenarios and then adjust the function parameters to fit the data (e.g., McDaniel & Busemeyer, 2005). The second system is thought to learn through generalizing from training examples to novel data, based on the degree of similarity between training and novel stimuli (e.g., DeLosh, Busemeyer & McDaniel, 1997).

The *extrapolation-association model* (EXAM) of function learning (DeLosh, et al., 1997) fuses together a system of associative learning with a rule based response to create predictions that accurately reflect human responses in tests of extrapolation. The EXAM model was derived from a series of studies evaluating rule-based and associative-learning model fits to human extrapolation test results. One-hundred and eight participants were assigned to one of nine conditions which varied by function (linear, quadratic and exponential) and density (the number of unique stimulus magnitudes/inputs presented during training). The training phase consisted of 200 correct response (output) feedback trials which were immediately followed by a transfer phase consisting of 45 trials of novel stimulus magnitudes presented without feedback.

Training performance was measured as the absolute deviation of participants' predictions from the correct function value (linear, quadratic, or exponential) for each training trial. Findings showed that learning was best for linear functions at the beginning of training, followed by exponential and worst for quadratic. However, as trials progressed, the learning for all three functions converged. In terms of transfer, participants deviated more significantly from the correct function in the extrapolation regions compared to the interpolation regions and showed underestimation in the linear condition and overestimation in the exponential and quadratic conditions.

EXAM was then evaluated against two rule-learning models: the polynomial hypothesis-testing model (Carroll, 1963; Brehmer, 1974) and the log-polynomial hypothesis-testing model (Koh & Meyer, 1991); and the associative-learning model (ALM) (Busemeyer, Byun, DeLosh & McDaniel, 1997). DeLosh et al. (1997) concluded that EXAM's extrapolations most accurately corresponded with the pattern of over and underestimation empirically observed in the first experiment. The accuracy of EXAM is due to it combining associative learning of stimulus-response pairs with a response generation process that is based on linear interpolation and extrapolation. It is a network model involving a set of input nodes and associated criterion (response) nodes. When a stimulus is presented, a corresponding input node is maximally activated, along with related input nodes which also become activated in accordance with a generalization gradient. The activated input nodes then activate the associated criterion nodes, which sets up the response.

During training, connection weights between input and output nodes become modified by the associative learning process which acts to increase the association between particular input and output nodes. Thus, function learning develops as a process of gradually increasing the associations between input values and output values that represent the correct response for a given input stimulus.

When a novel input value is encountered (in the extrapolation region of the transfer phase of an experiment), a response mechanism is activated which is comprised of two components: the first being the retrieved output which is based on the associative learning process; the second being the activation of a linear response rule. The linear response rule is evoked as a means of deciphering the best response from the multiples of maximally activated input nodes (training stimuli) in a new situation. The output values associated with the multiples of activated input nodes are retrieved and a linear estimate is then made on the basis of these output values. In essence, a response to a novel stimulus occurs as the result of a linear regression performed on the retrieval of the multiple input-output pairs that become activated when people encounter unknown values in new environments.

An alternative model to EXAM is the *population of linear experts* (POLE) model (Kalish, Lewandowsky & Kruschke, 2004) in which learners associate ‘linear experts’ with training stimuli and use the ‘experts’ associated with input stimuli to generate output values. Like EXAM, POLE is based on an associative-learning mechanism and assumes extrapolation is based on a linear rule. The difference lies in POLE assuming that input values are associated with multiple linear rules, each differing in their slopes and intercepts. This assumption is based on the premise that speed and efficiency are the most important characteristics of function learning and the most expedient method for learning input-output pairings, is a positive linear relation. POLE involves a total of six parameters which must be fitted to the data and dictates that complex functions are learnt through a process of deconstructing the functions into small sections which can be estimated using different linear functions. Linear rules are therefore used for both training and extrapolation stimuli in POLE, whereas in EXAM, training stimuli is not assumed to be confound to linear rules only.

In a comparison of the EXAM and POLE model ability to predict (as opposed to fit) transfer performance, McDaniel, Dimperio, Griego and Busemeyer (2009) tested learning and extrapolation in undergraduate students exposed to dense versus sparsely sampled sets of cue values presented in concave and convex functions (experiment 1); and two linear segments of equal cue density presented without tick marks and separated by a wide gap of untrained cue values (experiment 2). The results of the transfer phase showed that extrapolations were flatter than the function in the untrained, upper segments, but remained accurate between the linear segments in experiment 1. In experiment 2, transfer was consistently flatter for the untrained region between the linear segments. McDaniel et al. concluded that EXAM could deliver the best predictors of human extrapolation with an additional small modification to the initial response bias to set it to a positive linear function which would more accurately reflect peoples' positive linear bias in function learning. EXAM's assumptions captured human tendencies notably well in experiment 2 where it reconstructed a flat line in the transfer region between the two linear segments (as opposed to extending and intersecting the linear segments in the transfer region).

Although the EXAM and POLE models posit different mechanisms by which functions are learned, both models support the theory that people learn functions based on a propensity to apply positive linear relations to stimulus-response pairings. This reflects and affirms the empirical findings discussed throughout the review relating to peoples' tendencies to interpret numbers, compute values and predict outcomes in accordance with linear functions and arithmetic operations.

In sum, EXAM incorporates an associative-learning processes (akin to exemplar-based models of categorization) and a response process based on a linear extrapolation rule that operates on the retrieved associations. Although training values are retrieved that most closely match the novel stimulus, the slope of the retrieved training values forms a basis from which people then linearly extrapolate to generate predictions of future outcomes. Phenomena such as the exponential growth bias (Stango & Zinman, 2009; McKenzie & Liersch, 2011), problems comprehending fuel economy based on MPG (Larrick & Soll, 2008), and biases in interpretation of percentage price changes (Chen, Marmorstein, Tsiros & Rao, 2012) are examples of everyday decisions which occur in accordance with these learned mechanisms, shaped by the processes underpinning our understanding of environmental cause and effect.

2.5 Informational Frames in Consumer Choice Domains

The strong tendency to assume linear functions when computing values and judging likelihood is also widely exhibited throughout consumer decision making. In this context, the tendency make choices in accordance with linear functions is based on peoples' interpretation of information when presented in different numerical formats. As described in the studies exploring the effects of percentage information biases on investment decisions and financial judgments, misinterpretation of pricing and promotional disclosures can also significantly impact peoples' choices in everyday consumer settings.

For example, the framing of price promotions is shown to significantly influence how consumers process the information and perceive value (Chen & Rao, 2007; Chen, Marmorstein, Tsiros, & Rao, 2012; Kruger & Vargas, 2008). Chen and Rao (2007) showed that people make erroneous price calculations by additively processing values when successive percentage discounts or surcharges are applied to consumer products. Numerical framing has also been shown to significantly influence consumer choices for economically equivalent offers. For example, choices differ depending on whether a particular product is framed as 50% more expensive than an alternative product, or 33% less expensive (Kruger & Vargas, 2008). Similarly, Chen, et al., (2012) also found that shoppers prefer price offers which deliver 50% more for free as opposed to a 33% reduction in price.

Values framed as fractions can also pose problems for the rationality of value judgments. As exemplified in the case of the US food chain, A&W for example, the retailer was forced to discontinue their new 1/3 pounder shortly after its release, later learning that consumers perceived the '4' in the McDonalds 1/4 pounder rival burger as larger than the '3' in the '1/3' pounder. Thus, less meat was perceived to be in the A&W option, creating the belief that they were being short changed by A&W (Green, 2014).

Systematic errors are also shown when evaluating fuel economy framed in miles per gallon (MPG). Larrick and Soll (2008) found people showed a strong tendency to assume a linear relation between fuel consumption and MPG when evaluating the efficiency of different vehicles, when in fact there is an exponential relation. For example, making a small trade by replacing a 12 MPG with a 14 MPG vehicle represents better value compared to a larger trade of a 28 MPG for a 40 MPG vehicle. Thus, the non-linear increase in fuel consumption with MPG means the saving in the case of the larger 28 to 40 MPG trade is only 107 gallons over 10,000 miles, whereas the difference between the smaller 12 to 14 MPG trade is 120 gallons (10,000/12=833 gallons; 10,000/14=714 gallons).

Further experiments assessing rankings of pairs of 'old and new' vehicles and willingness to pay (WTP) for more fuel-efficient vehicles based on highest fuel consumption reductions showed a clear linear function in participants' choices, causing them to undervalue the impact of small MPG improvements on trades for inefficient vehicles. However, reframing MPG as gallons per mile (GPM) was shown to be effective in reducing the so called *MPG Illusion*. Under this framing condition, 64% of respondents chose the most fuel efficient option which have a smaller MPG improvement but a larger fuel saving. Thus, by removing the computational hurdles associated with exponential relations and rate data, making information comparable in a linear format, consumer choice may be greatly improved.

2.6 Frequency Formats and Cognitive Algorithms

The difficulties people experience in interpreting percentage formats and non-linear functions are likely to stem from the underlying cognitive structures which developed in humans in pre-mathematical environments. There is evidence to suggest that these cognitive structures are designed to make fast, heuristic judgments based on absolute counts or frequencies of events sampled naturally from the environment. As a consequence, people are redisposed to encode events in absolute numbers and form predictions by adding and subtracting numerical values which leads to the tendency to perceive and predict linearly.

There is a wealth of evidence suggesting that cognitive algorithms are programmed to function using absolute values and frequency data. The cognitive anomalies associated with percentages and non-linear functions are consistently shown to disappear when information is presented in frequency formats using (concrete) absolute values (Gigerenzer & Hoffrage, 1995, 1999; Sloman, Over, Slovak, & Stibel, 2003). With the advent of probability theory, human reasoning has been considered analogous with mathematical *optimizing* models (Daston, 1988). However, the more recent concept of human rationality as *bounded,* governed by limited computational capacities, information and time (Simon, 1955), recognises *cognitive algorithms* as incompatible with mathematical models. Rather than *optimizing,* humans seek to *satisfice* via fast and frugal heuristic strategies adapted to meet environmental demands (Simon, 1955).

The effectiveness of natural frequency formats is based on the concept that cognitive algorithms are evolved to make probabilistic inferences based on simple frequency counts, accrued in absolute values. Thus, the mathematical expression of probability is not equivalent to the psychological interpretation of probability (Fiedler, 1988; Gigerenzer, 1991, 1994). The effectiveness of frequencies is associated with the fact that they retain the original numerators and denominators, displaying the base rate information. Conversely, relative or normalized frequencies (single point percentages) are bounded by zero and one, making the base rate information difficult to access. In probabilistic inference tasks, presenting non-normalized frequencies of events sampled from a single population is shown to significantly increase performance in term of Bayes Rule compared to when data is presented in percentages. Using Bayes Theorem, the posterior probability of an event (i.e., the hypothesis) can be estimated in light of new information, according to the following mathematical equation:

$$p(H|D) = p(D\&H) / p(D\&H) + p(D\&{-H})$$

where p(H|D) represents the posterior probability that the hypothesis (H) and its complement (-H) is true given the observed data (D).

For example, when predicting whether a person with a positive test result actually has a disease, the natural frequencies representation of the problem is as follows: In 1000 people, 40 are infected and 30 test positive, and of the 960 uninfected, 120 also test positive. The probability, p(H|D) is therefore A / A + B where A represents the 30 infected people that test positive and B represented the 120 uninfected people that test positive = 30 / (30 + 12) = 0.2 (20%). The base rates for both the hit rate (40/1000) and for the false positive (960/1000) are thus clearly communicated.

However, when considering the problem presented in normalized probabilities, Bayesian reasoning becomes significantly more difficult. For example, out of 1000 people, 40 will be infected, out of 1000 infected, 750 will test positive, and out of 1000 uninfected, 125 will also test positive. In this format, the base rates must be multiplied by the conditional probabilities to determine the probability of infection p(H|D). Figure 2.2 shows an example of a Bayesian probability task presented in a natural frequency format versus single event probabilities.

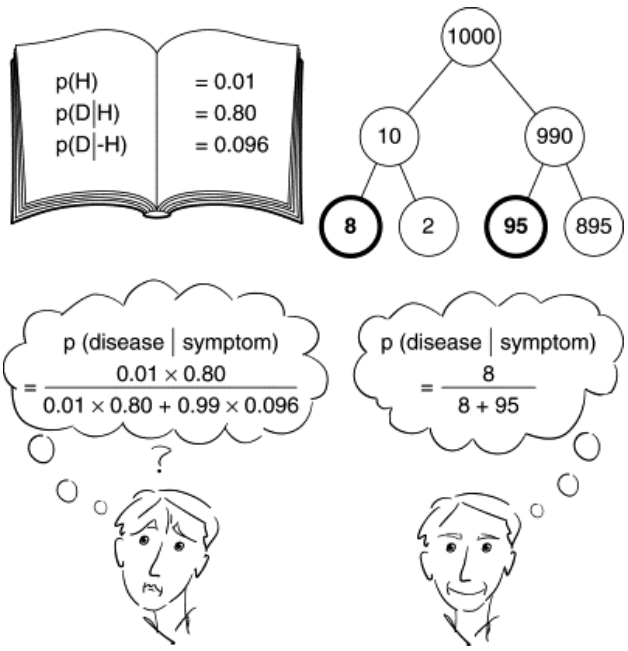


Figure 2.2 Natural Frequecy Vs Probability Formats

An example of a Bayesian probabilistic inference task presented in natural frequencies (right) versus single event probabilities (left). By presenting the problem in the natural frequency format, Bayesian rationality is significantly improved by the exposure of the base rate information which facilitates prediction, p(H|D) of disease (H) given the presence of the symptom (D).

The facilitative effects of frequency formats on Bayesian reasoning were first demonstrated among physicians in the domain of diagnostic decision making (Gigerenzer & Hoffrage, 1995, 1999; Gigerenzer, 1996; Hoffrage & Gigerenzer, 1998; Hoffrage, Lindsey, Hertwig & Gigerenzer, 2000). A wealth of studies have subsequently followed which confirm the effects of frequencies in numerous experimental settings (e.g., Cosmides and Tooby, 1996; Brase, 2002, 2008) and applied contexts such as screening for Down syndrome (Bramwell, West & Salmon, 2006), juror verdicts based on DNA evidence (Hoffrage et al., 2000; Koehler, 1996, Lindsey et al., 2003) and HIV test result interpretations among AIDS counsellors (Gigerenzer, 2002). Teaching people frequency representations shows robust long-term learning effects, including the ability to teach others (Sedlmeier & Gigerenzer, 2001; Kurzenhauser & Hoffrage, 2002), and frequencies enable children to perform Bayesian reasoning as accurately as adults (Zhu & Gigerenzer, 2006). Moreover, frequencies can also support Bayesian reasoning in more complex environments involving multiple cues and cue values (Krauss et al., 2002; Hoffrage et al., 2015).

2.7 Icon Arrays and Denominator Neglect

Findings relating to graph literacy and judgments of medial risk show that dynamic icon arrays (i.e., the use of symbols to convey sample frequencies) are particularly effective in increasing probabilistic inference and the accuracy of risk judgments, specifically among individuals with higher graph literacy (Okan, Garcia-Retamero, Cokely, Maldonado, 2011). In the context of medical screening, findings show that statistical data is widely misinterpreted by both expert physicians and patients (Garcia-Retamero, Wicki, Cokely & Hanson, 2014; Schwartz, Woloshin & Black, 1997; Wegwarth, Schwartz, Woloshin, Gaissmaier & Gigerenzer, 2012).

Inaccurate judgments of statistical information in this context are related to incomprehension of probability formats, associated with lower levels of numeracy and risk literacy (Cokely, Galesic, Schulz, Ghazal & Garcia-Retamero, 2012; Galesic & Garcia-Retamero, 2010; Gardner, McMillan, Raynor, Woolf & Knapp, 2011; Keller & Siegrist, 2009; Lipkus, Samsa & Rimer, 2001; Peters, 2012). Converting probability information into graphical frequency formats (icon arrays) however, is shown to significantly facilitate risk judgment across high and low numeracy groups (Garcia-Retamero & Cokely, 2013; Galesic, Garcia-Retamero & Gigerenzer, 2009; Zikmund-Fisher et al., 2008; Gaissmaier, Wegwarth, Skopec, Muller, Broschinski & Politi, 2012; Zikmund-Fisher, Witteman, et al., 2014; Okan, et al., 2012; Garcia-Retamero & Galesic, 2010). Figures 2.3 and 2.4 give examples of how icons arrays are used to increase risk judgment accuracy in medical decision making contexts.

Visual frequencies also facilitate inference in cancer screening decision making. In a comparison of icon arrays with frequencies in numerical formats combined with text information, Petrova, Garcia-Retamero and Cokely (2015) confirmed previous findings, showing the de-biasing effect of the format even in complex highly emotive situations involving counterintuitive information. Cancer screening risk comprehension was significantly higher where icon arrays were presented, even among participants with a low belief in the severity of the consequences of cancer.

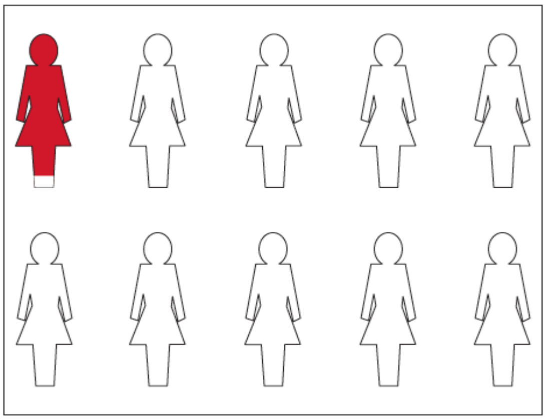


Figure 2.3 Cancer Risk Conveyed Via Frequency Format

An example of the frequency format used to convey a lifetime risk of breast cancer for a 50-year-old woman. The lifetime risk of 9% is portrayed in a frequency format with a denominator of 10 (Schapira, Nattinger, Colleen & McHorney, 2001).

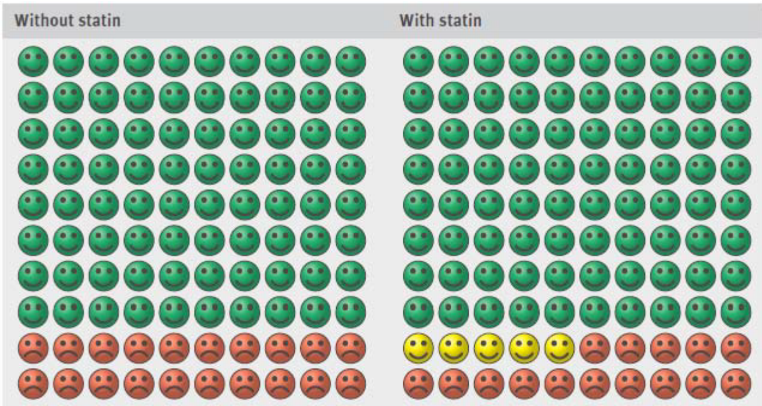


Figure 2.4 NHS National Prescribing Centre Frequecy Communication

NHS National Prescribing Centre: Reduction in Cardiovascular Risk from Taking Statins (Should I be taking statins?). This is an example of a frequency representation used to by the NHS to communicate the probability of cardiovascular disease in the general population when statins are taken (right) versus when they are not (left). As can be seen, the total number of smiley faces represent the sample population (100 people). The green faces represent the frequency of those without cardiovascular disease (80 out of 100), in relation to those with cardiovascular disease (20 out of 100). The yellow smiley faces on the right represented the reduction in the frequency of cardiovascular disease when statins are taken (5 out of 80, thus indicating a 5% reduction in the risk of cardiovascular disease).

The process underpinning the effectiveness of icon arrays is the increase in attention to base rate information. This acts to eliminate *denominator neglect*, recognised as the tendency to focus on numerators and ignore denominators in ratio information. For example, people judge the risk of cancer as a greater when described as “killing 1,286 out of 10,000 people” compared to “24.14 of 100 people” because the numerator (1,286 versus 24.14) is weighted, irrespective of the size of the denominator (10,000 versus 100) (Reyna & Brainerd, 2008; Yamagishi, 1997). Similarly, Denes-Raj and Epstein (1994) showed that when tasked with selecting a winning red bean from a bowl of white beans, people opt for a bowl containing seven red and 100 white beans, rather than a bowl containing one red and 10 white beans. Again, this exemplifies the strong tendency to ignore denominators when estimating event probabilities.

By making the focal and non-focal feature frequencies easily comparable, icon arrays give context and meaning to the data which increases statistical rationality in individuals with low numeracy and ability for statistical inference (Garcia-Retamero & Galesic, 2009; Okan et al., 2012). Specifically, icon arrays convey spatial features and conceptual relations between data which are shown to be important to visual encoding of the focal feature (Cleveland & McGill, 1984, 1986; Simkin & Hastie, 1987; Newman & Scholl, 2012; Okan et al., 2012), and the preservation of natural correspondences to the physical environment (Lakoff & Johnson, 1980; Tversky, 2001, 2009; Tversky, Kugelmass, & Winter, 1991; Fischer, 2012). These factors promote active elaborate processing of numerical information, which combined with text prompts, can enhance statistical inference (Okan, Garcia-Retamero, Cokely & Maldonado, 2015).

The effectiveness of frequencies in supporting elaborate processing was demonstrated in an analysis of psychiatrists’ judgments of the likelihood of psychiatric patients reoffending six months after discharge. Slovic, Monahan and MacGregor (2000) found that, when framed in frequencies (i.e., “20 out of every 100 mental patients typically commit a violent act after their release”), 41% refused discharge. However, when framed in percentages (i.e., “a 20% chance that patients will commit a violent act after their release”), 21% refused. The researchers concluded that frequency framings evoked a clearer perception of an individual patient, which in turn increased their risk perception, whereas percentage information was more ‘benign’ in that it did not create the same imagery and was therefore not as meaningful.

2.8 The Importance of Comparison to Statistical Inference

When considering user experience and interface design, it is recognised that people derive value from numerical information by ascribing meaning through a process of comparing attributes and choice alternatives (Roller, 2011). Judgement performance is shown to differ considerable depending on whether numerical information is presented and evaluated simultaneously or independently. Based on an analysis of presentation mode, Hsee, Loewenstein, Blount and Bazerman (1999) posit the *evaluability hypothesis,* in which all judgments are described as made in either a joint evaluation mode (JE), in which multiple attributes associated with a particular value are viewed simultaneously and evaluated comparatively, or a separate evaluation mode (SE), in which attributes are presented and evaluated in isolation. Attributes differ in the extent to which they may be evaluated, with *difficult-to-evaluate* attributes exerting more influence on judgments when presented in joint evaluation mode, which can lead to preference reversals between values presented in JE versus SE.

In the extended *general evaluability theory*, Hsee and Zhang (2010) clarify evaluability as the extent to which a person has relevant reference information necessary to assess the utility of a particular value. Evaluability is thus considered key to judgment performance, and is heightened in situations when decision makers possess knowledge about the domain and the attributes are either presented in joint mode, or they are inherently evaluable (i.e., they are assessable without learning/knowledge or social comparison, such as when assessing ambient temperature for example).

In the context of probabilistic inference, it is therefore important to consider not only the informational format, but also the nature of the surrounding context or reference information. Additional context information has the potential to create noise which can limit judgment performance in more complex prediction environments (Harvey, Bolger & McClelland, 1994). However, strongly correlated causal data which is specified to the judgment task is shown to facilitate prediction performance (Becker, Leitner & Leopold-Wildburger, 2007, 2008). For example, highly reliable trend change information is shown to positively impact judgment accuracy (Remus, O'Connor & Griggs, 1995) and increases in the reliability of temperature reference data was found to improve the accuracy of soft drink sales forecasts (Lim & O'Connor, 1996). This suggests that the type, format and specificity of surrounding contextual data is important to peoples' ability to interpret and apply reference information when forming judgments and determining preferences.

Findings suggest that the ability to contextualize data through comparison against a reference point of some sort is shown to facilitate statistical inference in many different contexts. For example, participants rated a gamble involving a 7/36 chance of winning $9 as less attractive compared to a gamble involving a 7/36 chance of winning $9 in joint mode with the added potential for a small loss – a 29/36 chance of losing 5c (Bateman, Dent, Peters, Slovic & Starmer, 2007). Researchers suggested that people preferred the latter gamble because in the first case, it was possible to evaluate the probability as it has known upper and lower bounds, however there was no means of contextualizing the monetary gain as 'good' or 'bad'. Therefore, a relatively low probability (7 out of 36 chances) was deemed unattractive. In the joint evaluation mode however, the addition of the 5c loss created a reference point, giving the $9 gain meaning which made the win/lose ratio attractive, i.e., the relatively small chance of loss seemed worth the gamble for the comparatively high value gain.

Slovic, Finucane, Peters and MacGregor (2002) demonstrated the how percentage and frequency formats presented in joint mode can influence decisions by mediating the affect heuristic (the tendency to make 'good' or 'bad' judgments). When informed that new airport safety equipment would "save 98% of 150 lives", participants were significantly more in favour of the expenditure compared to when they were informed that it would "save 150 lives". In isolation, "150" has no meaning and is difficult to evaluate as being positive or negative. However, when put into the context of "98%", participants were able to assign meaning to the data, and evaluated "150" as highly positive, indicated by "98%" being so near the upper bound of the percentage scale.

Similar effects are also found when assessing peoples' sensitivity to the scope (i.e., the size, frequency or duration) of an outcome. For example, when choices are presented in isolation, people are willing to pay the same amount to save 2,000 birds from oil spills as they would 200,000 birds (Desvousges, Johnson, Dunford, Boyle & Wilson, 1993), and donate the same amount to help one victim as they would to help multiple victims (Kogut & Ritov, 2005). When framed in the context of other information, these biases are eliminated. Joint evaluation mode is also shown to impact forecasting judgments, particularly when causally correlated information cues are provided (Becker, Leitner & Leopold-Wildburger, 2007, 2008).

Preference reversals between joint and separate evaluation mode are also apparent in decisions where attributes differ categorically and incrementally. For example, Hsee (1998) showed that in SE, people were willing to pay more for a 24-piece intact dinnerware set than for a 40-piece set containing some broken pieces, however in JE, preferences reversed. Similarly, people in SE paid more for a 7-oz ice cream served in a 5-oz cup (overfilled) than for an 8-oz ice cream served in a 10-oz cup (under-filled), yet their preference reversed in JE. When presented with a choice between a set of 13 mint condition baseball cards, or a set containing 10 mint and three poor condition cards, in SE, the latter choice received higher bids, whereas in JE the 13 mint condition cards were more popular (List, 2002). This demonstrates that when no reference point was provided, the incremental difference of whether the card set contained 10 or 13 was irrelevant to perceived utility. However, when evaluated in the context of the other choice, the categorical difference of the number of mint cards provided a measure of value.

Further analysis of frames and presentation modes is necessary to understand how to utilise evaluative judgment processes to improve judgments involving categorical, or ‘real’ change perception. Hsee and Zhang (2010) thus postulate that in JE, probability estimates will be less linear than in SE. This is due to the incremental differences in probability estimates that a person will express when comparing multiple attributes in JE (i.e., on a continuum), versus the categorical (0/1) probability estimate which will be made when evaluating an attribute in isolation (SE).

2.9 Data Formats and Proportion Change Judgments

The framing of numerical data is also shown to influence peoples’ ability to detect actual changes in the environment distinct from random fluctuations in sample sizes. Fiedler, Kareev, Avrahami, Beier, Kutzner, Hütteret (2016) tested participants’ detection of genuine change by comparing responses to pairs of stimuli which varied either in sample size, *n* (a random fluctuation), proportion, *p* (a genuine environmental change) or both *n* and *p*.

In 49 trials, participants were shown 24 pairs of stimuli which each involved different frequencies of shaded and dotted rectangles representing the *focal* and *non-focal features*, respectively. Each symbol appeared in a 1 second interval and all symbols remained onscreen for 3,000 milliseconds per trial. In half of the pairings, large and small increases and decreases in *p* (the focal feature) occurred across different sample sizes (*n* = 8 and *n* = 16), and in the other half, *n* changed but *p* did not. Participants then indicated per trial whether the sample was drawn from the same set as the previous, or whether it was drawn from a different set based on whether *p* had changed between each pairing (i.e., the frequency of shaded rectangles).

Fiedler et al., (2016) found that peoples' detection of change was influenced by actual *p* changes, with *p* increase judgments occurring most frequently followed by no-change, then decrease judgments. However, *p* change judgments were consistently higher for cases where *n* increased (8 to 16) and lower where *n* decreased (16 to 8). Moreover, judgment accuracy was highest when *n* changes occurred in the same direction as *p* changes, indicating a strong bias to detect real-world changes based on superficial fluctuations in samples.

Difficulties in distinguishing changes in sample frequencies from true population changes can have important consequences in real-world contexts, for example when judging student performance, public opinion votes or disease rates. To test the effects of *n* on *p* in a more realistic setting, Fiedler et al., (2016) examined the effects of format by comparing the symbols (control), to a *probability-plus-n* format in which participants observed normalize percentages of the focal feature plus the relevant *n* per trial (e.g., "Drawn were 8 (16) symbols, 75% shaded and 25% dotted"), a *probability-only* format which omitted *n*, and a *descriptive natural frequency* format in which the frequencies were presented in absolute values (e.g., "Drawn were 12 shaded and 4 dotted symbols").

A robust replication of the findings occurred for the control condition, with increasing *n* supporting *p* increase detection but hindering decrease detection, and vice versa for decreasing *n*. Surprisingly, the same was found in the frequency condition, with main effects for both *p* and *n* changes indicating that *p* change judgments scores were subject to the same biases as those shown in the control condition. However, presenting focal and non-focal feature frequencies as normalized percentages (probability-plus-*n* condition) showed a strong main effect for *p* change judgments (with no *n* changes effect or interaction), indicating the elimination of the bias. The probability-only format yielded the same results, suggesting that the sample size was ignored when percentage was presented.

Communicating the task using the symbols may have been unduly abstract or confusing, particularly when pairings were presented sequentially with the focal feature positioned randomly making comparison more difficult. To investigate whether the bias held when inductively experienced frequencies were presented in a more salient context, the rectangles were replaced with smiling and frowning faces and participants imagined themselves as politicians. They then had to judge the popularity of different campaign speeches based on public reactions (i.e., frequencies of smiling audience faces). This time the trial pairings were presented simultaneously on the right and left of the screen across four *n* combinations (i.e., *n* left = 8 vs. *n* right = 16 and vice versa, and *n* right = 8 vs. *n* left = 8, and *n* right = 16 vs. *n* left = 16). The smiling and frowning faces were positioned randomly and shown onscreen for 2,500 milliseconds. Despite the fact that visual-spatial changes in proportions (from left to right) were involved, as opposed to temporal changes (as with the rectangles), the findings were the same. Again, *p* change judgments were effected not only by *p* changes from left to right, but also by *n* changes from left to right. Both *n* left and *n* right samples contributed to the effect of *n* change on *p* change judgments.

Therefore, whether meaningless geometric symbols were presented sequentially, or smiling human faces in a socially meaningful context were presented simultaneously, proportion change judgments were biased by *n* changes in exactly the same pattern (i.e., *n* increases facilitated *p* increase detection but hindered *p* decrease detection, and vice versa for *n* decreases). Moreover, *n* changes were strong enough to influence *p* judgments even when no *p* change occurred, and in some cases *n* changes fully overrode the influence of *p* changes on participants change judgments.

On first inspection, these findings are surprising given the body of evidence relating to the effectiveness of frequency formats, and in particular icon arrays, in aiding inductive reasoning (e.g., Garcia-Retamero & Cokely, 2013; Gigerenzer & Hoffrage, 1995). No advantages were found from icon array formats although it seemed likely that semantically meaningful face icons combined with simultaneous sample presentation would have facilitated deeper level processing and increased judgment accuracy. Spatial comparison of the focal feature frequencies (i.e., via simultaneous presentation), should have been more effective in conveying spatial features and conceptual relations, thus yielding higher performance compared to when the proportional changes were assessed temporally (i.e., when the symbols were presented in sequence).

Anomalies in *p* judgments occurred when proportional changes were inductively learned from an experienced sample (i.e., observing frequencies of symbols or smiley faces). However, confusions between sample and proportional changes were eliminated when normalized percentage formats were presented. It is therefore possible that the data formats tested were inappropriately applied or measured in this particular judgment task, or that design oversights contributed to the findings in this particular context. For example, the 3,000 millisecond viewing time combined with the non-consecutive display of the focal feature (i.e., symbols and smiley faces) may have reduced peoples' ability to inductively learn proportional changes. Schapira, Nattinger & McAuliffe (2006) demonstrated a similar effect when comparing frequency displays communicating lifetime risk of breast cancer. When the focal feature was consecutively highlighted versus randomly highlighted within the population sample (see figure 2.5), risk was judged as lower by both numerate and innumerate participants.

In the case of Fiedler et al's., (2016) numerical frequencies condition, the 3 second observation time might have also hindered responses. For example, to determine whether a proportional difference actually occurred, the frequencies had to be summed per trial to compute *n* before comparing per pairing. Again, the sequential presentation of each pairing might have made the comparison more difficult under the time restriction.

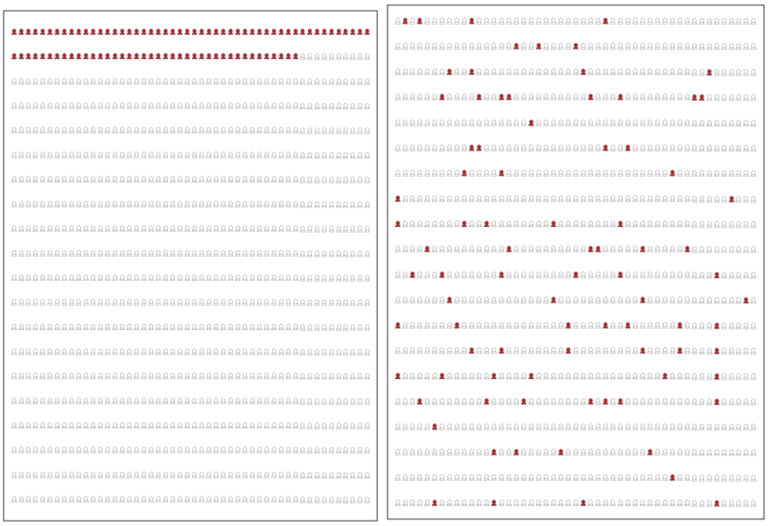


Figure 2.5 Consecutive Vs Random Highlighting of Focal Features

The left panel shows an example of consecutive highlighting of the frequency of the focal feature versus random highlighting (on the right) as used by Schapira, Nattinger & McAuliffe (2006) when comparing the effectiveness of icon arrays in communicating the 9% lifetime risk of breast cancer in 50-year-old women.

2.10 Reasoning in Accordance with Probability Theory

Contrary to Fiedler et al's., (2016) findings, the effectiveness of normalized percentage formats over and above natural frequency representations are not widely supported. Standard normalized percentage formats are consistently shown to yield biased probability estimates, suggesting that rather than reasoning in accordance with probability theory, people employ various heuristic strategies (Cesana-Arlotti, Téglás & Bonatti, 2012; Gigerenzer & Gaissmaier, 2011; Shafir & Leboeuf, 2002; Ariely, 2009; Kahneman, 2011). The systematic biases associated with probability formats support the notion of human judgment as incompatible with mathematical models of probability. This is indicative of the view that rationality has evolved to function based on environmentally adapted heuristic strategies which facilitate human survival (Gigerenzer & Todd, 1999).

However, using a simple model based on probability theory's addition law (equivalent to the equation for disjunction), $P(AvB) = P(A) + P(B) - P(A \wedge B)$, Costello and Watts (2014) showed that people do possess the ability to make unconditional probability estimates in accordance with the laws of probability. From this perspective, judgmental biases such as conservatism, subadditivity and the conjunction and disjunction fallacies (Hilbert, 2012; Tversky & Koehler, 1994; Tversky & Kahneman,1983) are the result of random variation or noise in the judgment process which disappear when predictions for everyday repeated events (i.e., weather events) are averaged. The following simple probability model for events A and B was used to test the assumptions of the heuristic perspective of probabilistic reasoning by creating conjunctive ("and") and disjunctive ("or") probabilities, $P(A \wedge B)$ and $P(AvB)$, as well as the individual probabilities P(A) and P(B).:

$$X_E (A,B) = P_E (A) + P_E (B) - P_E (A \wedge B) - P_E (AvB)$$

where P_E (A) represents a person's estimate for P (A) and P_E (B) represents the person's estimate for P (B), etc. There are there are an equal number of positive and negative terms in the expression which mean that the *d*, or noise terms are cancelled out. The rationale therefore, is that when averaged across predictions, the various biases on the individual expressions will cancel out, resulting in $X_E (A,B) = 0$, in accordance with probability theory's law of addition:

$$X(A, B) = P(A) + P(B) - P(A \wedge B) - P(AvB) = 0.$$

Participants made predictions, P_E (A), P_E (B), P_E (A^B) and P_E (AvB) for 12 pairs of A, B weather events from two sets (cloudy, windy, sunny, thundery and cold, frosty, sleety) which created pairings of low, medium and high probability events. Participants were asked what the probability of a given weather event would be on a randomly selected day in Ireland, for either a single (cloudy), conjunctive (cloudy and cold), or disjunctive event (cloudy or cold) (*probability format*), or on how many days they thought the weather event would occur from a 100 randomly selected days (*frequency format*).

There was no effect of question format on predictions, and X_E (A, B) values for each participants' predictions across the 12 pairs of A, B events were symmetrically distributed around zero with an average X_E value of 0.66 (*SD* = 27.1) which is very close to that predicted by probability theory (the predicted mean of 0 lay within the 99% confidence interval of the observed mean). Conjunctions and disjunctions were recorded at rates of 49% and 51% respectively, however even participants with high conjunction and disjunction fallacy rates still yielded values of X_E close to zero.

In a repeat experiment, the results held for estimates of event probabilities for conjunctions with negations, A ^ ¬B (A and not B), and B ^ ¬A (B and not A) where the derived sum equalled zero, and also for non-symmetric expressions involving an unequal number of positive and negative terms which generated left over noise terms (*d*). Averaging across the symmetrical models yielded a mean of -0.01 (SD = 29.2) (predicted mean = 0), and when one/two noise terms were left after cancellation, the same units of bias were observed in participants' judgments. The mean probability estimates for the biased models showed values closely clustered around the predicted values, with a mean less than 0.001 SD from the predicted values.

In this view, systematic biases in probabilistic judgments are therefore the result of noisy retrieval from memory and that by taking the mean of people's estimates across repeated events, biases vanish to yield judgments in agreement with probability theory. This presents an interesting perspective on how we might approach probabilistic decision making in many situations. However, the importance of informational format to decision processes and performance are still opaque. It is likely that different probabilistic decision situations activate different judgment processes which are better supported by some frames and formats and not others – sometimes percentages and single point probabilities may be appropriate. The interplay between reasoning and data format is likely to be more complex than the frequency hypothesis suggests, particularly as the type and variety of data we regularly interact with is rapidly increasing in all domains. The *probabilistic revolution* is thus ensuing at an increased rate and it is plausible that the human capacity for reasoning in accordance with probabilistic models is strengthening as our scientific and technological focus is being increasingly underpinned by machine learning approaches and mathematical modelling.

Like Fiedler et al., (2016), Costello and Watts (2014) found no difference in probability estimates between framing questions in a numerical frequency formats compared to a probability or percentage data formats. Although on face value this is counterintuitive to the widely supported frequencies hypothesis, it does not rule out the possibility that in a broad sense, people do in fact reason in accordance with the laws of probability (as these laws are conceived by human minds), however, the cognitive processes behind the estimations are still based on recall of raw frequencies drawn from experience. The would account for why biases are cancelled out when judgments are averaged across successive estimates for repeated events. When sampling from memory this would be expected because of the noise in recall which would correspond to natural variation in the environment, including the degree to which events co-vary and interact with one other.

It is likely therefore that the recall of frequency data is likely to be the default method of human statistical inference based on the automatic encoding and storage of event frequencies in memory (Hasher & Zacks, 1984). Evidence of probability estimates among children and animals using frequency data which align with optimizing models further support this perspective (Cesana-Arlotti et al., 2012; Kheifets & Gallistel, 2012).

It is therefore likely that no one information format may be 'good' or 'bad'. Instead, it is probable there is an interplay between how a particular data format is interpreted in a particular situation, the cognitive strategies activated by the task, and the measure used to determine judgment rationality. Thus, rather than perceiving inaccuracies in inferences as a sign of cognitive flaws and biases, it is more likely that such discrepancies represent the variations in event occurrences as they were encoded in memory (i.e., noise associated with the different co-dependencies and co-occurrences of events). As Costello and Watts suggest, averaging over a sample of a persons' probability judgments for a given set of events over time may counteract biases observed when considering individual judgments in isolation. This method may be effective in yielding an overall probability estimate which aligns with probability theory. However, judgment 'rationality' may differ considerably to rationality in the mathematical sense when considering specific real-world contexts and the measures of performance unique to those contexts.

In sum, the findings discussed in the above review suggest that people possess a strong tendency to additively process numbers and assume linearity in the environment based on the ease and efficiency associated with learning and applying linear functions. These tendencies suggest that cognitive algorithms have developed to process information in concrete, absolute formats, based on event frequencies naturally sampled from the environment. The result of being predisposed to 'think linearly' and predict events based on linear relations, means that people are better able to interpret and use absolute values which require arithmetic operations and tend to make errors where percentages are concerned. Percentage format biases involve mishandling base rate information (i.e., failure to apply geometric operations) which translates into problems interpreting non-linear data and predicting future outcomes which following exponential trends.

Although the propensity for additive processing and linear predictions is highly robust through human decision making, it is not necessarily the case that people cannot reason in accordance with statistical models based on probability theory. By altering the informational format from single point probabilities into frequencies using graphical frames and creating context in numerical judgment situations for comparative analysis of the data, people are shown to make mathematically accurate probabilistic inferences. Aside from incomprehensible probability formats, random noise involved in the encoding and recalling of events (i.e., potentially associated with event co-dependencies and co-occurrences) may also account for peoples' judgment irrationality in the mathematical sense. As demonstrated when taking the average of a set of human judgments, mean predictions align with those of probability models.

These results suggest that peoples' probabilistic judgement processes may not be inherently 'flawed'. Instead, the biases that exist result from the developmental processes of human cognition, yielding a propensity for handling data in absolute terms and anticipating the world in a linear fashion. Modern day probability theory thus presents a barrier to peoples' interpretation of information, but not necessarily the human ability to rationalise. Depending on the situation, informational format and surrounding context, people can be facilitated to form effective judgments which are both statistically and environmentally 'rational'. The onus is thus on matching data formats and frames to fit with task demands and the motivational factors and cognitive strategies activated in certain environments to best support the characteristics of human judgment processes.

In the following three chapters, five experiments are discussed which aim to address the questions raised throughout the review relating to how human rationality is differentially effected by numerical data framings and formats. To examine the robustness of biases across different judgment domains and levels of expertise, the probabilistic inferences of professionals and laypersons are assessed in noisy, real-world environments and consumer choice scenarios. Experiment 1 assesses the judgment performance of professional retail forecasters employed by a major UK Supermarket. Employees are tested in their ability to forecast product sales following linear and exponential trends when observed in absolute and percentage formats. This design thus assesses prediction performance based on interpreting *matches* versus *mismatches* between numerical functions and formats. Despite experience, employees displayed a robust tendency to additively process percentages and predict sales by linearly extrapolating trends based on the last two observed data points. This bias resulted in systematic under-forecasting (trend-damping) of increases and over-forecasting (anti-damping) of decreases.

Experiment 2 furthers the analysis of expert probabilistic judgment in the complex domain of humanitarian aid forecasting. Aid professionals are compared to non-experts in their ability to forecast real-world refugee camp data. The effects of noise on judgment performance is tested by presenting target time series data in *sparse* (i.e., single time series) versus *rich* (i.e., additional non-causal time series) contexts. Experts' predictions were no more accurate than novices' and both groups formed judgments by linearly extrapolating the data in both contexts. The tendency to linearly extrapolate increased in noise, particularly so among experts and correlated with forecasting error. One way in which noise impacted judgments was by influencing the trend direction of extrapolation. When all cues trended in the same direction, experts and novices tended to linearly project the 'common trend' and this tendency was predictive of forecasting error.

Experiment 3a applies insights drawn from the previous chapters to the domain of consumer financial decision making. Using a randomized controlled trial design, mortgage choices are compared between three disclosure conditions involving a control (standard industry price comparison format) versus total mortgage costs disclosed in current versus future interest rates with either a ‘current rates’ default (condition 2) or ‘future rates’ default (condition 3). Combined with simultaneous onscreen presentation, the future rates default was significantly more effective in optimizing loan choices compared to the control or sequential presentation with a current rates default.

Experiment 3b tests the framing effect identified in experiment 3a in conjunction with a behavioural disclaimer. The disclaimer is added to each condition to promote consideration of risk associated with BoE rate rises by increasing rate comparisons and depth of evaluative analysis of choice alternatives. The robustness of the framing effect was verified by finding no effect of the disclaimer on the proportion of fully optimized choices per condition. However, there were changes in proportions of choices made per rate frame and the degree of decision optimality (i.e., choice scores) alternated between rate frames with the addition of the disclaimer manipulation.

Experiment 4 extends the rate framing effect to monthly mortgage repayment decisions. In condition 1, total loan costs and interest charges are framed over the full term (20 years) versus a reduced term (10 years) combined with monthly repayment amounts necessary to clear the balance over each term option. In condition 2, the same disclosure is used except with the addition of current versus future rates for the 20 versus 10-year term costs. The combined disclosures of terms and rates in condition 2 significantly increased repayment amounts above the least effective choice (i.e., most optimistic and unrealistic) scenario of the minimum necessary to clear the balance over 20 years in current rates. The combined disclosures were thus more effective in counteracting riskier financial decision making associated optimistic perceptions of rate variability. As expected, higher temporal preference and lower levels of financial literacy were predictive of the tendency to make smaller (less optimal) monthly repayment decisions. Educational level interacted with the framing manipulations in both conditions, indicating the potential advantages of the combined disclosure in counteracting poor financial judgment associated with lower educational levels.

Chapter 3

The Linear Prediction Heuristic in Expert Retail Forecasting

This chapter seeks to address the theoretical question of whether people are prone to treat percentage points as whole numbers and make predictions by linearly extrapolating numerical data. Based on these possibilities, we predict a specific pattern of results among professional retail forecasters employed by a major UK Supermarket when assessed in their ability to forecast product sales trending linearly vs. exponentially and observed in absolute vs. percentage formats. Performance was worst when observed formats 'mismatched' functions and best when formats and functions 'matched'. A robust tendency to linearly extrapolate based on the last two observed data points was shown, leading to systematic under-forecasting of increases and over-forecasting of decreases.

3.1 Background and Rationale

As discussed in the background literature review in chapter 2, the evidence relating to the tendency to make erroneous judgments where percentage and rate information is concerned is numerous and widespread. Both consumers and experts alike are shown to engage in cognitive short-cuts by arithmetically processing percentages when geometric operations are necessary.

In professional settings there is a high demand on experts to make fast, accurate forecasting decisions, often based on highly variable data collated across multiple sources. In retail contexts for example, experts forecast future sales and make stock ordering decisions for thousands of individual products as well as whole product categories.

Effective forecasting decisions in this context involve the application of rich tacit knowledge based on individual experience in the field combined with interpretation of statistical model data. For example, seasonal variability, individual product lead times, promotions and pricing variation, competitor activities, product trends and supplier negotiations are all important 'un-modelled' data points which shape forecasting decisions. Although a proportion of experts' judgment is based on knowledge which exists outside of statistical models, individual stock and sales forecasting accuracy still depends however, on peoples' interpretation and application of the data generated by statistical models. Thus, when it comes to utilizing numeric information in the formation of a forecasting estimate, the 'rationality' of judgments in this particular is context is dependent on the ability to correctly interpret trends in data in different numeric formats which are generated by different statistical models and forecasting systems. The forecasters are then required to synthesize the various data points to make judgments which are informed by the multiple data sources.

To mirror the real-world demands of the forecasting task, experiment 1 involves the retailers making point forecasts for three-monthly sequences of sales data presented in absolute or percentage formats. Akin to the differences in the trends and formats of the model outputs, the sequences in both formats trend either linearly or exponentially. The rationality of judgments is thus dependent on the forecasters statistical understanding of the effects of formats and functions. Specifically, the experiment tests the ability to correctly distinguish the effects of linear versus exponential functions and to correctly project the trends in situations when the observed numeric format differs to the trend function.

Working with different statistical systems to forecast thousands of products across numerous different categories effectuates the propensity for costly errors. For example, interpreting a 10 percent decreases as a 10-point decrease in stock or sales in one display or data set, or summing a +10%, -10% rise and fall to predict no change from base can result in significantly inaccurate predictions of available stock or future sales. Such errors can have dramatic financial impact, especially when dealing with large orders or products involving high manufacturing costs and long lead times. Figure 3.1 shows an example of a commercial forecasting tool in the retail domain, demonstrating the complexity of the prediction environment. The tool encompassed a vast array of data for all general merchandise product categories stocked by the retailer. Forecasters are required to synthesize multiple cues across graphical and tabulated data frames and in percentage and absolute formats to formulate forecasting estimates for stocks and sales levels.

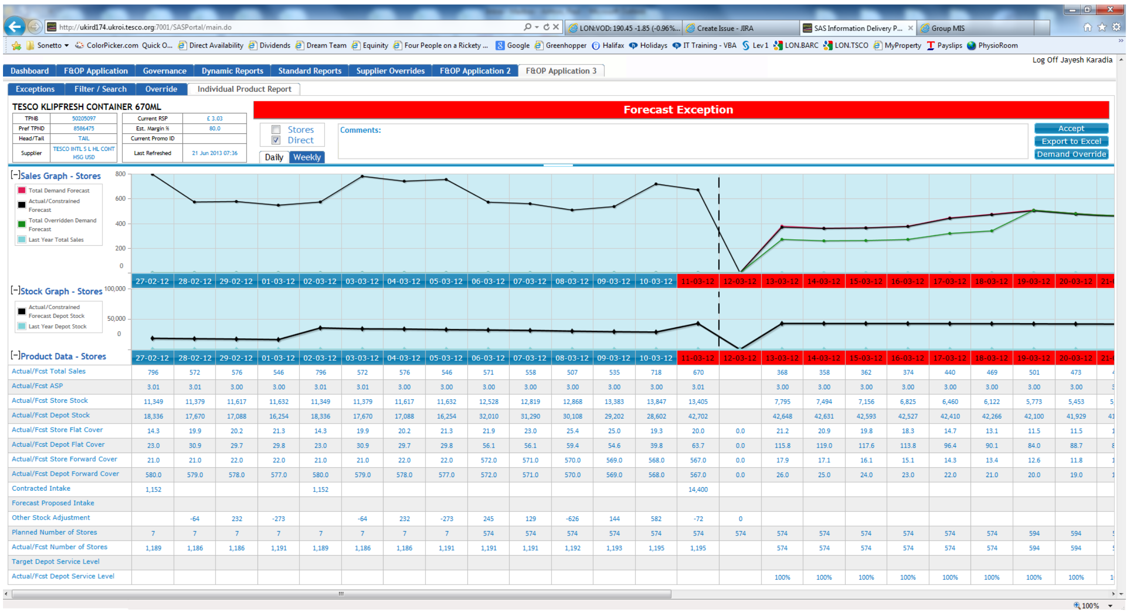


Figure 3.1 Professional Retail Statistical Forecasting Tool

Example of a statistical forecasting tool used by professional retail forecasters to formulate sales predictions and make stock ordering decisions. The forecasting environment for these retailers was complex and varied, involving multiple data cues framed in percentage and absolute values. Judgments were formed based on the synthesis of domain knowledge, time-series and tabulated data points which heightened the propensity to err, based on the process of data comparison across numerical formats and visual representations covering different time horizons.

Given the high task demands involved in professional retail forecasting, and that forecasters deal with multiple data sources daily, it is intuitive to assume that experts will accurately process sequential percentage changes and correctly interpret the effects of non-linear functions. However, the robustness of the tendency to additively rather multiplicatively process percentage information suggests that the human propensity to form predictions based on linear extrapolation will persist, despite experts' specialist forecasting training and experience. To test the hypothesis that experts will exhibit the same interpretational errors and prediction biases observed in consumer and non-expert settings, experiment 1 examines expert merchandise forecasters' ability to forecast sales trending in linear versus exponential functions when viewed in percentage versus absolute number formats.

By alternating linear versus exponential growth functions with the numerical format of the sales figures, comprehension of the effects of numerical functions is revealed through the correct/incorrect use of additive versus multiplicative processing when forming predictions. For example, figure 3.2 shows an exponentially increasing sales trend over four months and what the deleterious effects of additively instead of multiplicatively processing the percentage points can be when forming an estimate for the fourth month of sales.

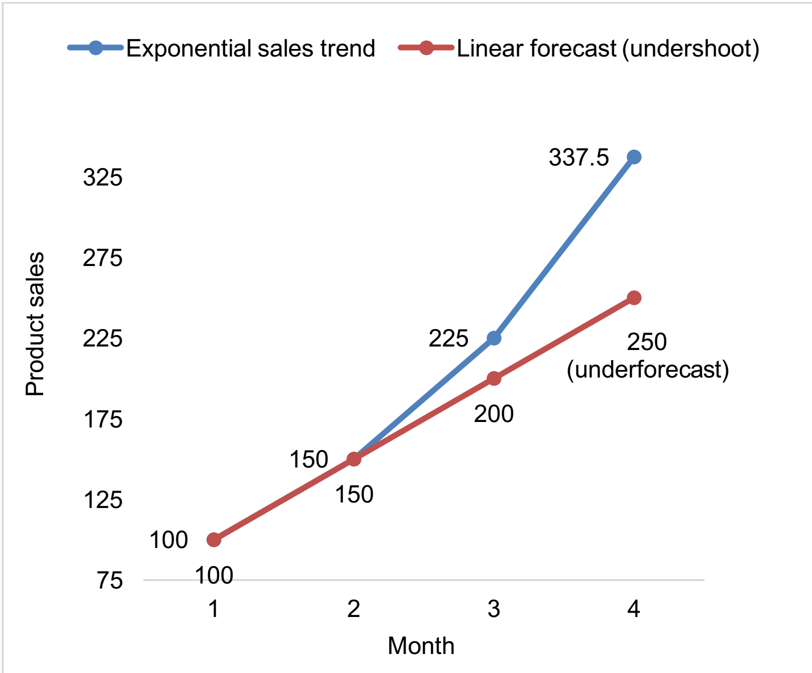


Figure 3.2 Arithmetically Processing Exponential Trends

An example of an exponential sales trend increasing in +50% increments at each time point (blue line) with the sales figures displayed in absolute values and the effects of a sales forecast for the fourth month based on additive processing of the sales figures (the red point). The correct prediction for the fourth month is 337.5 units, which is computed multiplicatively by following the exponential trend (100x0.5, 150x0.5, 225x0.5). However, if month four is computed arithmetically by adding +50 points to the base value at each time point (100+50, 150+50, 200+50=250), a significant under-forecast results. In a retail context, linearly extrapolating an exponential trend in this way can have severe financial impacts which increase in accordance with the magnitude of the base measure. For example, one product unit may represent several hundred/thousand individual products, thus compounding the costliness of the prediction error in terms of the stock deficit.

3.2 Experiment 1

Professional retail forecasters are trained and experienced in monitoring stock and sales levels and making ordering decisions based on predicted sales for thousands of products spanning different categories. The demand on professionals in commercial settings to make effective judgments is thus high, as the financial consequences for poor judgment can have highly detrimental business impacts.

The patterns and biases in professional forecasting judgments can create effects over time which significantly shape business outcomes and impact financial performance. To examine experts' propensity for linear biases (based on additively processing percentage data and linearly extrapolating non-linear trends), the prediction performance of professional retail forecasters employed by a major UK supermarket is assessed for exponential and linear sales trends which alternate between percentage and absolute number formats. It is expected that experts will treat the percentage points as absolute numbers and additively process the onscreen figures (i.e., add or subtract the values) even when inappropriate. It is predicted that this will lead to the formation of linear forecasts based on the arithmetic difference between the last two observations in each sequence. Findings are discussed in relation to the robustness of the linear prediction heuristic which is shown to characterize judgments in this expert domain.

3.2.1 Method

Participants

One hundred and thirty-seven professional retail forecasters employed by a major UK supermarket (42.7% female, M_{age} 34.1 years, SD_{age} 9.3 years) were recruited via a random sampling technique using a mailshot sent to their company email accounts. Each employee possessed an average of three years professional forecasting experience and their full-time duties involved forecasting and ordering all non-food items sold by the supermarket. To compensate for participation, a coffee voucher was emailed to participants on completion of the experiment.

Design

In a repeated measures 2 x 2 design, participants underwent four trials in a percentage format condition and four trials in an absolute number format condition. In the percentage condition, all the observed trial stimuli were shown in percentage format, whereas in the absolute condition, all the observed stimuli were presented in absolute values. In both conditions, two trials increased and decreased in exponential functions and two trials increased and decreased in linear functions. Thus, each condition involved a *match* between the observed numerical format and sequence function in two trials and a *mismatch* between the observed format and sequence function in the other two trials. Trials were repeated (4 x 2) and randomized per condition and conditions were counterbalanced per participant.

Materials

Observed Trial Stimuli

Table 3.1 shows the stimuli presented onscreen in each of the four trials per condition. Each trial involved a sequence of three numbers which related to the changes in product sales over three consecutive weeks. Trials 1 and 2 show the *matched* increase and decrease sequences for the absolute condition (i.e., linear trends observed in absolute values) and trials 3 and 4 show the *mismatched* sequences (i.e., exponential trends observed in absolute values). In the percentage condition, the opposite is shown, with trials 1 and 2 displaying the *mismatched* trial sequences (i.e., linear trends observed in percentages) and trials 3 and 4 displaying the *matched* sequences (i.e., exponential trends observed in percentages).

Table 3.1 Observed Stimuli Per Trial and Condition in Exp 1
The observed stimuli per trial relating to product sales for three consecutive weeks in the absolute and percentage format condition.

Trial 1	Trial 2	Trial 3	Trial 4
Linear increase	Linear decrease	Exponential increase	Exponential decrease
Absolute condition			
+203.125 units	-203.125 units	+100 units	-337.5 units
+203.125 units	-203.125 units	+150 units	-225 units
+203.125 units	-203.125 units	+225 units	-150 units
Percentage condition			
+101.56%	-20.06%	+50%	-33.3%
+50.39%	-25.01%	+50%	-33.3%
+33.51%	-33.51%	+50%	-33.3%

Figure 3.3 shows the onscreen display of the trial stimuli. Each trial involved each three number sequence presented on a black background with each individual number displayed one at a time, transitioning down the screen from top to bottom for a period of three seconds per individual number. After the three second transition period, each individual stimuli value was removed from view and replaced by the next number in the sequence which then transitioned from top to bottom in the same mannor, until all three numbers had been viewed. After the three stimuli had been observed, a response box was displayed at the bottom of the screen and participants were prompted to enter their sales forecast for the fourth week to the nearest whole number (decimal points, minus values, zero's and non-numeric values were not permitted) and press then 'return' when satisfied with their response.

Figure 3.3 The Trial Stimuli in Exp 1

The visual presentation of the trial stimuli in both conditions showing an example of a mismatched exponential increase trial in the absolute format condition. In the absolute condition the individual stimuli (sales figures) were followed by the term 'units' and in the percentage condition the stimuli were followed by a '%' sign. All increasing sequences were presented in green with a preceding plus sign and all decreasing sequences were presented in red with a preceding minus sign.

Generation of the Trial Stimuli

The trial stimuli in each condition were generated using a 'design function' (shown in figure 3.4) created specifically for the purpose of testing the hypotheses relating to the processing of numeric data in accordance with arithmetic versus geometric operations. The design function had a start value of 200.00 and an end value of 1012.50. In both conditions, the increasing trials were generated by moving across the function from left to right, and the decreasing trials were generated by moving from right to left. The italicized values relate to the linear and exponential changes in base rates between the start and finish points, each of which remained hidden throughout all trials. Table 3.3 displays all the values shown per trial and condition which were generated by the design function, including the unobserved base rate values as both an absolute value and percentage point.

The correct predictions for the matched trials in the absolute condition were generated using the arithmetic mean of the three observed stimuli points (e.g., 203.125+203.125+203.125 / 3 = 203.125 for linear increases), whereas the correct predictions for the mismatched trials in the absolute condition were computed using the geometric mean (e.g., 100*1.5*1.5*1.5 = 337.5 for exponential increases). In the percentage condition, the correct predictions for matched trials were produced using the geometric mean of the three observed values (e.g., 200*0.5*0.5 = 50 for exponential increases), whereas the correct predictions for the mismatched trials were computed by applying the arithmetic mean to absolute value changes from base (e.g., 203.125+ 203.125+203.125 / 3 = 203.125 for linear increases). The arithmetic means for percentage condition mismatched trials were then converted back into percentages of the target value base rates to attain the correct predictions in the same format in which the observed trial stimuli was presented.

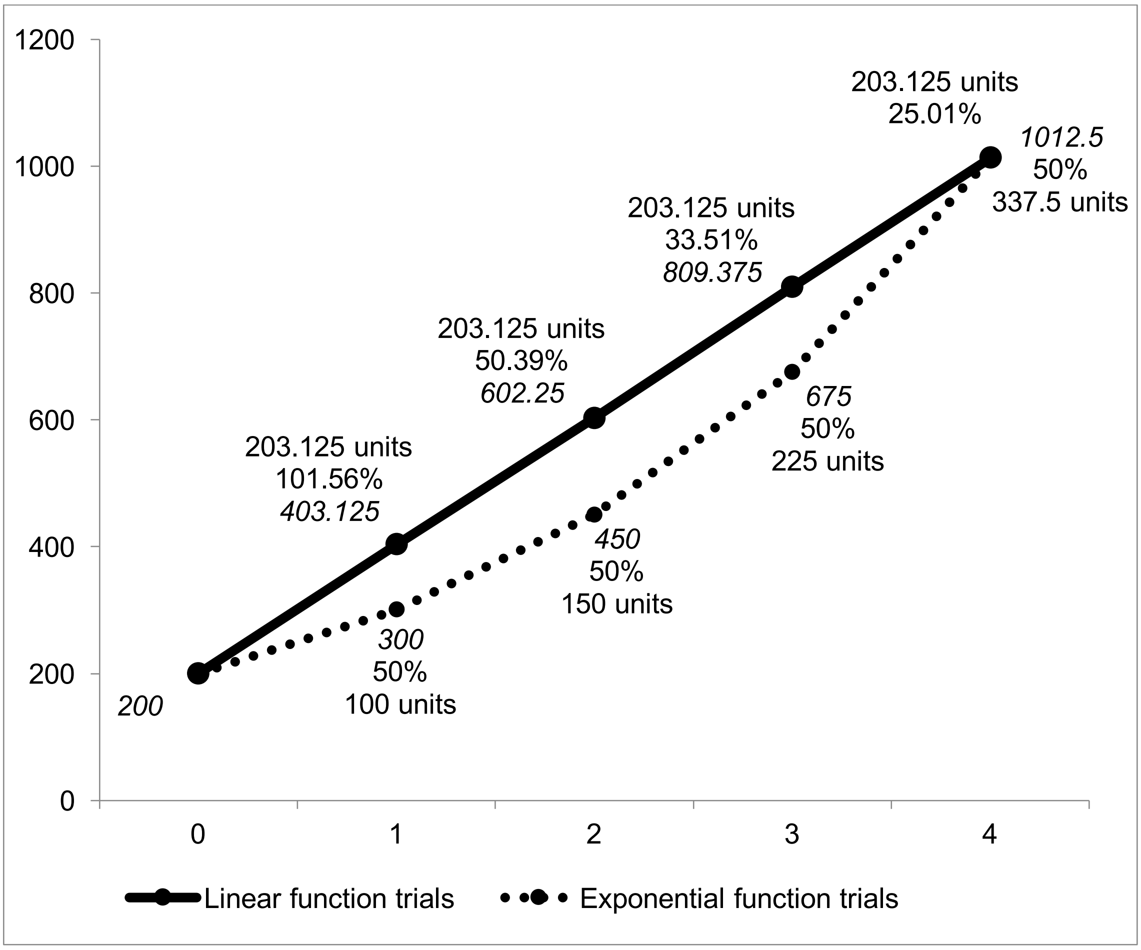


Figure 3.4 The Design Function used to Generate Stimuli in Exp 1

The design function which was developed by the researchers to generate the trial stimuli used to test the hypothesis. The solid line shows the linear trials and the dotted line shows the exponential trials. Moving from left to right across the design function generated the increasing trials in both conditions, and moving from right to left generated the decreasing trials. The italicized values are the linear and exponential changes in base rate between the start value of 200.00 and end value of 1012.50. The base values remained hidden from participants in all trials. The values labelled as 'units' were the trial stimuli observed onscreen in the absolute condition, and the values labelled with '%' signs were the trial stimuli observed onscreen in the percentage condition. All the observed and hidden stimuli values generated by the design function per trial and condition are shown in table 3.2. In the percentage condition, the far right-hand column shows the conversion of the onscreen percentages into absolute values. The method used to convert the observed stimuli in the percentage condition into absolutes can be viewed in appendix 1. Participants responses to the percentage condition trials were also converted into absolute values using the same method to simplify the analysis by having all values in the same format as the design function.

Table 3.2 All Values Generated by the Design Function in Exp 1
All observed trial stimuli and hidden base rate values generated by the design function per trial and condition.

Absolute condition				
	Trial stimuli (Observed)	Target forecast (4^{th} point) (Hidden)	Base rate changes in absolute values (Hidden)	Base rate changes in percentages (Hidden)
Linear increase	203.125 203.125 203.125	203.125	200 403.125 606.25 809.375	2.015625 % 1.503876 % 1.335052 % 1.250965 %
Exponential increase	100 150 225	337.5	200 300 450 675	1.5 % 1.5 % 1.5 % 1.5 %
Linear decease	-203.125 -203.125 -203.125	-203.125	1012.5 809.375 606.25 403.125	0.799383 % 0.749035 % 0.664948 % 0.496124 %
Exponential decrease	-337.5 -225 -150	-100	1012.5 675 450 300	0.666667 % 0.666667 % 0.666667 % 0.666667 %
Percentage condition				
	Onscreen percentages (Observed)	Target forecast (4^{th}point) (Hidden)	Base rate changes in absolute values (Hidden)	Conversions of onscreen percentages into absolutes values (Hidden)
Linear increase	101.56 % 50.39 % 33.51 %	25.01 %	200 403.125 606.25 809.375	203.125 203.125 203.125 203.125
Exponential increase	50 % 50 % 50 %	50 %	200 300 450 675	100 150 225 337.5
Linear decrease	-20.06 % -25.01 % -33.51 %	-50.39 %	1012.5 809.375 606.25 403.125	-203.125 -203.125 -203.125 -203.125
Exponential decrease	-33.3 % -33.3 % -33.3 %	-33.3 %	1012.5 675 450 300	-337.5 -225 -150 -100

Procedure

The experiment was conducted at the supermarket's head-quarters on the employee's personal computers during an ordinary working day. An invitation to participate was sent via mailshot to the employees work email addresses which contained a link to the online experiment. After clicking on the link, participants were provided with the following information on the landing page of the experiment before giving informed consent:

"In this brief experiment you will observe eight trials involving sales of general merchandise products over a three-week period. Green numbers indicate that the sales increased that week whereas red values indicate that they decreased. After observing the changes for each of the three weeks, you will be asked to predict how much the sales will increase or decrease in the fourth week."

After proving details of age, gender and years of professional forecasting experience, participants were then instructed to "press return when ready to observe the sales changes by week" which took them to the first trial.

As shown in figure 3.3, each trial involved the three sales figures for each week transitioning in isolation from the top to the bottom of the screen in a three-second timed interval. After the third sales figure had been observed, a response box was immediately provided at the bottom of the screen and participants were asked to make their prediction for the fourth week by typing in a whole number. No time restriction was applied to responses and participants were free to amend their predictions before submitting the response and moving to the next trial. At the half way point between the trials, a blank partition screen was presented informing respondents that the format would change for the remainder of the trials (e.g., if they had just completed trials in the percentage condition, they would now observe trials in the absolute condition and vice versa). The partition screen delivered the following instruction:

"So far you have seen sales changes in absolute units [percentages]. Weekly sales changes will now be shown in percentages [absolute units], as will the response. Please press RETURN to continue."

On completion of all eight trials, employees were thanked for their participation and provided with a remuneration code for a free drink.

3.2.2 Results

Outlier Removal

To simplify the analysis, participants' responses in the percentage condition were converted into absolute numbers to produce values which were all in the same format as the design function. The method used to convert the percentage condition trial stimuli and responses is shown in appendix 1. Next, outliers were removed by fitting human responses to the design function using absolute squared error per trial. Error per trial was then averaged per participant to give one data point each and the points were plotted. Six data points which fell over 1.5 standard deviations above the mean were removed, leaving 131 respondents in the dataset.

Human Performance in Relation to the Design Function

To determine an initial measure of performance in relation to the design function, participants mean predictions were compared to the design function fourth points represented by the broken lines. Figure 3.5 shows the mean human forecasts per trial and condition.

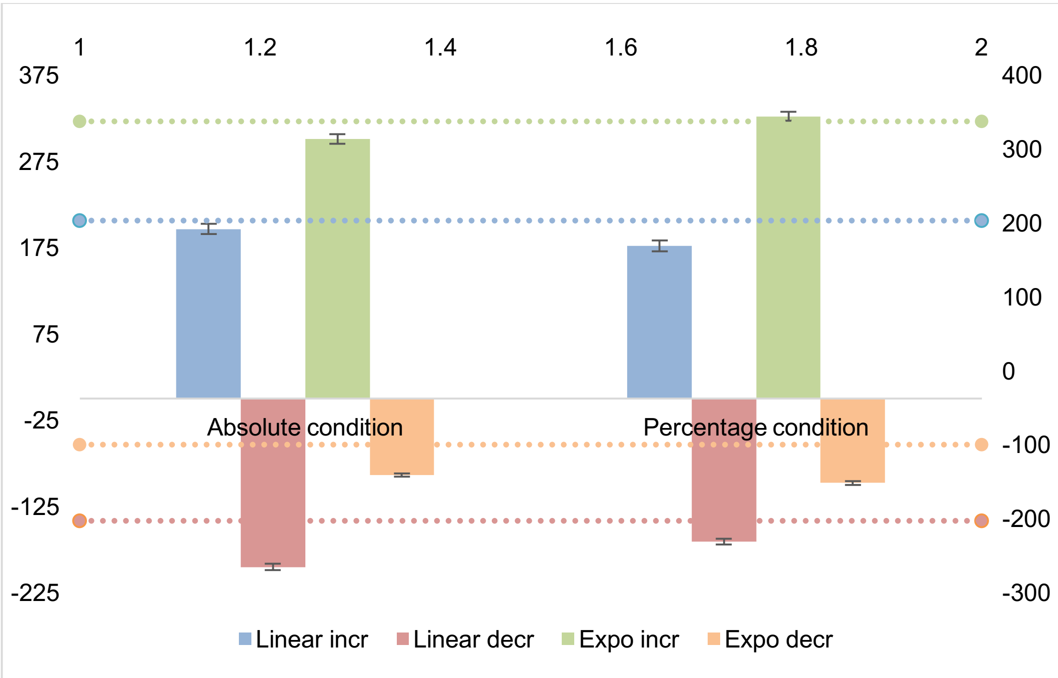


Figure 3.5 Human Performance in Relation to the Design Function in Exp 1

Mean participant forecasts per trial and condition in relation to the design function fourth points (represented by the dotted lines) with 95% within-participant confidence intervals. Each of the four colours of the dotted lines representing the design function correspond with the colours for the four different trial types. Overall, the largest differences in relation to the design function tended to occur for the mismatched trial types in both conditions (i.e., the trials in which observed absolutes increased and decreased exponentially, and observed percentages increased and decreased linearly). In both conditions, the pattern of responses indicated the general tendency to underestimate the increases and overestimate the decreases. This was particularly pronounced for the mismatched trial types. This suggested that prediction error was heightened when trend functions differed to the numerical format in which the points were observed.

To statistically examine performance in relation to the design function, a repeated measures ANOVA was conducted on the absolute forecasting error by trial type and condition. To test for the effects of condition order on prediction performance, percentage versus absolute condition stimuli order was also included as a predictor in the model. Figure 3.6 shows the absolute error per trial and condition. A significant main effect was found for trial type, $F(3,1032) = 5.77$, $p<.001$ and condition, $F(1, 1032) = 8.39$, $p<.01$ which indicated that increases were forecast with less accuracy compared to decreases (M=28.4 vs. M=16.8) and observed percentages were forecast with less accuracy compared to observed absolutes (M=19.2 vs. M=26.0). There was also a significant interaction between trial type and condition, $F(3, 1032) = 46.50$, $p<.001$ which showed that performance was heightened when the observed formats matched the trend functions and reduced when observed formats mismatched the trend functions (M=9.3 vs. M=35.9). There was no effect of condition order on absolute forecasting error, $F(1,1032) = 0.14$, $p=0.70$, nor an interaction between condition order and trial type, $F(3, 1032) = 0.31$, $p=0.81$. It was therefore assumed that condition order had no effect on forecasting performance. Thus for simplicity, the analysis proceeds from this point excluding condition order as a factor.

Post hoc pairwise comparisons with Bonferroni corrected p values confirmed that across conditions, exponential decreases (M=10.40) were forecast with significantly greater accuracy compared to exponential increases (M=27.99), $p<.001$, linear increases (M=28.91), $p<.001$, and linear decreases (M=23.20), $p<.05$. Overall, absolute condition trials were forecast with greater accuracy (M=19.24) compared to percentage condition trials (M=26.01), $p<.05$. The significant interaction indicated that forecasting error was significantly higher for trials when they were observed in mismatched functions.

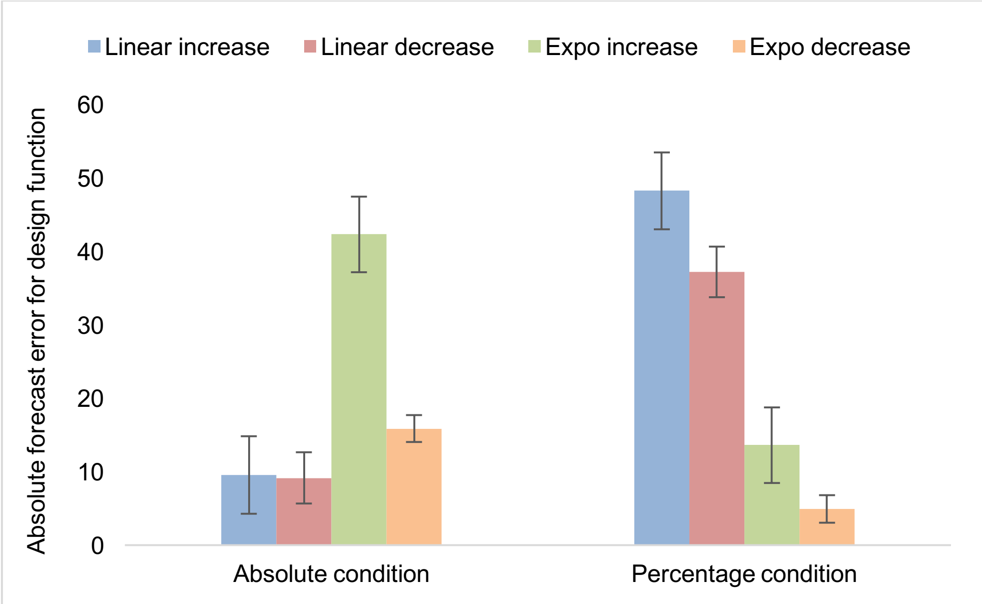


Figure 3.6 Absolute Error in Relation to the Design Function in Exp 1

Absolute forecasting error in relation to the design function per trial with 95% within-participant confidence intervals. The higher bars in each condition confirm the results of the plots in figure 3.5, showing that performance was worst on the mismatched trials which supports the significant interaction between trial type and condition (i.e., accuracy depended on the observed formats being matched to the sales trend functions in both conditions).

Performance in the Mismatched Trials

Figure 3.7 shows the mean forecasting error for matched versus mismatched trial types. Paired t-tests indicated that there was significantly higher forecasting error in relation to the design function in the mismatched compared to the matched trial versions for the linear increases, $t(130) = 7.21$, $p<.001$, linear decreases, $t(130) = 7.95$, $p<.001$, exponential increases, $t(130) = 5.48$, $p<.001$, and exponential decreases, $t(130) = 5.74$, $p<.001$. In terms of directionality, error was negative in relation to the design function target point for the increasing trials (indicating an 'under-forecast' effect), and positive for the decreasing trials (indicating an 'over-forecast' effect). In accordance with the hypothesis, this pattern of human forecasting error indicates the propensity to form linear predictions when processing percentage information based on the tendency to forecast towards the mean. This results in estimates which act to under-forecast increasing and over-forecast decreasing trends.

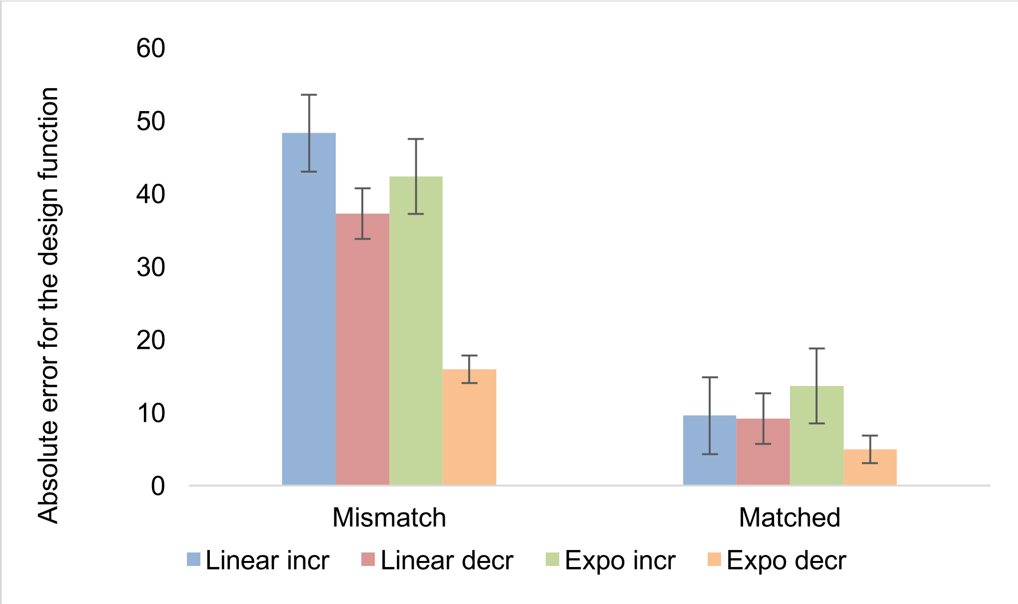


Figure 3.7 Error in the Matched Vs Mismatched Trials in Exp 1

Mean forecasting error for mismatched versus matched trial types in relation to the design function fourth points with 95% within-participant confidence intervals. The degree of error per trial relates to the extent of under-forecasting in the case of the increasing trials and over-forecasting in the case of the decreasing trials. Thus as shown, the greatest extent of underestimation occurred for the mismatched linear and exponential increase sequences. The higher overall error for all the mismatched trials compared to the matched, indicated that participants were less effective at predicting numerical outcomes when the necessary method of processing (i.e., additive or multiplicative) was incongruent with the expression of the observed units (i.e., percentage points or absolute units).

Model-based Analysis of Forecasting Error

In reference to the hypothesis, it is predicted that the linear extrapolation bias is robust to the extent that even trained and experienced expert retail forecasters will show a strong tendency to form predictions by projecting trends linearly. Regardless of whether percentage points or absolute numbers are observed, it is expected that experts will treat all numerical stimuli as absolute values which will lead to predictions which follow linear trends, irrespective of whether the actual growth functions are linear or exponential. In this view, the linear prediction bias is expected to be so pronounced that it will influence peoples' interpretation of growth functions and override the ability to accurately distinguish between linear and non-linear trends (and thus make effective estimates of sales and stock values).

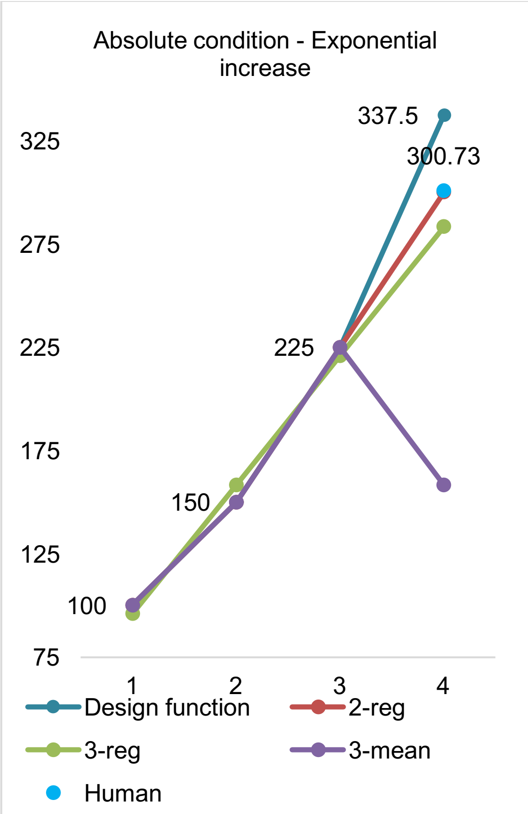
To test the propensity to project trends linearly, experts' predictions were examined in relation to three formal models; two different linear models and one averaging model. This was done by fitting participants' estimates to each model per trial and condition using absolute error. The first of the linear models was a '2-regression' model, based on the last two observable data points per trial. This was generated by applying linear regression to the second and third onscreen points in each trial. The second linear model was a '3-regression' model which incorporated all three observable stimuli and was generated by regressing on all three onscreen points per trial. The third model was a '3-mean' model which was included as a measure of potential non-linearity of forecasts. This was created by averaging all three observed stimuli per trial. Both the 3-regression and 3-mean models use all of the information disclosed per trial to formulate their predictions. They were therefore included in the analysis for completeness and to provide comparison to the hypothesised method of human forecasting which involved use of only the last two observed data points per trial. (Full details of how the models were generated can be viewed in appendix 2).

The degree to which participants' predictions fit the 2 and 3-regression models thus indicates the degree of forecast linearity as dependent on the use of either all three data points, or only the last two observable points per trial. Predictions which are reflective of the 3-mean model suggest people are forming judgments by linearly extrapolating using all the available data. However, predictions which are best fit to the 2-regression model indicate that people are applying a more streamlined version of the linear heuristic strategy, based on use of minimum data points (i.e., only the last two observations of the three stimuli points per trial).

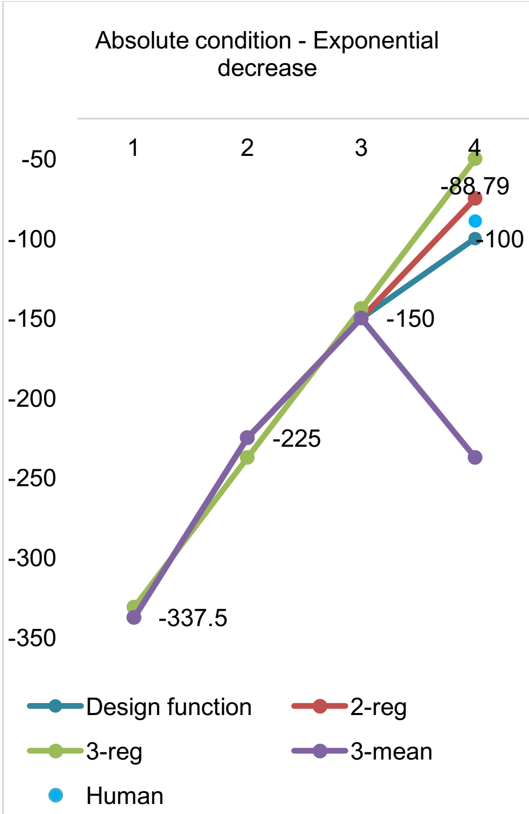
Performance which aligns with the 3-mean model on the other hand, indicates the tendency to apply non-linear extrapolation, derived from averaging the three observed data points per trial. It may be appropriate to apply this strategy in noisy environments or where cyclical trends exist such as in seasonally varying data. In the context of the experimental trials, it may therefore be less rationale to adopt an averaging forecasting strategy, as the presentation of a short three-point sequence in the absence of additional cues does not indicate high variance or the presence of a cyclical trend. Thus, the likelihood is that people will be most inclined to make errors in processing the percentage data, leading them to extract and project linear trends. This will hinder the capability to correctly identify and project exponential growth.

It is therefore expected that participants will employ minimal cues and make predictions which will be best fit to the 2-regression model based on the fast, frugal arithmetic processing of the last two data points observed onscreen per trial. Treating percentage points as absolute numbers will lead to adding or subtracting values even when inappropriate which will promote the propensity to forecast linearly based on the arithmetic difference between the last two observations in each sequence. This method is more likely to occur that a 3-regression forecasting method which increases the cognitive demands associated with processing more data points, thus making for a less efficient strategy. The results of the model fits are shown in figure 3.8 which display participants' forecasts in relation to each model prediction in the absolute condition trials (panels A - D), and in the percentage condition trials (panels E - H).

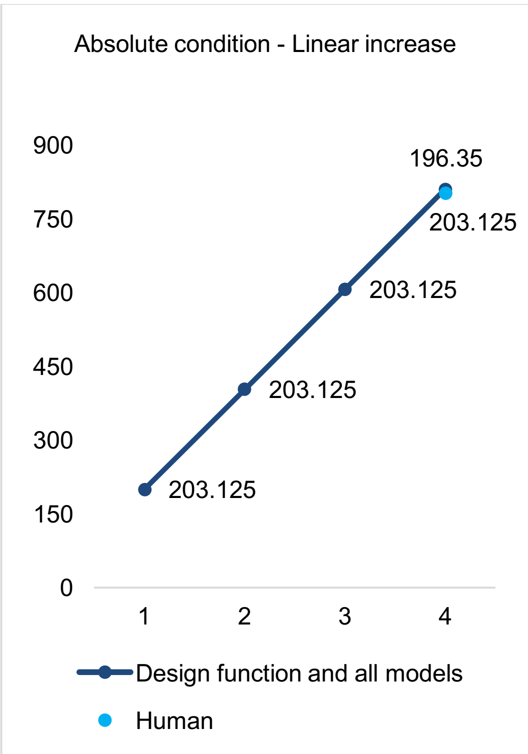
Figure 3.8 (A – D) The four panels below show the human and model predictions for absolute condition mismatched exponential increase (A) and exponential decrease trial (B), the matched linear increase (C) and linear decrease trial (D).



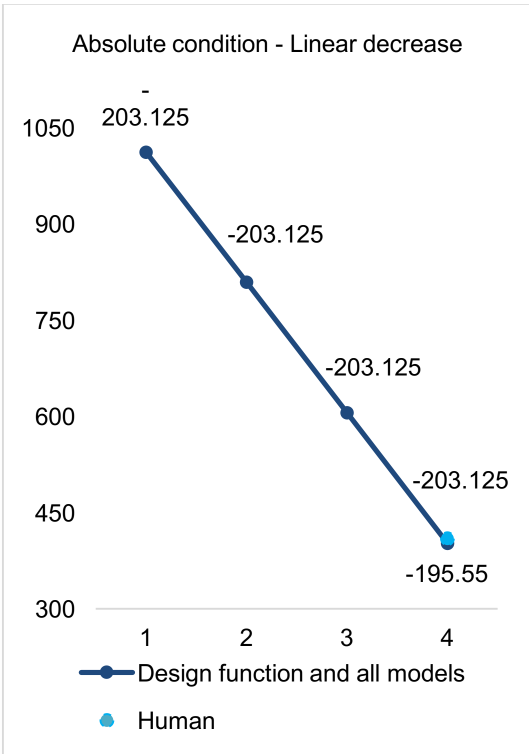
A (Exponential increase)



B (Exponential decrease)

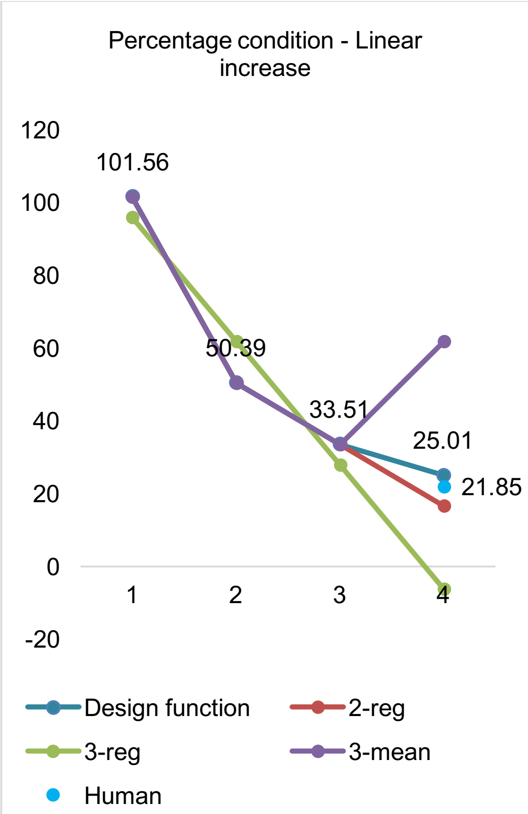


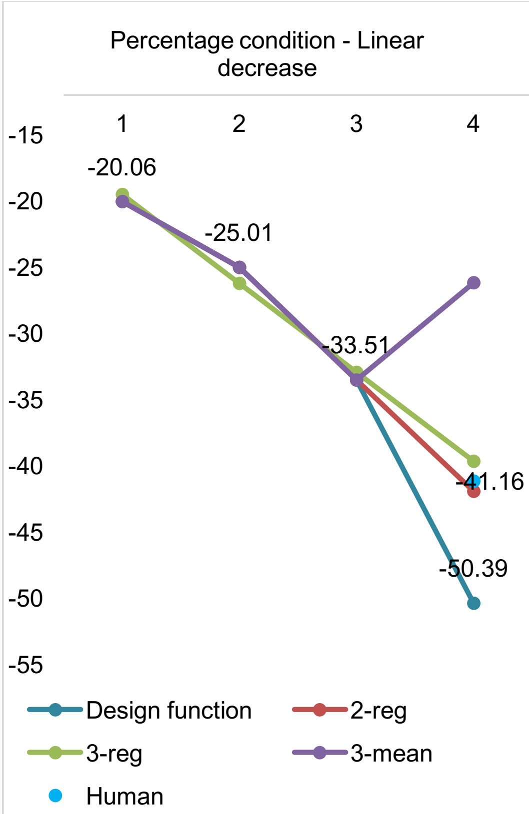
C (Linear increase)



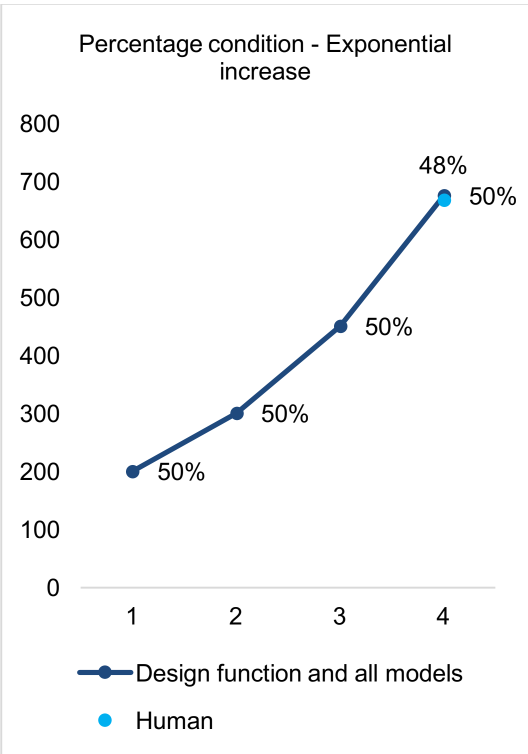
D (Linear decrease)

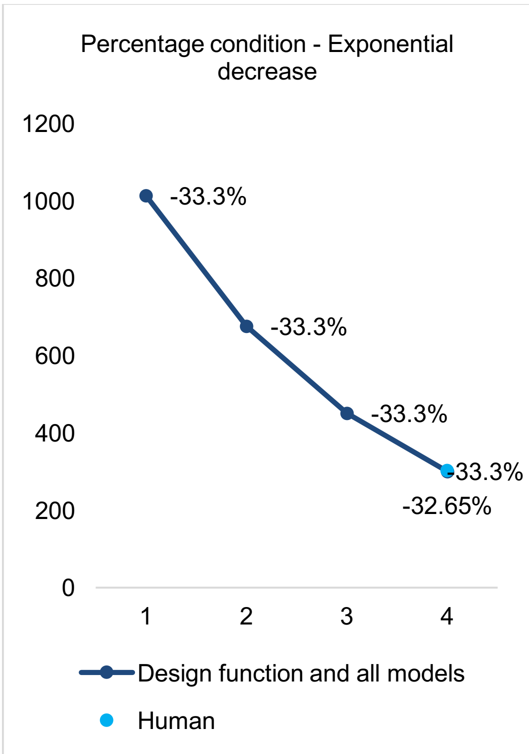
Figure 3.8 (E – H) The four panels below show the human and model predictions for the percentage condition mismatched linear increase (E) and linear decrease trial (F) and the matched exponential increase (G) and exponential decrease trial (H).





E (Linear increase) *F (Linear decrease)*





G (Exponential increase) *H (Exponential decrease)*

Figure 3.8 Human Performance in Relation to Formal Models in Exp 1

Participants' forecasts in relation to the 2-regression, 3-regression, 3-mean model and design function fourth point predictions per trial in the absolute condition (A) and percentage condition (B) with 95% within-participant confidence intervals. The first three data points marked on the design function in each plot are the stimuli points *observed* by participants in each trial. The data point marked on the dashed line in each plot is the participants mean fourth point forecast for that trial. In the mismatched trials in the absolute condition (i.e., the exponential increase and decrease trials) and the mismatched trials in the percentage condition (i.e., the linear increase and decrease trials), the design function reflects the actual stimuli observed in the trials. Whereas in the matched trials in both conditions, the design function reflects the *hidden base rate* changes in absolute values, and the values marked on the design function in these trials are the *observed stimuli* per trial. (Only the design function is shown in the matched trials because each observed stimuli value was the same per sequence which meant that each of the model fourth point predictions were also the same as the design function).

As shown in figure 3.8, human accuracy in the matched trials was fairly high in both conditions, based on very similar predictions to the design function (and model) predictions in each case. In the mismatched trials however, human judgments were most similar to the 2-regression model in both conditions. This suggests that participants forecast trials across absolute and percentage format conditions using a method akin to linearly extrapolating based on the last two observed data points per trial.

To determine whether the 2-regression model could account for participants' method of prediction formation across all the trial types (i.e., in both matched and mismatched trials), the 2-regression, 3-regression and 3-mean model predictions were fit to human forecasts using absolute error per participant, per trial and condition. This yielded 24 individual error values per participant (i.e., 8 trials x 3 models). The errors were then averaged per participant, per condition to give 131 error values for each model per condition. Figure 3.9 shows the mean absolute error for all trials per model and condition.

A two-way repeated measures ANOVA conducted on the mean absolute forecasting error per model showed a significant main effect for condition, $F(1,774) = 66.89$, $p<.001$, a main effect for model $F(2,774) = 168.92$, $p<.001$, and a significant interaction between condition and model $F(2,774) = 14.91$, $p<.001$ yielded by the significantly heightened mean error for the 3-regression and 3-mean models in the percentage condition. Pairwise comparisons with Bonferroni corrections confirmed that mean error was higher across all models in the percentage condition (M=63.52) compared to the absolute condition (M=41.70), $p<.001$, and across both conditions, mean error for the 2-regression model was significantly smaller (M=21.46) compared to the 3-regression (M=46.78) and 3-mean models (M=89.63), $p<.001$. Thus, based on least absolute error, the 2-regression model was shown to best describe participants' predictions across each trial type.

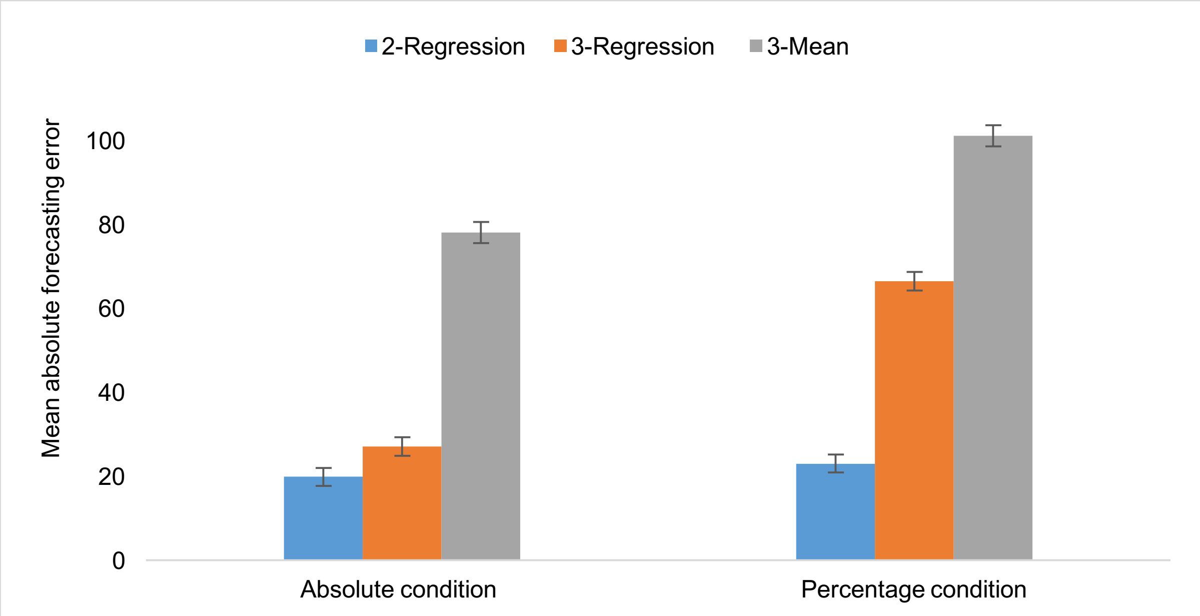


Figure 3.9 Mean Absolute Error in Relation to Formal Models in Exp 1

Mean absolute forecasting error per participant per trial and condition for the 2-regression, 3-regression and 3-mean model with 95% within-participant confidence intervals of the mean. Across trials in each condition, the smallest error occurred for the 2-regression model which indicated that participants forecast in a way most akin to linearly extrapolating using the last two observed data points regardless of the numerical format or trend function.

Further statistical verification of the best model fit across trials and conditions was provided by calculating the frequencies of least error counts for participants' forecasts across trials for each model. (The results of the frequency counts are shown in figure 3.10). A chi-square analysis conducted on the frequencies per model indicated that the 2-regression model achieved the highest frequency of least error counts across conditions $X^2(2, 262) = 6.67$, $p<.05$, confirming the expectation that estimates were formed in accordance with the tendency to predict linearly based on additively processing the last two data points observed onscreen in each trial.

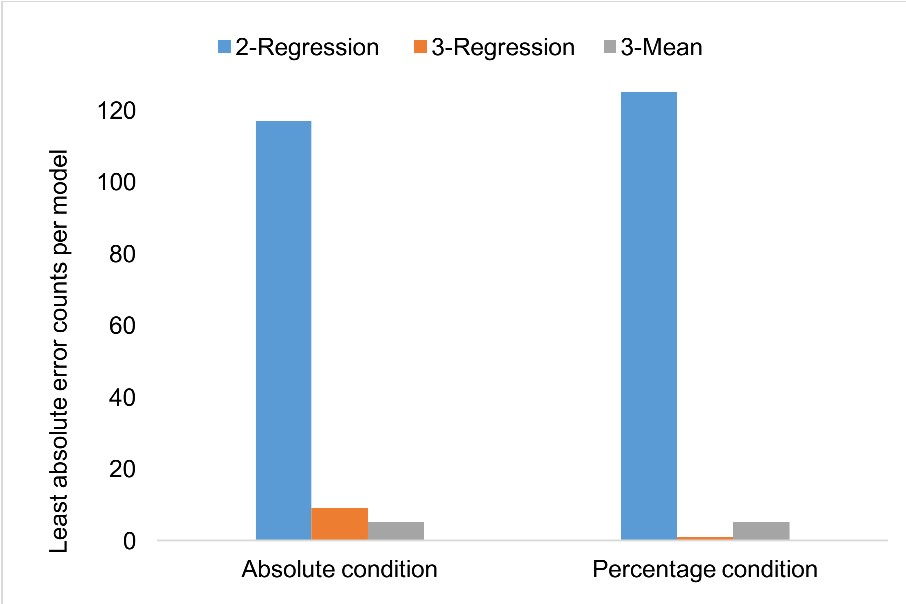


Figure 3.10 Least Absolute Error Counts Per Model in Exp 1

Least absolute error counts per participant, trial and condition for the 2-regression, 3-regression and 3-mean model. As shown, the 2-regression model achieved a significantly higher frequency of east error counts across trials in both conditions, supporting the hypothesis that experts would display the propensity to form linear forecasts when making judgments based on data presented in numeric formats, regardless of whether the data are trending linearly or exponentially.

Supplementary Analysis of Linearity

Although the 2-regression model was shown to be the best characterization of human forecasting estimates in terms of formal models, it was also possible that peoples' estimates may have followed a curve which was not completely linear, but involved a degree of non-linearity which was not undetected by the formal model fitting procedure. To explore this possibility, Shapiro-Wilks tests were conducted on the distribution of error in relation to the design function fourth point. The tests yielded significant results ($p<.05$) for each trial type which as suspected, indicated a degree of non-linearity in estimates across trials.

It is thus concluded that people exhibit a strong linear tendency in forecasting estimates which is best explained in terms of a 2-regression based on the last two points observed per trial. However, on examination of performance in terms of the distribution or error in relation to the target fourth point, it is apparent that judgments may not be strictly linear in nature and that people are likely to show some sensitivity to the acceleration of the trend, even though it is underused. Follow up studies are thus necessary to further explore the nature of the non-linearity in peoples' estimates and the factors contributing to subtle differences in the degrees of judgment linearity which may not be accurately captured with formal prediction models.

3.2.3 Discussion

The findings confirmed expectations, showing that despite expert domain knowledge, professional retail forecasters were insensitive to differences in functions and exhibited a consistent tendency to form linear predictions when presented with percentage information and non-linear growth trends. The results support the robustness of the linear prediction heuristic based on the propensity to process numeric values simultaneously, using arithmetic operations to form estimates of future events whether appropriate or not. Retail forecasters were shown to predict future outcomes for numbers in both linear and exponential trends by using the absolute difference between the last two observations per trial. This suggests that the bias occurred in the interpretation of the numeric format. The observed values in each trial were processed as absolute numbers, hence the same 2-regression method was applied even when a percentage sign was shown.

These results dovetail with wider literature relating to percentage format biases in which both experts and novices are shown to display a robust tendency to additively process percentage points where multiplicative operations are required. As outlined in review in chapter 2, this strong bias for arithmetic processing and linear trends is associated with inaccurate numerical estimates and predictions across many professional and consumer choice domains (e.g., Larrick & Soll, 2008; Stango & Zinman, 2009; Newall & Love, 2015). The results of experiment 1 therefore support the notion that the tendency to assume linearity in the environment is derived from the underlying bias to count and compute values in concrete terms, assuming absolute numbers with non-normalized base rates.

Compared to multiplicative computations, adding and subtracting numbers arithmetically provides a quicker and easier method of processing and is thus favoured as a more efficient heuristic strategy, particularly in complex domains. In this view, on-the-spot arithmetic calculations provide the premise for predictions which follow linear patterns when extrapolated into the future. The ability to perceive non-linear curves and accurately project exponential trends conversely, is dependent on comprehending sequential numerical processing. Thus, it is likely that people are primed to seek and project linear trends in the environment because they are congruent with underlying cognitive propensities to apply heuristic strategies which align with intuitive cognitive processing methods.

The reliance on a 2-point linear extrapolation strategy created a trend-damping and anti-damping effect in which increases were underestimated and decreases were overestimated. These effects were amplified in the trials in which the observed number format mismatched the trend function. This resulted in the judgments in the matched trials being significantly more accurate than those in the mismatched trials, because relying on the absolute difference between the last two observations when the values are the same will yield an accurate prediction of the fourth point. Conversely, applying this method to observed values which differ will lead to inaccurate judgments. Thus, it is apparent that the understanding of numerical functions in forecasting contexts of this kind is likely to be distinct or secondary to the interpretation of the numerical format. It is the biased interpretation of the format which is thus likely to be the key determinant of peoples' judgment performance.

This study highlights the importance of numerical format in forecasting judgment, indicating the problems which arise from the incongruence between the human propensity to additively process data and identify linear relations, and single point percentage data and non-linear extrapolations of statistical forecasting models in complex forecasting environments. These findings are salient in that they have been demonstrated among experts possessing specialist training, knowledge and experience in their field. It is therefore likely that the same tendencies and numerical format biases will also characterise the judgment processes of novices and consumers who do not possess any background knowledge or forecasting skill and experience.

Although the task demands and level of complexity involved in the experiment were representative of the experts' professional forecasting role, it is possible that the absence of additional contextual information impacted their judgmental performance. This possibility suggests that forecasting decision making in certain domains may be associated less with statistical rationality and ability to accurately interpret trends based on statistical models, and more with the application of tacit domain knowledge and un-modelled data sources.

Due to economic and practical constrains associated with conducting research in the real-world commercial setting of the Supermarket headquarters, it was not possible to continue experimentation with this particular group of retail forecasters. Ideally, the study would have been extended with the examination of other factors relevant to expert performance in addition to the interpretation formats and functions. For example, assessing other aspects of knowledge and factors important to performance such as promotional and seasonal effects and supplier relations would increase our understanding of 'expertise' and how it develops in such settings. Furthering the analysis of how retailers synthesize these 'unmodelled' domain specific features with statistical model output would thus provide a more comprehensive understanding of 'expertise' in human forecasting which would be relevant to other modern complex statistical environments.

To continue with the investigation of expert judgment in complex real-world settings, chapter 4 extends exploration to the field of humanitarian aid by focusing on the importance of contextual information and the effects of additional cues on judgment performance. In experiment 2, experts in the aid field and novices without forecasting experience are assessed in their forecasting performance for real-world refugee camp data. The emphasis is on the effects of varied informational cues in rich data environments and how different levels of complexity influence patterns in peoples' predictions and the propensity to extract linear trends in noise.

Chapter 4

Rich Data Contexts & Phantom Trend Forecasting in Humanitarian Aid Experts

In this chapter, humanitarian aid experts are compared to non-experts in their ability to forecast real-world refugee camp data. The effect of noise is tested by presenting target time series data in sparse versus rich contexts. Experts' predictions are no more accurate than novices and both groups formed judgments by linearly extrapolating the data in both contexts. The tendency to linearly extrapolate increased in noise, particularly so among experts and correlated with forecasting error. Error in noise was found to be associated with projecting 'common' trends (i.e., when all cues trended in the same direction). This suggests that the linear prediction bias is reinforced in more varied conditions, based on visual heuristic strategies employed to identify trend congruencies.

4.1 Background and Rationale

In the previous chapter, the examination of expert judgment involved assessing professional retailers' ability to interpret numeric data and compute estimates based on values presented in alternating frames and growth functions. From this perspective, performance was assessed in terms of participants' numerical skill and ability to interpret the data made immediately available within the experimental setting. It could therefore be argued that the trial stimuli and design may have influenced participants' responses, or yielded estimates which may not have accurately reflected the retailers' true forecasting abilities or propensities.

However, the nature of the data presented in the trials closely resembled that of the numeric information used by the retailers on a daily basis. Specifically, the task involved forming stock and sales forecasts by combining percentage points and absolute figures across multiple forecasting tools and systems. The trial stimuli therefore presented the same judgment task demands involved in the retailers' every-day forecasting role, thus generating an accurate assessment of what characterized expertise in this particular environment.

To explore the characteristics and importance of ‘expertise’ in other professional contexts, experiment 2 examines the more ‘intuitive’ and naturalistic judgment processes involved in the field of humanitarian aid. The forecasting performance of expert aid workers and novices is assessed when presented with sparse context data versus rich context trials which involve additional non-causal morbidity indicators to test how experts react and what that may indicate regarding the application of domain knowledge and intuitive judgment processes. Judgement ‘rationality’ in the aid setting is thus assessed in terms of the aid workers’ ability to accurately interpret the cyclical nature of the widely recognized target data trend and to project the trend independently of the rich context cues which are known among aid workers to be non-causally related to the target series.

In contrast to the professional retail domain in which the accurate interpretation of the effects of functions and formats is important to judgment performance, expertise in the aid setting is dependent on the ability to combine rich knowledge of unmodelled variables with statistical models of morbidity rates to formulate supply and demand estimates. Such tacit knowledge involves understanding how different indicators impact upon on one-another in refugee camp situations for example, and how seasonal factors can influence morbidity rates in specific regions.

Thus, where the professional retailers’ judgment is dependent more on the numeric data in the immediate task environment, expert performance in the humanitarian aid environment is shaped to a larger extent by their application of knowledge and experience of wider factors beyond the data frame. The experiment examines how non-causality related morbidity cues are handled in ‘noisy’ context trials in contrast with sparse data trials to investigate how experts apply knowledge of cues and seasonal variability to data presented in particular formats and frames. For example, ‘intuitive’ judgment processes or even optimistic tendencies may come into play which lead experts to make conservative estimates of morbidity rates or to assume that additional cues presented in the data environment are important to the target data in some way, despite the knowledge that they are unrelated in the real-world.

In situations of uncertainty, experts are often relied upon to make critical judgments. For example, expert aid workers forecast refugee counts in order to match camp supplies with needs. In highly variable environments, one danger is that people may detect trends when in reality the fluctuations arise from noise. Forecasting based on these phantom trends (e.g., linearly extrapolating when noise governs the system), leads to greater error than baseline models which forecast using the mean of past observations or the previous point. We evaluate whether aid workers show this pattern of linear bias within their expert domain. Expert aid workers' forecasts were no better than novices, and experts showed stronger linear extrapolation biases in richer (i.e., more realistic) contexts. Findings suggest that the harmful tendency to find meaning in noise may be accentuated by expert domain knowledge.

We rely on experts to accurately assess the status of real-world events and make consequential forecasts using rich, domain specific knowledge (Philips, Klein & Sieck, 2004). Within their field, experts' breadth and depth of knowledge and appreciation of contextual information can increase forecasting performance (Armstrong, 1985; Lawrence et al., 2006), facilitate integration of data cues (Einhorn & Hogarth, 1975; Hoch & Schkade, 1996), and assist in understanding relationships among predictor variables (Seifert & Hadida, 2013). In a retail environment for example, the more familiar experts are with products and contextual factors (such as seasonal differences in demand), the better their forecasting accuracy (Edmundson, Lawrence & O'Connor, 1988; Seifert et al., 2015) and adjustments to statistical forecasts (Goodwin & Fildes, 1999; Sanders & Ritzman, 2001).

Despite domain knowledge, there are however a number of settings in which experts' judgmental forecasts are found to be no more accurate than novices (Armstrong, 1980, 1991; Lawrence & O'Connor, 1993). Indeed, those with greater expertise have been shown to perform worse in the financial sector (Önkal & Muradoğlu, 1994; Wilkie-Thomson, Onkal-Atay, & Pollock, 1997). The high volatility of this domain is considered to make forecasting particularly difficult (Kasa, 1992; Makridakis & Taleb, 2009), potentially leading experts to chase trends which may simply be noise.

What explains these contrasting results, wherein experts surpass novices in some situations but in others are no better? There are likely multiple complementary explanations. Here, we test one hypothesis, namely that experts in noisy domains feel licensed by their rich contextual knowledge to extract patterns from time series that in reality are nothing more than noise. This possibility is consistent with prior work and makes an additional prediction: The richer the context, the stronger experts' propensity to infer trends. To foreshadow, we test this hypothesis with expert aid workers forecasting morbidity indicators commonly encountered in refugee camps throughout the world. In this complex, highly variable domain, we found that experts' forecasts were no better than novices and their tendency to extract misleading trends was heightened by richer contexts (i.e., when they were provided with more information about the camp situation).

Findings relating to how people learn predictive relationships between stimulus and response variables show that people possess a strong tendency to apply linear functions when forming judgments under uncertainty (DeLosh, Busemeyer, & McDaniel, 1997). For example, linear relations are assumed between everyday variables such as body weight and height, price and quantity, drinks consumed and blood alcohol level, etc. The tendency to extract linear patterns also commonly extends to domains where it is inappropriate such as retail forecasting (Love & Parker, in preparation), and in financial transactions where costs and benefits compound over time (McKenzie & Liersch, 2011).

Another example of the strong human propensity to form linear inferences is shown by Hohle and Teigen (2015) who identify a 'trend effect' in which novices are shown to linearly project experts' forecasts for natural events by extending the trend direction of past predictions based on two historical data points. The propensity to linearly extend current trends is shown to create significant over and underestimates of the risk of natural events compared to experts' prognosis. Overall, the most successful models of how people learn and generalize functional relations have a strong bias to extract linear patterns from observations (DeLosh et al., 1997; Kalish, Lewandowsky & Kruschke, 2004). On this basis, we expect to see a tendency among experts and novices to detect and project linear trends in richer contexts (i.e., where camp data variability is increased).

Finding linear trends in observations that arise from noise will lead to greater forecasting error. This conclusion readily follows from basic concepts in statistics and machine learning, such as overfitting and the bias-variance trade-off (e.g., Gemen, Bienenstock & Doursat, 1992). Consider a time series that is pure noise (i.e., observations are generated by a normal distribution). The best guess for the next observation is simply the mean of the past observations. If instead a linear trend model were fit to the time series, the idiosyncratic nature of the past observations (i.e., the training sample variance) would lead to poor prediction for the next observation because the inferred trend would diverge from the mean of the observed items (which is the best estimate of the next observation).

People's tendency to treat randomness as signal and extract phantom trends from noise (Harvey, 1995; Harvey, Ewart & West, 1997; Lopes & Oden, 1987) leads to judgment inaccuracy in noisy environments (Brehmer, 1978; Kahneman & Tversky, 1973) which increases with complexity (Lee & Yates, 1992). Our contention is that providing additional context (evoking a complex real-world scenario) will increase experts' tendency to find trends in observations. In support of this notion, providing people with additional cues (akin to context) in complex prediction tasks reduces performance (Harvey, Bolger & McClelland, 1994). These observations dovetail with the view that unadjusted expert forecasts are least useful in domains that are rich, complex, and uncertain (Green & Armstrong, 2007a, 2007b; Tetlock, 2005).

In sum, forecasting morbidities in a humanitarian crisis is a rich, complex, and uncertain domain. In this study, experts and novices forecasted rates of refugee camp morbidities after viewing a time series of real-world observations from refugee camps. It is necessary to note that traditional forecasting research tends to refer to 'noise' as random error around an observation in a time series. However, in the following experiment, we make reference to noise as additional time series cues which are non-causally related to the target time series (i.e., added data points which act to increase the amount of data observed, but have no impact on the target series). We manipulated how much real-world context was supplied on each trial and predicted that *rich* ('noisy') contexts with additional non-causal cues (as opposed to *sparse* contexts) would increase experts' tendency to infer linear trends, leading to forecasts no better than those of novices.

4.2 Experiment 2

4.2.1 Method

Participants

Participants consisted of 30 expert humanitarian aid workers (mean age = 32.42 years, *SD* = 8.53; 17 males, 13 females) and 36 novices (mean age = 27.31 years, *SD* = 4.8; 25 males and 11 females) recruited using a random sampling technique via LinkedIn (an online networking site for professionals, groups and organizations to liaise, seek jobs and share content). The sample size was determined by the time frame available for data collection. A three-week period was set and within this window as much data as possible was collected. The expert humanitarian aid workers all possessed varying degrees of professional experience in monitoring and forecasting morbidity trends (mean = 6.8 years, *SD* = 7.5), and 18 of them had direct experience of working in refugee camps. The novices each possessed an undergraduate or postgraduate degree in fields unrelated to forecasting or humanitarian aid work. For compensation, participants chose between the chance to win a 25 Euro Amazon voucher, to make a 30 Euro donation to charity, or to select neither.

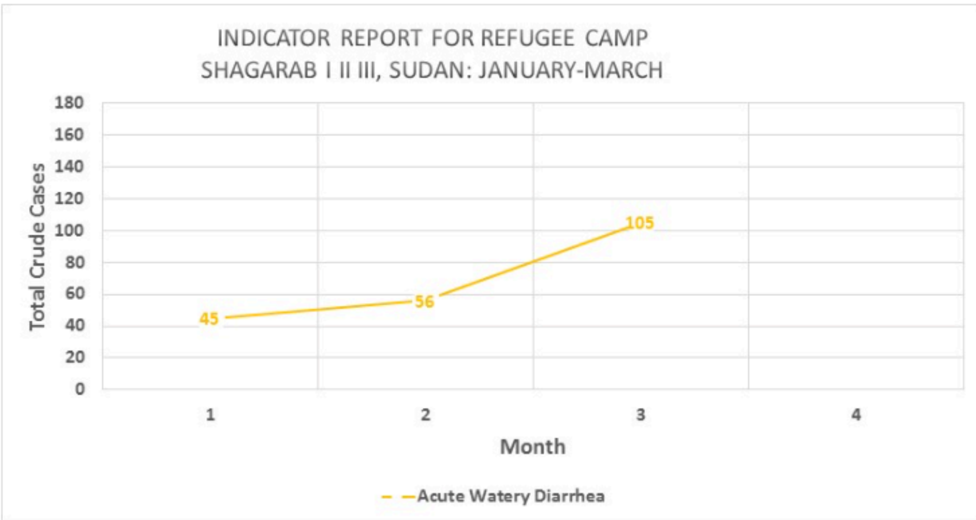
Design

Participants completed two blocks of 15 trials each involving an individual time series of three data points. Each data point related to the total number of refugee morbidities for one month, and the points were for three consecutive months. One block contained 15 trials in a sparse context and the other block contained 15 trials in a rich context and participants were tasked with forecasting the fourth month in every trial. The individual trials in each block were repeated across the sparse and rich context conditions, such that the same 15 individual time series of morbidity counts for three consecutive months were shown in both blocks. The order of blocks was counterbalanced across participants and the order of trials within each block was randomized for each participant.

The three-month time series stimuli used in each of the 15 trial sequences, as well as the surrounding context time series for the rich context trials (see Procedure for details), were randomly drawn from a sample of Health Information Systems (HIS) Detailed Indicator Reports downloaded from the United Nations High Commissioner for Refugees (UNHCR) Twine Project website. The downloaded pdf reports and exacted time series used in the study are included in the online repository of materials for this contribution, osf.io/dn3xy.

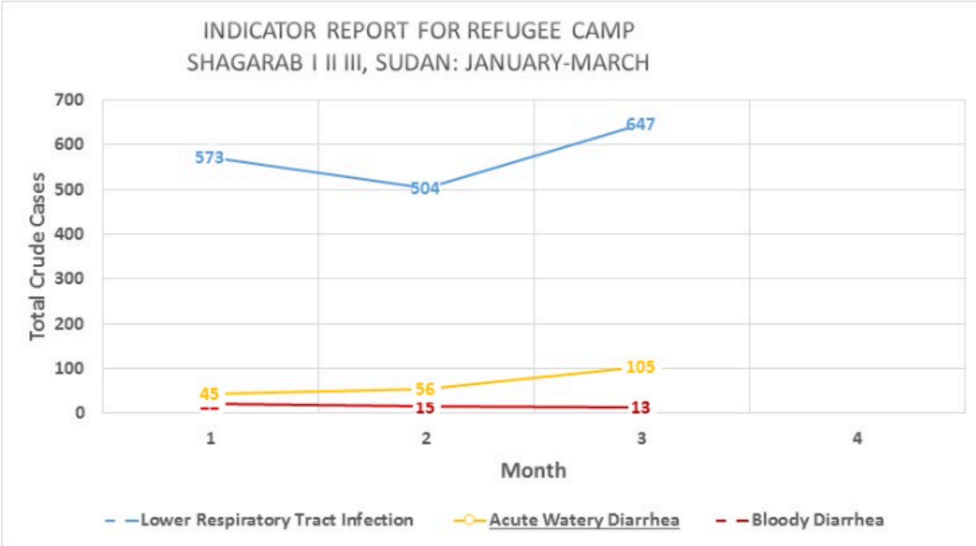
Materials and Procedure

On each trial, an individual time series consisting of three data points which represented total rates of ‘acute watery diarrhoea’ (AWD) per month for three consecutive months was shown. Participants forecast the value for the fourth month (i.e., the next data point) by entering an integer into a response box (see Figure 4.1). In the rich context condition, participants were also shown surrounding contextual information about other common camp indicators, namely ‘lower respiratory tract infection’ (LRTI) and ‘bloody diarrhoea’ (BD) (see Figure 4.1B). Each trial was response terminated and participants were given a brief break at the midway point in between the two trial blocks.



Using the relevant information presented, predict the number of Acute Watery Diarrhea cases for month 4:

Figure 4.1 (A) An example of the sparse context trial in which participants observed a single time series cue



Using the relevant information presented, predict the number of Acute Watery Diarrhea cases for month 4:

Figure 4.1 (B) An example of a rich context trial in which the same time series observed in the sparse trial was shown but with the addition of two non-causal time series cues

Figure 4.1 The Sparse and Rich Context Trials in Exp 2

An example of a sparse context trial (A) and a rich context trial (B). Participants were shown a time series consisting of three points and predicted the fourth point. On rich context trials, additional surrounding context was provided, modelled on reports expert aid workers receive on refugee public health in camp situations.

4.2.2 Results

On each trial, participants observed a time series consisting of three points and made a forecast for the fourth point. The main dependent measure was scaled error, which was calculated per participant per condition as follows in Equation (1):

$$\frac{\sum_{i=1}^{n}|Y_i - F_i|}{\sum_{i=1}^{n}|Y_i - Y_i'|} \tag{1}$$

where Y_i is the true value for the fourth point, Y_i' is the third point, F_i is the participant's forecast for the fourth point, and n is the number of trials, which was 15.

The numerator in Equation (1) is the sum of absolute forecasting error across trials, which is normalized by the denominator that is the sum of absolute forecasting error for a *naïve model* which uses the third point as the forecast for the fourth point. Scaled error is used because it is scale independent and easy to interpret (Hyndman & Koehler, 2006). In particular, a scaled error of less than 1 indicates that forecasts are more accurate than the naive model, whereas those greater than 1 are worse than the naive model's forecasts.

Analysis of Decision Environment

To evaluate our characterization of the task environment, we applied two models to the AWD time series. A *linear trend model,* which uses the best linear fit to the first three points to forecast the fourth point, and an *aggregation model,* which uses the mean of the first three points to forecast the fourth, were applied to the AWD time series. We hypothesized that the environment is noisy, and therefore the linear trend model would perform worst. As predicted, the linear trend model had significantly more error (M=1.23) than the aggregation model (M=0.96), $t(57.76) = 2.82$, $p<.01$, Cohen's $d=0.73$. Many trends detected by the linear trend model were likely phantom trends, whereas the aggregation model can average over noise.

Human Forecasting Error

The scaled forecasting error for novices and experts by trial type is shown in Figure 4.2 Forecasting error across groups was larger on rich context trials (M=1.30) compared to on sparse context trials (M=1.21), $F(1, 64) = 5.68$, $p<.05$, $\eta^2=0.85$. There was no significant effect of group, $F(1, 64) = 0.05$, $p=0.819$, $\eta^2=0.03$, nor a significant interaction between group and condition, $F(1, 64) = 0.79$, $p=0.376$, $\eta^2=0.12$. Planned t-tests indicated a significant difference in the forecasting error between the rich (M=1.33) and sparse contexts (M=1.20) for the experts, $t(29) = 2.10$, $p<.05$, $d=0.4$. However, for novices, no significant difference in error between rich and sparse contexts was found (M=1.28 vs. M=1.22), $t(35) = 1.24$, $p=0.221$, $d=0.17$.

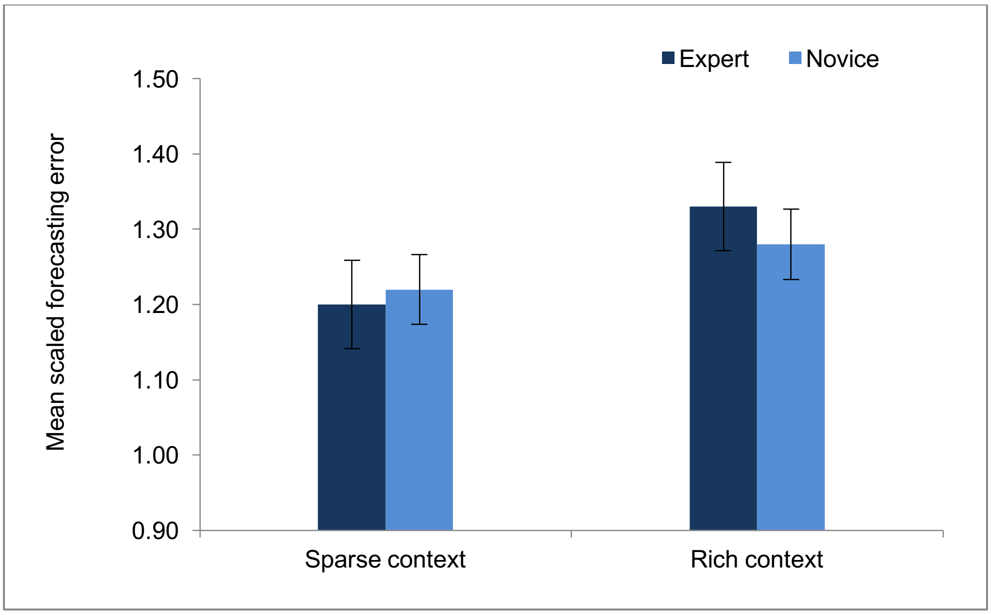


Figure 4.2 Mean Forecasting Error Per Condition in Exp 2

Forecasting error for the two populations and trial types are shown. As predicted, participants performed worse in rich contexts. Error bars are 95% within-participant confidence intervals of the mean.

Model-Based Analyses

The preceding analyses indicate that participants' performance levels were more akin to the linear trend model than to the aggregation model. To evaluate whether participants' forecasts were more congruent with the linear trend model than the aggregation model, we calculated the proportion of trials in which a participant's forecasts were closer to that of the linear trend model than the aggregation model. As shown in Figure 4.3, the proportions of match were above 0.5 in both contexts. As this finding indicates, both experts' and novices' forecasts were significantly more akin to the linear trend model than the aggregation model in the sparse context (M=0.57), $t(65) = 3.89$, $p<.001$, $d=0.95$ and also the rich context (M =0.60), $t(65) = 5.94$, $p<.001$, $d=1.46$. No statistically significant differences were found in expert versus novice compliance with the linear trend model, $F(1, 64) = 0.09$, $p=0.766$, $\eta^2=0.07$, nor a significant difference between sparse and rich contexts, $F(1, 64) = 2.81$, $p=0.09$, $\eta^2=0.69$, nor a significant interaction between group and context, $F(1, 64) = 0.99$, $p=0.32$, $\eta^2=0.24$.

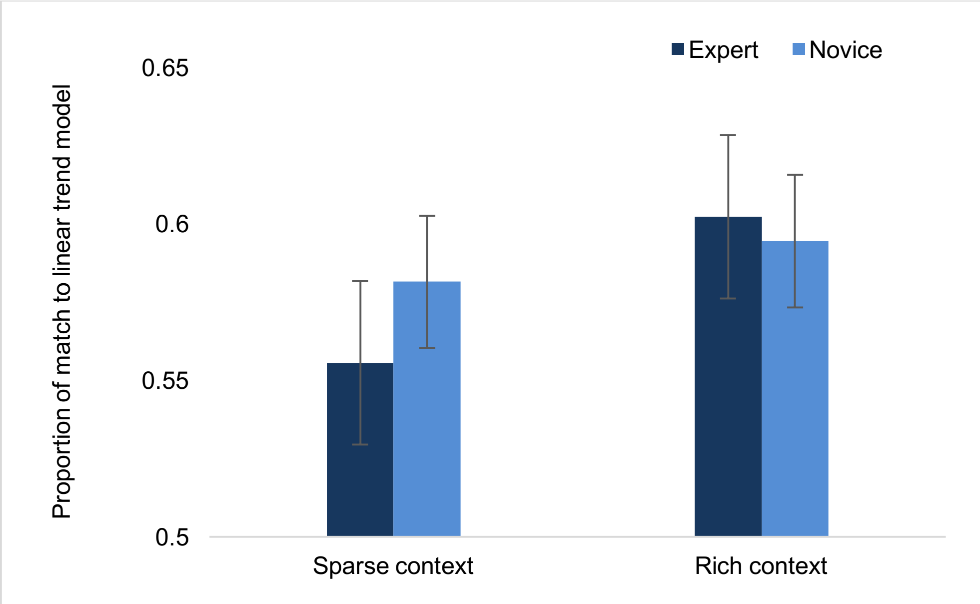


Figure 4.3 Judgments Fit by the Linear and Aggregation Model in Exp 2

Proportion of judgments which were better fit by the linear trend model compared to the aggregation model per group and trial type with 95% within-participant confidence intervals of the mean.

One interesting possibility was that the tendency to chase linear trends led to increased forecasting error. Consistent with this notion, a regression model showed that the proportion of judgments best fit by the linear trend model was predictive of scaled forecasting error across conditions when condition and participant group were included as factors in the model, $R^2 = .06$, $F(3,128) = 3.03$, p=.031, $B = 0.54$, $t(128) = 2.58$, $p<0.05$.

Human Tendency to Trend-Damp

In addition to the model-based analysis, participants' tendency to trend-damp the real-world AWD trend (i.e., draw towards the series mean) was examined per group and condition by comparing forecasts to the AWD fourth point per trial and condition. Experts were shown to trend-damped the AWD sequences in 57% of trials. A one-sampled t-test indicated that the proportion of experts' underestimates was significantly greater than 0.5 across conditions, $t(899) = 4.30$, $p<.001$. Novices trend-damped the AWD sequence in 56% of all trials which was also shown to significantly differ to 0.5 across conditions, $t(1079)$, = 3.92, $p<.001$. Moreover, experts were found to trend-damp in a significantly larger proportion of trials in the sparse (59%) compared to the rich context (56%) $t(407.68) = -21.57$, $p<.001$. A higher rate of trend-damping in the sparse (57%) compared to the rich context (55%) was also shown for the novices $t(501.42)$ = -

28.44, $p<.001$. A repeated measures ANOVA indicated that there was no overall difference between experts and novices in the tendency to trend-damp, $F(1,64) = 0.14$, $p=0.70$, or between either group in either condition, $F(1,65) = 3.54$, $p=0.06$.

It is important to note that the propensity to forecast towards the mean in this particular setting may have reflected an optimism bias whereby people tended to make more 'hopeful', and therefore more conservative estimates of morbidity rates.

Rich Context Trend Directions

Following completion of the above analysis, a potential limitation in the experimental design was identified which may have influenced the results. It was noted that the presentation of the data on different scales in each condition created a difference in the shapes of the curves of each stimuli sequence pairing between the sparse and rich context. For example, as shown in figure 4.1, the AWD sequence data presented on a scale of 0-700 in the rich context condition (B) generated a far flatter curve compared to that in the sparse context (A) in which the same data was presented on a scale of 0-180. Although the numerical data values for each month were shown on the time series in each trial, it was possible that this discrepancy in curve shape led people to make larger adjustments from the third point in the sparse context compared to the rich context.

Therefore, due to these scale differences, a follow-up analysis was conducted on the predictions in the rich context condition separately to those in the sparse context condition. The aim was to determine whether the trend direction of the additional cues in the rich context impacted forecast judgments and error. Participants' forecasts which increased from point 3 to 4 were coded as 1 and all else as 0 and entered into a logistic regression as the dependent variable. The predictor variables were entered as the number of cases in which all three additional cues (i.e., AWD, LRTI and BD) increased from point 2 to 3 (coded as 1 and 0 respectively), and participant type (humanitarian versus novice).

The cases in which all additional cues were increasing was shown to be predictive of participants' tendency to forecast an increase ($B = 0.74$, $z = 2.96$, $p<.01$, 95% CI [0.25, 1.24]), with no difference in the propensity between experts and novices ($B = -0.02$, $z = -0.15$, $p=0.87$, 95% CI [-0.29, 0.25]). A second logistic regression with participants' decreasing forecasts coded as 1 and all else as 0 also showed that the instances in which all three additional cues decreased from point 2 to 3 significantly predicted both humanitarians and novices' tendency to forecast decreases (B = 0.93, z = 3.48, p<.000, 95% CI [0.41, 1.47]). Again, there was no difference in the effect between participant groups (B = -0.14, z = -1.02, p=0.30, 95% CI [-0.42, 0.13]).

The propensity to project trends when all three were in alignment was then examined in relation to participants forecasting error. Absolute scaled error in the rich context trials was regressed on cases where all cues increased, all cues decreased, and participant type. The result showed that all cues increasing was significantly predictive of participants forecast error across rich context trials, R^2 = .008, F(2,987) = 5.23, p<.01, B = 0.44, t(987) = 3.20, p<.01, with no difference in the effect between experts and novices (B = -0.05, t(987) = -0.46, p=0.63). This suggests that the linearly of participants' forecasts may have been even more pronounced in the trials in which all cues were increasing, hence the significant effect on forecast error.

In sum, the results from the sparse context condition showed a linear 'trend' extrapolation effect which, congruent with the findings of Hohle and Teigen (2015), indicated that people interpret data as trended when given absolutely minimal evidence of existent trends (i.e., they projected linearly based on only a single three-point time series). However, when considering the effect of the rich context shown in the above secondary analysis, a new finding is delivered. Congruity between the trend direction of the target time series and that of the two additional series (noise) is shown to reinforce the propensity to linearly project the target series by continuing the trend in the same direction as the other cues. Moreover, the propensity to extend congruous trends is associated with increased forecasting error.

It is possible therefore, that rather than people over or under adjusting from point 3 to 4 in the rich context trials (due to the size of the data scale), the error in the rich context is associated with people complying with the trends of the different series. This suggests that visual analysis of trend directions in multi cues contexts may be more important to forecasting performance than cue causality. Thus to conclude, despite domain expertise, the bias to extend trends linearly is shown to be reinforced in complex varied data, based on the 'consensus' between trend directionalities of the cues observed.

4.2.3 Discussion

We trust in experts to deliver effective judgments under uncertainty using rich domain knowledge (Lawrence et al., 2006). In some situations, for example supply chain forecasting, experts' contextual knowledge increases judgmental performance (Armstrong, 1985; Edmundson, Lawrence & O'Connor, 1988; Seifert et al., 2015). However, in other situations, expertise is not shown to enhance forecasting accuracy, indeed sometimes leading to a reversal in performance (Önkal & Muradoğlu, 1994; Wilkie-Thomson, Önkal-Atay, & Pollock, 1997). Given the comprehensiveness of experts' knowledge, why do such disparities exist in the effectiveness of experts across different environments? The present study sought to answer this question by assessing the effects of sparse versus rich context on expert forecasting accuracy in the complex arena of humanitarian aid.

We predicted that experts would perform worse in rich contexts, which they did, mirroring novices. Similar to the results of Hohle and Teigen (2015), we found that participants showed a strong tendency to linearly extrapolate from past observations, which should lead to increased error in domains where there is a great deal of noise or there is no linear trend. As predicted, the strength of the tendency to linearly extrapolate positively predicted forecasting error across sparse and rich contexts among both experts and novices. The tendency to chase noise and linearly extrapolate was particularly strong among experts in richer contexts that are more akin to real-world environments. This tendency to linear extrapolate in noisy domains may partially explain why experts sometimes perform no better than novices. When examining the tendency to trend-damp, experts were significantly shown to underestimate increases in the target data fourth point (i.e., dampen the real-world trend). The proportion of the trials in which trend damping occurred was greater compared to those in which it did not and the tendency to trend-damp occurred to a significantly greater extent among experts compared to novices.

Based on a follow-up analysis of the effect of the rich context series on prediction performance, we present a new finding which shows that people linearly extend the trend of cues when all three are pointing in the same direction. Moreover, the propensity for error is greater when all three series are increasing. This suggests error in the rich context was associated with the propensity to seek trend alignment in noise and linearly extend the aligned trends, particularly when all cues are increasing, rather than simply linearly extrapolating the target series regardless of the directions of the other cues. It is likely therefore, that participants' judgments of increase were more linear than predictions of decrease, which is why error was shown to be associated with cases when the trends of all observed series were increasing.

There are several reasons for why experts showed a tendency to chase phantom trends despite domain knowledge. One hypothesis is that the linear prediction bias arises from experts' attempting to make sense of more complex, variable environments by assimilating cues and seeking patters which don't exist. Why we find experts more effective in some settings compared to others, could therefore depend on the degree of contextual complexity and uncertainty of the forecasting environment. Further examination of experts' judgmental processes in complex environments is necessary to determine the robustness of the linear prediction bias and how it may interact with domain knowledge. Future repetition of the study in a naturalistic camp setting would also further insights into experts' processing of data in rich contexts, and allay issues of validity associated with artificially replicating field environments (Lipshitz, 1993; Lipshitz, Klein, Orasanu & Salas, 2001).

One practical recommendation based on our findings, is to use experts to identify key factors and variables in complex prediction domains which are then used to build statistical forecasting models and systems, as opposed to relying on experts to form unaided judgments. For example, experts have proven effective system builders and data interpreters in a new approach to forecasting humanitarian aid demands ahead of climate-related disasters. Based on experts' interpretation of hazard data from previous aid responses, the approach is predicted to reduce response costs by up to 50% by releasing funds pre-event (Cousin, 2015).

Experts are also shown to be good at identifying important variables in weather simulations (e.g., barometric pressure) which are weighted by models to make actual predictions. Experts increase the predictive accuracy of weather simulations based on models of climatology by 10 to 25% (Lynch, 2006, 2008). Moreover, the judgmental improvements are shown to hold, despite decades of statistical model enhancements. The heightened model performance is attributed to experts' ability to better perform sanity checks than machine learning tools, and integrate historical experience and 'common sense' factors into the forecasts (Silver, 2013). Using experts in the appraisal of model forecasts may therefore be effective in minimizing the harmful tendency to follow false trends, whilst retaining some of the advantages of their detailed tacit knowledge.

In sum, our findings show a robust linear prediction bias in expert humanitarian aid workers in rich informational contexts akin to noisy real-word settings which predicts a decline in forecasting performance. Our results suggest that the tendency to find signal or meaning in noise is a robust phenomenon that is impervious to expert knowledge. Indeed, richer contexts which should elicit knowledge appear to make experts more vulnerable to finding patterns in noise, not less.

Chapter 5

Data Disclosures to De-Bias Optimistic Time-Cost Evaluations in Financial Decision Making

In this chapter, two experiments are conducted based on findings from chapters three and four to investigate the effects of financial information disclosures on loan product choices. Experiment 3a tests the effects of lessening informational cues and framing percentage data in absolute currency values in current versus future rates on judgment performance in an online price comparison environment. In chapter 3 the concept of framing data in absolute values was shown to lead to judgment improvements and is thus expected to increase financial choice in this setting. The Finding show that simultaneously displaying rate alternatives using a future rate default to anchor people on 'realistic' options increased the propensity to opt for more effective, high future rate choices (as opposed to low, 'optimistic' current rate choices).

Experiment 3b is conducted to verify the robustness of this framing effect by adding a disclaimer (akin to a real-world financial industry disclaimer) to each condition to test whether choice performance can be improved when attention and effort applied in the task of comparative analysis is increased by making the risk of rate variability more salient. Applying the disclaimer in the context of the standard industry disclosure thus tests whether people are capable of effectively using percentage and rate information when encouraged to increase attention and cognitive effort.

Results show that the framing manipulation identified in experiment 3a remains highly robust in 3b and that no difference was found in choice performance with the addition of the disclaimer in either the standard industry or framing manipulation conditions. These results held even when individual differences in optimism, financial literacy and numeracy were included as covariates. The fact that performance remained constant across conditions between experiments 3a and 3b indicates that the reframing of rates in absolute terms is key to judgment rationality, and that percentage format biases in complex environments cannot be overcome by increasing cognitive effort in comparative analysis tasks.

5.1 Background and Rationale

The findings from chapters 3 and 4 have shown that people are more capable of making accurate judgments when data is provided in concrete values with non-normalized base rates. In chapter 3, experiment 1, retail forecasters were shown to exhibit strong format biases which led them to additively process percentage points as if they were absolute values. The retailers' decision making strategy involved them employing minimal data, using only two values when shown three data points with which to predict weekly product sales. This strategy involved additively combining the last two observations in each sequence, using the absolute difference between the two points to generate predictions which were akin to a two-point linear regression model. The inclination to use only two numbers and combine them using an arithmetic operation therefore gave rise to estimates which were consistently linear in nature. In the case of non-linear increasing and decreasing trends, this resulted in forecasts which systematically underestimated increases and overestimated decreases in sales for the fourth week.

The same strong theme of a linear prediction heuristic was found to continue throughout chapter 4, in which humanitarian aid workers were shown to linearly extend trends in the context of real-world cyclical trends in refugee camp data. This robust linear prediction bias was evident in sparse data contexts (where no additional cues existed) and was shown to intensify in noisier data with additional cues akin to the real-world environment. Thus, the more complex the situation, the more people tended to project trends linearly which also correlated with an increase in judgment error. The directionalities of the trends in noise were specifically found to influence the direction of peoples' judgments. Although the contextual information had no relation to the target forecast data, the predisposition to seek linear patterns and extend trends linearly was shown to be influenced by a directional bias whereby people linearly extended the target data in accordance with the trends of other cues in noise when a consensus was detected between the directions of all observable data points.

These results suggest that where numerical information is more complex or varied, people employ fast heuristic judgment strategies which involve applying linear functions regardless of the context or causality of the data, or whether the trend of the target data is linear or not. The fact that this tendency is more pronounced in richer contexts where there are multiple cues and greater variability suggests that people are particularly prone to use any added information in a way that reinforces linear judgment tendencies. Thus, rather than using additional data in a sematic capacity (i.e., to help contextualize the target data based on the meaning of the surrounding cues), people tended to incorporate context information by applying a form of a visual heuristic strategy, based on seeking commonalities between the trends in noise and the target data. This leads to the detection of false relations and 'meanings' within the data environment.

In sum, fast, frugal heuristic judgment strategies based on linear functions and additive processing were applied by the decision makers in both the professional domains investigated. Where data was presented in numerical values (as opposed to time series) with no additional context cues available, retail forecasters were shown to 'make sense' of the information by applying a cognitively efficient arithmetic strategy to two cues to yield linear judgments akin to a 2-point regression model. Where context information was provided and data was presented in time series format, humanitarian aid workers employed a strategy which employed all available data in their judgments based on similarities in trend directions. Having visually identified congruence between trends, they then projected the target data linearly in the same direction. Thus, using the non-causally related cues to inform judgments combined with the strong propensity to apply a linear function in the context of cyclical data meant that prediction inaccuracy was high, and greater in the context of noise.

When considering the domain of financial judgment, the ability to form an optimal decision is recognised as a notoriously complex and cognitively demanding task which involves marshalling and computing multiple sources of information conveyed to consumers predominantly in percentage and rate formats. Judgment 'rationality' in the financial choice context is complex as it can be dependent on individual circumstances which may lead to unintuitive results. For example, an individual may opt to put off larger payments into the future in the knowledge that they are due some inheritance or an increase in income later on. In this circumstance, showing a propensity for delay discounting would not be considered 'irrational' in the context of the individuals' wider life factors.

Due to the complexities involved in measuring individual life circumstances in the context of financial decision choice and behaviour, the 'rationality' of peoples' judgments in experiments 3a and 3b was measured from the economic perspective. From this standpoint, choices which act to minimize total repayment costs over the full loan term are categorized as 'rational' or fully 'optimal' judgments, whereas other choices are classified as 'irrational' or 'suboptimal'. An interval scale of measurement of choice optimality is also included in the analysis which provides further data relating to the degree of rationality of peoples' judgments, as oppose to a strict correct/incorrect indicator.

Based on the robustness of judgment biases shown among the retail forecasters, there is a premise for assessing the tendency to additively processes percentages in financial decision making. It is possible that data manipulations designed to mitigate such biases in product choice situations will improve the effectiveness of consumer judgment. There are several mechanisms via which percentage format biases and the linear prediction heuristic could impact the rationality of peoples' financial judgment. As previously discussed, the tendency to additively processes rate information leads to people underestimating the increasing costs associated with rising interest rates. This simple bias alone can have grave consequences when selecting between credit product options which are subject to base rate and individual lender rate hikes. The potential for erroneous financial choices is further increased when considering additional factors such as peoples' limited attentional capacity, the nature of contextual cues and the complexity of the decision environment.

It is likely that the type of context data and how people use additional information in the process of making comparisons and evaluating choice alternatives will be important to decision performance. For example, framing choice alternatives in terms of absolute costs and displaying them simultaneously in a joint presentation mode could help to communicate the effects of compound interest growth without depending on peoples' ability to correctly extrapolate non-linear functions. This format could facilitate comprehension of the effects of rate changes over time and support the evaluative analysis of different cost options.

Individual difference factors such financial literacy and levels of numeracy are also likely to influence the effectiveness of peoples' financial judgments and behaviours. Both financial literacy and numeracy are associated with the accuracy of financial and medical risk assessments and thus warrant assessment. Interactions between framing manipulations and these individual difference factors could indicate for example that the framings are successful in mitigating negative effects of individual differences. Individual differences in trait optimism is also potentially relevant to financial judgment rationality.

In comparison to intertemporal effects like delay discounting or a more general present bias, trait optimism it potentially more important to financial judgment and behaviour because it encompasses a broad range of behaviours and characteristics which are shown to be related to the effectiveness of peoples' financial actions and judgments via indirect mechanisms. For example, at extreme levels, optimism is associated with saving less, planning over shorter horizons, working fewer hours, and holding more individual stock (Puri & Robinson, 2007). Conversely, moderate levels of optimism are associated with more optimal financial behaviour via the mechanism of increases in self-control.

Findings from the Competition and Markets Authority 2015 investigation into the Payday lending market provide further support for optimism as a key factor underpinning consumers' suboptimal financial choices and behaviours. In the 'shopping around' working paper (https://www.gov.uk/cma-cases/payday-lending-market-investigation#working-papers), the CMA cites 'over-optimism' as the reason for people significantly underestimating the likelihood of not repaying a loan, or of needing to take out further loans to cover repayment costs. The findings were based on qualitative responses combined with the high frequencies of late loan repayments, and loans which were either rolled over or never repaid in full. This led the CMA to conclude that "Customers would pay greater attention to the fees and charges associated with repaying a loan late (or with future loans), if their expectations about their ability to repay were more accurate".

It is possible therefore, that in the context of financial choice, a high temporal preference (i.e., the propensity to opt for low current costs and delay repayments into the future) is reflective of an optimistic bias which leads people to underestimate future costs and downplay the risk of future rate hikes. The negative effects of such a bias may be further compounded by people overestimating their ability to make higher repayments in the future. Although delay discounting or general present bias measures may be used to elicit temporal preference, such measures commonly involve a one-shot response to an individual delay discounting question which cannot account for individual circumstances or other factors mediating intertemporal effects. Measures of trait optimism on the other hand, assess more stable propensities and characteristics such as impulsivity which in turn are shown to impact financial judgment. From this perspective, optimism is potentially a more comprehensive and reliable assessment of intertemporal effects in this particular context, capable of accounting for the complexities in the mediating effects involved in financial judgments.

Anchoring is another mechanism which has the potential to significantly influence peoples' financial rationality. The anchoring and adjustment heuristic typically involves the presentation of an initial value or piece of information, commonly a numeric value on a dimension, which influences peoples' subsequent estimates and choices in a given judgment domain (Tversky & Kahneman, 1974). The anchor creates an initial basis for an estimates and people then make adjustments away from the anchor in the context of additional information. The anchor therefore sets the premise for a decision or choice which is altered according to peoples' interpretation and processing of the surrounding contextual data. In this sense, a numeric anchoring effect may be created by the presentation of default value in a numeric judgment environment such as online financial choice.

When presented with a set of choice alternatives, the default option is often considered as the choice that is accepted if no action is taken (i.e., it is the option a decision maker will obtain without engaging in further informational processing). However, a default setting may be used which involves a particular informational content being disclosed to the decision maker on first viewing of a data environment. A decision maker is then free to incorporate the default information in the judgment process as opposed to forming a choice based on taking no action. In this sense, a default setting still requires the decision maker to process the information to create a choice by comparing to other data points in the environment, however, based on the saliency of the default information, is likely that it will influence judgments to yield estimates and choices which align with the default data.

Default settings are commonly shown throughout behavioural science research to have a powerful effect on the decisions people subsequently make, increasing the likelihood that a particular selection is made. A default is commonly considered distinct from an anchor in that it provides an instantaneous choice in the absence of the comparative analysis of choice alternatives, whereas an anchor creates a bias in peoples' judgment processing which leads them to make choices which are more congruent with the anchor than the other values or options in the data environment. Although they are not always considered synonymous with anchors, it may be possible however, to apply a numeric default to elicit an anchoring and adjustment effect in the context of relevant additional information.

This is the case in the current experiment in which a default setting is used to create an anchoring and adjustment mechanism in the context of an online mortgage price comparison choice. It is expected that mortgage costs which are disclosed in rank order with a default view set to future interest rates will lead people to make more effective loan product choices compared to the disclosure of choices when the default view is set to current interest rates. Moreover, the effect of the future rate default is predicted to be enhanced by the subsequent disclosure of the current versus future rate choice alternatives side-by-side. By anchoring people on the most optimal loan choice (in the future rate frame) followed by simultaneously presentation of the current and future rate cost alternatives, it may be possible to encourage the direction of peoples' choices by facilitating the process of comparative analysis.

It is possible therefore, that the size of the values in the default view in relation to the values in the non-default view (i.e., the extent of the range between the upper and lower bound), might influence the amount of adjustment people make away from the default (or 'anchor') information which could result in more optimal loan choices (i.e., a lesser downward adjustment from the default) where a larger absolute difference between the upper and lower range is perceived.

In sum, the ideas postulated above suggest that removing percentage information and framing financial interest rates in absolute currency costs in easily comparable, simultaneous choice alternatives with a strategic default value could be effective in increasing mortgage choice optimality compared to data disclosed in standard industry formats. It is expected that current versus future rates framed in total cost alternatives will improve peoples' mortgage choices by heightening awareness of the potential for future rates rises (i.e., financial risk) and the financial effects of those rate increases (i.e., how compound interest amounts over time). This framing manipulation is predicted to mitigate optimistic biases associated with opting for variable rate loans which have low rate costs in the present compared to fixed rate loans which have higher initial rates but are not subject to the same rate variability and therefore represent optimal choice over the full term.

In the following review, each of these themes are discussed in turn, focusing on how each factor relates to financial rationality and how data frames utilizing these factors and mechanisms may by employed to enhance the effectiveness of financial judgment in an online price comparison context.

Understanding APR's and the effect of compound interest are key to making good financial decisions. Whether deciding to make a purchase on credit, how to manage savings, or judging which mortgage will work best in the long-term, the ability to correctly apply rate information is key to financial risk assessment and financial well-being. Confusion and error in consumer financial choice is frequently associated with the overall opaqueness of financial data communications. Based primarily on APR and percentage information, industry disclosures are shown to persistently create barriers to consumer decision making. Various findings point to how the reframing of finance industry data communications can be effective in reducing investor biases (Bateman, Dobrescu, Newell, Ortmann & Thorp, 2016; Beshears, Choi, Laibson, & Madrian, 2009; Choi et al., 2010; Fisch & Wilkinson-Ryan, 2014; Hung, Heinberg, & Yoong, 2010; Koehler & Mercer, 2009; Mercer, Palmiter & Taha, 2010; Newall & Love, 2015).

There is also a lack of evidence to support the effectiveness of standard finance industry disclaimer information on improving financial judgment and choice. For example, Newall (2016b) showed that adding a disclaimer manipulation in the form a social comparison 'nudge' ("Some people invest based on past performance, but funds with low fees have the highest future results") significantly increased investor fee sensitivity compared to the standard industry disclaimer ("Past performance does not guarantee future results"). When considering the negative effects that financial advisers can have on client biases' (Mullainathan, Noeth, & Schoar, 2012), these findings suggest that behaviourally-informed financial data frames and communications could be the most effective means of improving consumer financial judgment (Erta, Hunt, Iscenko, & Brambley, 2013).

With reference to the effects of interest rate variability, interest on savings and mortgages costs are subject to fluctuations in the Bank of England base rate. Previous rate forecasts suggested an increase from the static 0.5%, observed since 2009, to 2.0% by 2019. However, slowing global economy and negative inflation in 2015 has meant that the first rate rise to 0.75% is now predicted for 2019, increasing to 3% by 2025. This is not without the possibility however, of two rate hikes to 1.25% by the end of 2017 (Oxlade, 2016). Although the forecasts vary, it is indisputable that when the rate does incline, even the slightest of increases will dramatically effect credit repayments, translating into unmanageable monthly mortgage repayments for millions of UK homeowners (Wearden, 2015). Raising the saliency of compound interest and the impact of rate fluctuations on loan repayments is therefore highly important to the next generation of mortgage seekers.

As discussed in chapter 2, the errors and irrationalities in judgments involving percentage and rate information are numerous and widespread, stemming from the propensity to additively process numerical information. It is commonly assumed for example, that +10%, -10% = 0, based on adding and subtracting values, whereas multiplicative computation is necessary to derive the correct answer (-1%) (Newall, 2016a). In the context of financial decisions, it is well recognised that consumers have difficulty understanding the cost of credit (e.g., Lee & Hogarth 1999) and interest compounding (Eisenstein & Hoch 2007; Stango & Zinman 2009). The tendency to linearly processes compound interest rates can thus lead to harmful underestimates of the cost of borrowing (Stango & Zinman, 2009; McKenzie & Liersch, 2011) and slow debt repayment, as well as the miscalculation of investment fees and downside financial risk (Newall & Love, 2015; Newall, 2016a).

The tendency to additively process numbers and form linear predictions is likely to be associated with the cognitive ease and efficiency in relation to the demands of multiplicative processing and non-linear judgment (DeLosh, Busemeyer & McDaniel, 1997). The propensity to reason more effectively using concrete (absolute) values, suggests that simplifying financial product data by removing rates and percentages could significantly increase decision effectiveness. Figure 5.1 below shows an example of the standard informational format adopted by the finance industry to convey loan product information to consumers. Price comparison tools such as this are designed for people to make judgments regarding the most cost effective loan option based on comparing the interest rates and individual product attributes across loan providers, amounts and borrowing terms. As displayed in figure 5.1, standard price comparison tools disclose the representative APR which is the overall cost for comparison (4.4% in this particular example), the interest rate for the initial fixed period (1.39%) and the remainder of the term (4.74%), a max loan to value (70%), and a monthly repayment cost based on the initial fixed period only (£355.31). This data is delivered in combination with the industry disclaimer; “Your home may be repossessed if you do not keep up repayments on your mortgage” which appears at the bottom of the calculator shown in figure 5.1.

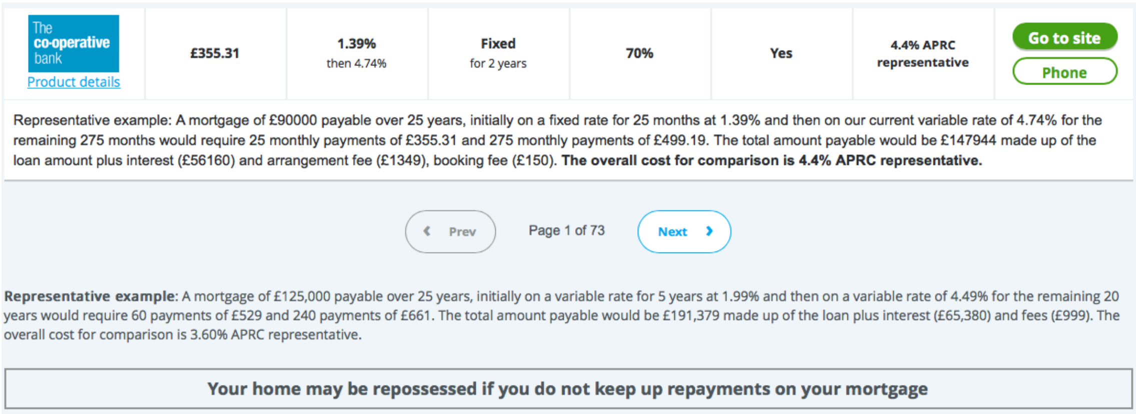


Figure 5.1 Example of a Standard Industry Mortgage Price Comparison Site

As figure 5.1 demonstrates, this format creates a highly complex decision environment, making it difficult for decision makers to synthesize the data points in the process of forming judgments of future costs. This format assumes an understanding of representative APR and compounding and requires decision makers to compute rate changes over the full term. The addition of the disclaimer does little to draw consumers' attention to the risks of base rate increases, added fees and lender rate rises. Thus, online price comparison websites represent a complex task environment, in which the synthesis of multiple data points and comparison of choice alternatives is particularly difficult to achieving optimal choice. In addition to future rate hikes, the compounding of arrangement fees, mortgage alteration charges, or an increase in the term by taking a repayment holiday for example, could result in greater monthly and final repayment costs than estimated based on the data provided.

One way to simplify the choice environment and de-bias judgments would be to present the total repayment cost of a mortgage over the full term in currency format in the current interest rates and future base rate increases. This would eliminate the need for multiplicative computation of interest rates over fixed and non-fixed horizons, and highlight the monthly and total cost differences between current and future rates. For example, a standard variable rate with a small initial fixed rate may seem like a more attractive choice at current rates compared to a fixed rate mortgage with a much higher rate for the initial period and full term. However, when factoring in base rate increases (separate from lender increases), the standard variable rate cost far exceeds that of the fixed rate over the full term, thus making it a far riskier option.

Reducing the quantity of loan attributes presented in the decision task is also likely to help alleviate the computational challenge associated with synthesizing multiple cues. The limitations in the amount of data that people are able to attain to at any one time are well recognised throughout Psychological literature (e.g., Broadbent, 1958), associated with heuristic judgments, often excluding important information (Simon, 1955). Recent research into peoples' limited attentional capacities have shown a neural basis for restrictions in the ability to prioritize information (e.g., Mecklinger et al., 2003), encode data into working memory (e.g., Todd & Marois, 2004; Scalf et al., 2007, 2011b), and respond to task-relevant material (e.g., Dux et al., 2006; Erickson et al., 2007).

The investigation of attentional capacity for multiple data points has thus focused on the limitations in cognitive resources (e.g., Intriligator & Cavanagh, 2001; Lavie & Robertson, 2001; Mitchell & Cusack, 2008; Xu & Chun, 2009), giving rise to models from a 'resource-limited' perspective (e.g., Alvarez and Franconeri, 2007) based on evidence of peoples' ability to individuate and identify single items, but failure to simultaneously perform these operations on multiple group members or individual items. Bounded attentional capacity in the context of financial choices means that peoples' quick, heuristic judgments about how to allocate it can lead to the exclusion of important information from the decision process. Rich, complex decision environments with multiple cues are therefore likely to hinder judgments by using up attentional resources and distracting people from the important points.

From the perspective of behavioural economics, increasing complexity by adding more choice alternatives has shown to reduce judgment performance in investment decision making (Benartzi & Thaler, 2002) and increasing the number of pension plan alternatives can lead to a decrease in the number of enrollers (Sethi-Iyengar, Huberman & Jiang, 2004). In general, probabilistic judgments are shown to be negatively affected by increases in informational cues (e.g., Harvey, Bolger & McClelland, 1994). Forecasting studies show that accuracy deteriorates as the environment becomes noisier because people tend to interpret the noise as signal, leading to erroneous predictions (Harvey, 1995; Harvey, Ewart & West, 1997; Lopes & Oden, 1987; Brehmer, 1978; Kahneman & Tversky, 1973; Lee & Yates, 1992).

In consumer domains, 'choice overload' created by increasing choice alternatives is shown to demotivate decision makers, leading to choice dissatisfaction and regret (Iyengar & Lepper, 2000). The increased difficulty of evaluating multiple attributes in richer choice domains is likely to create greater uncertainty and burden of responsibility for making an 'optimal' choice. In terms of mortgage choices, price comparison calculators represent noisy domains, as not all product attributes are necessary or useful to computation of total or monthly repayment costs. The removal therefore of redundant product attributes is likely to decrease the opaqueness of the data and increase consumer data comprehension, sense if control and judgment confidence.

In addition to removing data cues to simplify the environment, presenting the rate alternatives simultaneously is also likely to increase the evaluability of attributes. Numerical decisions are shown to be more effective when data is presented and processed in joint rather than in separate evaluation mode (Hsee, Loewenstein, Blount & Bazerman, 1999). Insights from user design indicate that people ascribe meaning and derive value from numerical data through making comparisons against a reference point, or within the context of defined upper and lower bounds (Roller, 2011). Judgments in many contexts are shown to vary significantly between data evaluated simultaneously versus independently, with preference reversal commonly occurring between the two formats (Hsee, 1998; Hsee et al., 1999).

The ability to evaluate and contextualize data through comparison against reference information is shown to be important to statistical inference in many situations (Hsee & Zhang, 2010). For example, judgments of risk, sensitivity to scope, and the stability of preferences are all shown to be differentially effected by data framed and evaluated in joint decision mode (simultaneously) verses information framed and processed separately (e.g., Bateman, Dent, Peters, Slovic & Starmer, 2007; Slovic, Finucane, Peters & MacGregor, 2002; Desvousges, Johnson, Dunford, Boyle & Wilson, 1993; Hsee, 1998). In the context of financial judgment, the ability to easily compare choice alternatives is therefore very important to optimising choice. This suggests that side-by-side presentation of the current and future rate costs for each loan choice could be more effective in creating a comparative context and promoting evaluative processing than framing current versus future rates in isolation.

Independent of framing manipulations however, the effectiveness of financial judgment and choice may be related to individual differences in financial literacy (Mitchell & Lusardi, 2011). For example, judgments of downside financial risk have been found to positively correlate with both financial literacy and numeracy (Newall, 2016a), and lower levels of numeracy and risk literacy are shown to relate to poor statistical inference and incomprehension of probability data in judgments of medical risk (e.g., Cokely, Galesic, Schulz, Ghazal & Garcia-Retamero, 2012).

In general, however, findings relating to the effects of financial literacy on financial behaviours are mixed. Evidence suggests that interventions to increase financial literacy have not led to improvements in financial behaviour, and that associations between literacy and behaviour may reflect other factors such as numeracy (Fernandes, Lynch & Netemeyer, 2014). In the context of credit card repayments for example, high numeracy is associated with the tendency to overestimate the repayment amount required to pay off a loan in three years, and vice versa for low numeracy consumers (Soll, Keeney & Larrick, 2013). Interestingly, a reverse effect of financial literacy has been found in investment decision making, with people higher in financial shown to be more prone to select mutual funds with higher fees rather than opting for more effective low fee alternatives (Newall, 2016a). Although the relation between financial literacy, numeracy and behaviour is unclear, it makes sense to include such measures when examining financial judgments as all result are useful in further distinguishing the mediators and moderators of financial behaviour in different financial decision contexts.

Other potentially relevant individual difference variables shown to influence financial judgment and behaviour are optimism and temporal preference. From the dispositional perspective, optimism is regarded as generalized positive expectations about future events which remains constant across context (Scheier & Carver, 1985; Scheier, Carver & Bridges, 1994). Optimistic bias, conversely, can vary from one situation to another and involves overestimates of the probability of a favourable outcome or underestimates of a negative outcome (Weinstein, 1980). Over-optimism is thus shown to have negative effects, specifically with respect to personal risk miscalibrations for illness and mishap which can lead to neglect of precautionary measures (Weinstein, 1980; Weinstein & Klein, 1996). From an economic perspective, it is possible that optimism biases align with overconfidence in the likelihood of future financial events which can thus lead to suboptimal decisions based on decreases in the overall utility associated with a particular financial choice.

Higher levels of unrealistic optimism measured in terms of 'underestimated future borrowing behaviour and generalized wishful thinking' (Vitaliano, Carr, Maiuro & Becker, 1985) are shown to associate with decreased sensitivity to APR's and increased sensitivity to annual fees. In the context of consumer credit behaviours, these findings suggested that higher unrealistic optimism may be predictive of selecting credit cards which are suboptimal in terms of an individual's usage behaviours (Yang, Markoczy & Qi, 2007). Higher levels of optimism have also been found to relate to generally suboptimal financial behaviours when examined in a wider context of financial and demographic characteristics.

Puri and Robinson (2007) developed a measure of optimism based on the miscalibration of life expectancy using data from the Survey of Consumer Finances (SCF) which collects data on subjective and actual life expectancy among many other demographic and financial features. The measure was validated based on strong correlations with the revised version of the Life Orientation Test (LOT-R; Scheier, Carver & Bridges, 1994), and generalized positive expectations about the economy and future income growth. Puri and Robinson (2007) found that moderate optimism was predictive of optimal financial habits and choices such as paying credit card balances on time, planning over longer horizons and saving more. However, extreme optimism was related to saving less, shorter planning horizons, fewer working hours, and holding more individual stock.

The authors explain the apparent dichotomy between extreme and moderate optimism in terms of important differences in self-control. For example, extreme optimists were found to be more likely to smoke, work fewer hours and hold less liquid assets compared to moderate optimists who were more likely to work harder and less likely to be day traders. These results indicate that the link between optimism and financial temporal preference may be related to behavioural differences in self-control. For example, moderate levels of optimism are related to a better ability to plan long-term and execute those plans which are both skills which correspond with higher levels of self-control. It is therefore possible that higher self-control instantiates a preference for higher future rewards based on the ability to delay gratification and forgo short-term rewards which do not ultimately maximise utility. In turn, these propensities create more prudent financial choices and behaviours. The increased self-control of moderate optimists also means that they have less need to rectify financial problems compared to extreme optimists who are more likely to overcome self-control problems by locking their money into illiquid assets (e.g., Laibson, 1997; Strotz, 1955).

Optimism has also been explored in the context of credit card repayment behaviours, testing the effects of minimum required payment information. Navarro-martinez, Salisbury, Lemon, Stewart, Matthews & Harris (2011) showed that presenting minimum repayment information had a detrimental effect on decisions to pay, leading to lower monthly payments (Steward, 2009; Navarro-martinez et al., 2011). Disclosing future interest costs in addition to minimum repayment amount was found to increase consumers' likelihood of paying the minimum. However, this improvement did not hold when time to pay off the balance was added. The only manipulation found to increase the likelihood of paying the minimum amount was increasing it from 2% to 5% of the balance. However, this was only effective among borrowers with a moderate to high propensity to pay minimum amounts in general.

These results suggest that cost sensitivity, based on temporal preference (i.e., individual differences in the propensity to pay now versus in the future), is likely to interact with informational framings to generate differential financial behaviours. A low propensity to make minimum payments may therefore be viewed as a present-focus or optimistic bias regarding future ability to repay, characterized by people seeking to minimise losses in the short term by delaying payments into the future. Other studies have also shown that optimistic biases in the propensity to consider immediate concerns versus future concerns influence credit card debt (Joireman, Kees & Sprott 2010) and fiscal responsibility (Joireman, Sprott & Spangenberg 2005).

As Stewart (2009) observed, presenting a minimum repayment requirement on a credit card statement is likely to create an anchor from which adjustments are made to form subsequent repayment decisions (Tversky & Kahneman, 1974). Numerous findings in behavioural economics research show that default values can have a powerful anchoring effect which impacts subsequent decisions (Johnson & Goldstein, 2003). Analysis of retirement planning decision making for example, showed that consumers interpreted a default set by the employer of 401k as an implicit recommendation, and were subsequently more likely to select it (McKenzie, Liersch & Finkelstein, 2006). Thus, it is likely that the propensity to focus on an initially observed value will vary with optimistic biases or differences in temporal preference to differentially influence financial choice and behaviour. For example, Haws, Bearden & Nenkov (2012) found that repayment amounts among consumers who were low in spending self-control were increased when supplemental information relating to minimum repayment amount was disclosed.

Anchoring can also work adversely however, particularly where data is presented in isolation without a reference point for evaluation. Mussweiler, Strack & Pfeiffer (2000) demonstrated the importance of making an alternative repayment amount salient as an alternative anchoring point. For example, in Navarro-martinez et al's (2011) study, it is likely that the repayment value ($38.74) was interpreted as a loss and without another amount for comparison, participants subsequently adjusted down from the anchor when deciding how much to repay. As the minimum amount was increased from 2% to 5% of the balance, the increase in the propensity to pay the minimum thus reflected the downward adjustment from the 5% minimum towards the 2% minimum. The fact that this occurred in those more prone to make minimum payments suggests that the effects of anchoring may be stronger for those with a stronger optimistic bias, or who are less future-focused.

A possible solution to the negative effects of minimum payment requirements on repayment decisions may be to consider the effects of percentage versus currency frames. For example, findings show that investors sensitivity to product fees is increased when costs are framed in currency rather than percentage, due to the tendency to down weight the smaller percentage values (Newall & Love, 2015). This so called 'peanuts effect', in which people tend to discount smaller costs and rewards (Weber & Chapman, 2005), is found throughout many behaviours which involve incurring either a small loss or gain. When considering health behaviours, this bias explains the detrimental effects in repeat behaviours such as smoking (Loewenstein, Asch, Friedman, Melichar & Volpp, 2012) and other diet or exercise behaviours where small effects of repeated actions may be downplayed.

Based on the general misinterpretation of percentage formats, reframing the minimum repayment cost ($38.74) in percentage may therefore be effective in leading consumers to perceive costs as 'small' in the context of the 'large' balance framed in currency (e.g., 2% of $1,937.28). The fee may now be evaluated differently in relation to the balance, thus increasing willingness to pay. Conversely, people may be more likely to avoid making payments when costs are framed in current (i.e., seemly large values in relation to the balance) compared to a percentage (i.e., seemingly 'small' values compared to the balance).

The simultaneously display of the balance in currency costs and percentage formats could therefore be particularly effective in improving financial decisions by countering optimistic biases among those who are more prone to delay repayment or pay less than the minimum requirement. Other means of mitigating the negative effects of disclosing minimum repayment requirements may be to frame them in the context of alternative repayment options. For example, Salisbury (2014) showed that proving costs and loan repayment duration information for an additional higher amount in conjunction with the minimum costs, proved successful in increasing monthly credit card repayments above the minimum required.

Levering the psychological process involved in financial judgments is key to delivering disclosures which target the biases underpinning suboptimal choice (Loewenstein, Sunstein, & Golman, 2014). Behavioural "nudges" (Thaler & Sunstein, 2008) in the form of disclaimer manipulations are also shown to be effective in altering perceptions of cost in financial judgements. For example, Newall (2016b) compared the effects of three different disclaimers on choice optimality when selecting between mutual funds offering monotonic trade-offs between maximising past returns and minimizing fees. The three disclaimers were designed to increase sensitivity to fees utilizing mental accounting (Thaler, 1985), sensitivity to costs by evoking loss aversion (Kahneman & Tversky, 1979), and evaluation of own behaviours relative to others using social comparison (Buunk & Gibbons, 2007). When compared to the industry disclaimer, "Past performance does not guarantee future results", the social comparison disclaimer, "Some people invest based on past performance, but funds with low fees have the highest future results", was found to significantly increase sensitivity to investment fees, thus reducing the tendency to maximise past returns when selecting mutual funds.

Standard financial industry disclaimers are found largely to be ineffective in improving financial judgment and behaviour. When assessing the optimality of investment decisions for example, Mercer, Palmiter and Taha (2010) found that "Past performance does not guarantee future results" was no more helpful in guiding investment decisions than providing no disclaimer at all. Among those which have shown to be effective, emphasizing the importance of fees is the key feature facilitating the improvements in investment decisions (Mercer et al., 2010; Fisch & Wilkinson-Ryan, 2014). In the case of mortgage products, the industry disclaimer, "Your home may be repossessed if you do not keep up repayments on your mortgage" is also likely to little to draw attention to the risks associated with variable interest rates, or increase consumers' consideration of the need for careful financial planning when entering a long-term credit agreement.

Among other factors, several experimental manipulations show the importance of simplification when designing effective disclosure information. For example, uptake of Earned Income Tax Credit, retirement plans, and a reduction in interest payments, late fees and over-limit fees have been attributed to simplification of mandatory disclosures (Bhargava & Manoli, 2013; Clark, Maki & Morrill, 2014; Agarwal, Chomsisengphet, Mahoney & Stroebel, 2013). As demonstrated by Salisbury (2011), the provision of comparison data is also fundamental to the effectiveness of disclosures. Being able to accurately assesses the cost trade-offs associated with different interest rates when selecting between different financial products can significantly enhance decision making. Standardized information formats (e.g., absolute currency costs) create a meaningful context and reduce the cognitive workload associated with comparing rates and costs over different time horizons independently (e.g., Hsee et al., 1999).

Financial costs presented in currency as opposed to percentages are also made more salient when presented comparatively, as is the case for energy savings when framed in monetary values (Newell & Siikamäki, 2013). With respect to payday loans, presenting prospective borrowers with the monetary cost of loans across varying terms compared to the relatively lower costs of credit card debt, was shown to be effective in reducing loan amounts and decisions to take-up loans (Bertrand & Morse, 2011). This format was more effective than disclosures involving comparisons between the APR's of payday and other loan types, thus exemplifying the importance of concrete (non-rate) formats to evaluative judgment processes. Choices of investment funds with fee alternatives framed in dollars is also shown to impact fund choice to a greater extent than with fee alternatives presented in percentage points (Hastings & Tejeda-Ashton, 2008).

In sum, financial judgments involving long-term credit agreements such as mortgage product choices are subject to cognitive and behavioural biases which limit judgment effectiveness. Strong evidence for the effects of numerical format biases based on additive processing, and the tendency to seek and apply linear trends in complex multi-cue environments were shown in experiments 1 and 2 in chapters 3 and 4. Experiment 3a involves an application of these findings to a real world environment by testing the effects of framing manipulations designed to overcome percentage format biases and the linear prediction heuristic in the context of financial judgment. In experiment 3a, a comparison is made between data disclosed in a standard industry financial disclosure with a manipulation involving interest rates reframed in concrete currency costs over the full mortgage term. The rates disclosed in currency format are then shown in current versus future rate alternatives.

Compared to standard industry formats, it is predicted that mortgage choices will be most effective when current and future rates frames are displayed simultaneously with a default set view to future rates. It is expected that moderate improvement in choice effectiveness will be shown when rates are displayed sequentially (on separate screens) with a default set to current rates. These effects are based on mitigating the tendency to underestimate costs by arithmetically processing percentages and linearly extrapolating non-linear growth which is shown to lead to erroneous financial judgment.

To test the robustness of the framing effect examined in 3a, experiment 3b is conducted as a replication study to assess the effect of a disclaimer manipulation, akin to a real-world financial industry disclaimer which is designed to increase attention and cognitive effort applied in the task of comparatively analysing the choice alternatives in each condition. The disclaimer aims to achieve increased attention and effort by making salient the effects of rate variability and the financial risks associated with selecting low current rate options. The purpose of experiment 3b therefore, is to identify whether simply prompting people to attend more carefully to financial data and apply greater cognitive effort when comparatively analysing multiple cues and attributes is sufficient to improve judgment performance when viewing data in standard industry rate and percentage formats. For example, if this were to be the case, we would expect to see greater choice effectiveness in the control condition in experiment 3b compared to experiment 3a, thus indicating that people possess the cognitive capacity to rationalise effectively using rate and percentage formats and that more complex judgment processes can be activated by sufficiently prompting or motivating people to do so.

Based on the robustness of the linear prediction heuristic and the saliency of percentage format biases shown in chapter 3 and 4 it is expected however, that choice performance in experiment 3b (with the addition of the disclaimer) will not significantly differ to performance in experiment 3a. It is predicted that the positive effects of reframing current versus future rate data in absolute (concrete) currency costs on financial judgment will hold, showing that framing methods to overcome format biases and attentional limitations are necessary to human rationality and that people do not possess the cognitive capabilities to synthesize percentage data in complex probabilistic domains. Thus, warning people about the risks of rate variability to prompt engagement in deeper comparative analysis of rate data will not be sufficient to yield judgment improvements.

As explained in the above review, measures of financial literacy, numeracy and trait optimism are also included in experiment 3b to assess possible interactions between relevant individual difference variables and the framing manipulations. The findings relating to the influence of individual differences on financial rationality and behaviour are somewhat mixed. Therefore, no specific predictions are made regarding the direction of the effect of financial literacy and numeracy on judgment performance and the measures are included for investigative purposes to further examine the possible mechanisms underpinning the effectiveness of the framing manipulation in the context of online financial choice.

With regard to the effects of trait optimism, it is expected, based on Puri and Robinson's (2007) findings, that higher trait optimism will be predictive of lower choice effectiveness and interact with frame preference within the two framing manipulation conditions (i.e., the rate frame in which people choose to make a selection). This prediction is derived from the view that lower temporal preference (or greater 'self-control' from Puri and Robinson's perspective) will instantiate a preference for higher future rewards (at the expense of lower costs in the present) which is shown to be associated with more prudent financial choices and behaviours that act to increase utility. In this sense, lower trait optimism may be associated with a tendency to delay gratification and forgo short-term rewards (in the form of lower present loan costs) and opt for higher cost, fixed rate loans based on future rates options which yield higher overall utility when considered over the full loan term in the context of increasing interest rates.

5.2 Experiment 3a Current Vs Future Interest Rate Disclosure Effects on Financial Choice

In this initial experiment, a mortgage price comparison calculator was built which resembled a standard industry mortgage price comparison tool displaying real-world loan product data (see figure 5.1 below for a screenshot of the tool). Just as in real-world comparison websites, the tool presented key information to the consumer relating to APR, initial and subsequent interest rates, and estimated monthly costs based on initial fixed rate periods. Each row corresponded to a lender and the columns provided the different attributes which could be used to ranked and order the product alternatives. Although this *standard disclosure* provides all the necessary information to make a mortgage decision based on a given term and borrowing amount, the ability to make an effective choice is dependent on correctly processing percentages, understanding compounding and synthesizing multiple data points whilst controlling for risk associated with additional charges and future rate variability.

Decision making performance in the *standard disclosure* condition is thus compared to choice effectiveness when data is framed in an *interactive/optimistic* and an *interactive/realistic framing* condition in which currency costs are framed in *current* versus *future interest rates* with differential default views.

It is important at this point to delineate the use of the term 'framing' to describe the data manipulations tested in experiments 3a and 3b, and how the test of framing is distinct from the examination of informational presentation undertaken in experiment 4, chapter 6. Specifically, a framing effect typically refers to the impact a piece of information has on peoples' preferences, judgments and choices when the same information is presented in different ways. For example, when given the same data, creating a positive informational frame is shown to increase peoples' propensity to avoid risk, whereas a negative frame is associated with an inclination to seek risk (Tversky & Kahneman, 1981). The data presented in each condition in experiment 3a and 3b is exactly the same - i.e., the loan attributes used to populate each trial were extracted from the same set of mortgage products selected from those available on the market in 2015. Thus, the examination of the alternative informational formats in the three conditions in experiments 3a and 3b are tests of framing effects in the traditional sense.

In experiment 4, chapter 6 however, the data presented in condition 2 differs to that show in condition 1 with respect to the future rate information. Both conditions 1 and 2 disclose full and reduced term repayment information in current interest rates, whereas only condition 2 involves the full and reduced term repayment information in future rates. In this sense, experiment 4 involves the examination of the impact of two different data manipulations on repayment judgments as opposed to a 'framing effect' based on the alternative presentation of the same information in each condition.

The current rates are those set by the mortgage lenders during April 2015 (at the time of writing) and the future rates are set at a +1.5% increase applied in 0.5% increments over three consecutive years from the second year of the loan term. This +1.5% rate increase was designed to reflect the Bank of England base rate increase forecast in early 2015 to rise to 2% by 2016. Thus, simulating a rise of +1.5% between 2016 and 2018 represented a realistic and conservative estimate of future repayment costs at the time of study design.

Combined with the disclosure of current versus future rate costs, the optimistic and realistic framing conditions also greatly simplified the decision environment by making APR and rate information viewable only on demand, thus reducing computational barriers and helping to counteract attentional limitations. To recap, the findings reviewed in the above section indicate that people tend to treat percentages as absolute values and compute future outcomes by adding and subtracting percentage points as if they were whole numbers. When extrapolated into the future, computations based on arithmetic numerical processing thus lead to the linear projection of growth, which in the case of exponentially increasing compound interest rates can lead to significant underestimations of future repayment costs. When considered in the context of financial judgment therefore, the tendency to linearly extrapolate translates into a form of optimistic bias, leading people to assume lower future costs and thus less risk associated with potential rate increases.

It is possible that judgments based on arithmetic processing are a key factor underpinning peoples' optimistic ('unrealistic') financial estimates. Applying additive methods to rate and percentage information will lead people to make underestimates of costs, and thus predict 'manageable' future repayments. This is likely to result in people discounting higher future costs and opting for the most favourable choice in the present (i.e., lowest current costs). An alternative view is that people approach financial decision making with a pre-existing optimistic bias, leading them to assume that rates are unlikely to rise and that if they do, they will somehow be able to meet the increased financial demands in the future. From this perspective, the way optimism may impact financial judgment could be multifaceted. However, based on the strength of evidence for processing inaccuracies stemming from format biases, it is probable that the former viewpoint is the main precursor to peoples' optimistically biased financial judgments and choices.

The use of rate and percentage data in a complex decision domain involving multiple cues and data points such as financial price comparison, therefore places demand on attentional resources which exceeds peoples' cognitive capacities. Hence, it is predicted that the important manipulations of reducing the quantity of data points necessary for decision making and replacing rate formats with absolute currency costs framed in current and future rates will significantly increase the effectiveness of financial choices compared to the standard format condition. Specifically, the simultaneous display of rate alternatives combined with a default set to future rates in the realistic framing condition is expected to be the most effective overall by overriding the optimistic tendency to discount higher future costs (or unfavourable outcomes) in preference for smaller (more favourable) present costs.

5.2.1 Method

Participants

One-hundred and seventy-nine participants aged over 18 years were recruited via Amazon Mechanical Turk and paid \$2.00. The average age was 36.8 years (SD = 10.02, range = 21 to 67 years), 40.40% were female and 40% were educated to a minimum level of a college degree. The sample size was determined by a power calculation which showed 159 participants necessary for 80% power based on a one-way ANOVA (α = 0.05, d = 0.25, 1-β = 0.8). The budget was set accordingly and data collection proceeded until the budget was reached.

Design

The experiment was conducted as a randomized controlled trial using an independent groups single factor design. Participants were randomly assigned to one of three conditions: 1) *standard disclosure* (control); 2) *interactive/optimistic framing*; and 3) *interactive/realistic framing*. In each condition, participants underwent eight trials (two of which were attention checkers) in which they selected what they believed to be the best mortgage from five alternatives for a given loan amount and term. Trials were randomized per participant in each condition.

Materials

There were eight unique trials per condition in which five mortgage choices for different term and amount combinations were presented. All trial stimuli were extracted from real-world mortgages available on the UK mortgage market at the time the experiment was designed during April 2015.

Each trial in all three conditions involved a choice between one or more standard variable rate (SVR), discounted variable rate (DVR), and fixed rate mortgages. Each condition differed in the presentation and framing of the mortgage data across all eight trials. Figure 5.2a shows a screenshot of the control condition 1 (*standard industry disclosure)* in which the five loan options were presented in an interactive price comparison tool resembling a traditional, real-world mortgage price comparison website. The tool design, user interface and the specific product attributes presented were exactly the same as the attributes provided on commercial websites available in 2015.

Mortgage type	Max LTV	Initial rate	Subsequent rate	Initial average monthly cost	Overall cost comparison	SELECT MORTGAGE
Discounted variable rate A	85%	1.99% for 2 yrs	then 4.75%	£391.29	4.30% APR	Select A
Discounted variable rate B	85%	1.85% for 2 yrs	then 4.99%	£391.03	4.60% APR	Select B
Discounted variable rate C	80%	1.69% for 2 yrs	then 5.95%	£390.72	4.50% APR	Select C
Fixed rate D	85%	5.58% for 5 yrs	then 5.73%	£449.58	6.40% APR	Select D
Fixed rate E	95%	6.09% for 2 yrs	then 5.73%	£399.06	6.20% APR	Select E

Figure 5.2a The Standard Industry Disclosure Cond 1 (Control) in Exp 3a

The standard industry disclosure condition (control) in which all the loan product attributes and the design and layout of the information directly matched real-world mortgage price comparison websites. As in the real-world context, loans could be ordered by the individual attributes using toggle keys in the header. Although all the product attributes are made available, making an optimal decision depends on synthesizing multiple variable rate data and is thus particularly difficult.

Total repayment costs over full term. Show Future rates

Mortgage type	At CURRENT rates	Price rank order	Monthly cost	SELECT MORTGAGE
? Discounted variable rate A	£659,626	1	£2,209.20 average for full term	Select A
? Standard variable rate B	£662,759	2	£2,198.75 average for full term	Select B
? Discounted variable rate C	£674,316	3	£2,247.72 average for full term	Select C
? Fixed rate D	£797,247	4	£2,657.49 fixed for lifetime	Select D
? Fixed rate E	£798,068	5	£2,660.23 fixed for lifetime	Select E

Details for Discounted variable rate A : 1.99% for 2 yrs then 3.94% and available on properties with a Max LTV of 80%. The overall cost for comparison is 3.7% APR.

Click '?' to see individual loan interest rates and APR's

The optimistic framing condition 2 in current rate frame (default view)

Total repayment costs over full term. Back to Current Rates

Mortgage type	AT FUTURE RATES	Price rank order	Monthly cost	SELECT LOAN
? Discounted variable rate A	£768,709	1	£2,562.36 average for full term	Select A
? Standard variable rate B	£772,544	2	£2,575.15 average for full term	Select B
? Discounted variable rate C	£783,893	3	£2,612.98 average for full term	Select C
? Fixed rate D	£797,247	4	£2,657.49 fixed for lifetime	Select D
? Fixed rate E	£798,068	5	£2,660.23 fixed for lifetime	Select E

Details for Discounted variable rate A : 1.99% for 2 yrs then 3.94% and available on properties with a Max LTV of 80%. The overall cost for comparison is 3.7% APR.

Click '?' to see individual loan interest rates and APR's

The optimistic framing condition 2 in future rate frame (non-default view)

Figure 5.2b The Optimisic Framing Condition 2 in Exp 3a

The optimistic framing condition 2 showing the default frame set to total repayment costs in *current rates* (the default view) and the alternative frame showing total repayment costs in *future rates* (the non-default view). Participants were free to click backwards and forwards between the rate frames in the two separate screens before making a mortgage choice in either screen. Loans in both frames were rank ordered by total loan cost over the full term, thus dramatically simplifying the comparison problem compared to the standard disclosure. Participants could click to see individual loan interest rates and APR's if desired (otherwise, all rate and APR data was removed to simplify the decision process). The default set to current rates is predicted to exacerbate the optimism bias, anchoring participants on best option in current rates, thus making it moderately more effective compared to the control.

Total repayment costs over full term. show CURRENT rates

Mortgage type	At FUTURE rates	Price rank order	Monthly cost	SELECT MORTGAGE
? Fixed rate D	£393,389	1	£1,311.30 fixed for lifetime	Select D
? Fixed rate E	£394,797	2	£1,315.99 fixed for lifetime	Select E
? Discounted variable rate A	£422,050	3	£1,406.83 average for full term	Select A
? Discounted variable rate B	£423,134	4	£1,410.45 average for full term	Select B
? Discounted variable rate C	£424,201	5	£1,414.00 average for full term	Select C

Click '?' to see individual loan interest rates and APR's

The realistic framing condition 3 in future rate frame (default view)

Total repayment costs over full term. Back to FUTURE Rates only

Mortgage type	At CURRENT rates	Price rank order	Monthly cost	At FUTURE rates	Price rank order	Monthly cost	SELECT MORTGAGE
? Fixed rate D	£393,389	4	£1,311.30 fixed for lifetime	£393,389	1	£1,311.30 fixed for lifetime	Select D
? Fixed rate E	£394,797	5	£1,315.99 fixed for lifetime	£394,797	2	£1,315.99 fixed for lifetime	Select E
? Discounted variable rate A	£360,867	1	£1,202.89 average for full term	£422,050	3	£1,406.83 average for full term	Select A
? Discounted variable rate B	£361,737	2	£1,205.79 average for full term	£423,134	4	£1,410.45 average for full term	Select B
? Discounted variable rate C	£362,222	3	£1,207.41 average for full term	£424,201	5	£1,414.00 average for full term	Select C

Click '?' to see individual loan interest rates and APR's

The realistic framing condition 3 in current rate frame (non-default view)

Figure 5.2c The Realistic Framing Condition 3 In Exp 3a

The realistic framing condition in which the default was set to total repayment costs in *future rates* (the default view). When participants clicked to see costs in *current rates* (the non-default view), the two repayment alternatives were displayed onscreen side-by-side as opposed to on separate screens (as in the optimistic framing condition). This was designed to facilitate evaluative judgment processes by increasing sensitivity to the effects of rate changes over the loan term, and how these variances differentially effected costs depending on mortgage type. This framing is predicted to be the most effective by anchoring participants on the most realistic cost scenario, and facilitating the comparative analysis of the different mortgage types and in the context of the rate alternatives to a greater extent than the sequential disclosure in the optimistic framing condition.

In the *optimistic framing condition 2* shown in figure 5.2b, participants viewed an interactive and simplified version of the traditional loan price comparison tool in the control condition. Unlike the traditional disclosure, the optimistic framing condition had all rate information removed and total loan repayment costs over the full term (in UK currency) were displayed in price rank order. The default screen (i.e., the view that participants viewed first in each trial) was set to the rank ordered prices in *current rates* (the first screenshot), hence the 'optimistic' framing. In the header, a tab labelled 'show future rates' enabled participants to click to a new screen where they viewed loans rank ordered in *future rates* (the second screenshot). They could then alternate between the two screens to view the current and future rates separately by clicking on the tab in the header.

The same as in control condition, the loans could also be ordered by individual attributes using toggle keys in the header. However, in contrast to the control condition, computation of APR information was not necessary for decision making and was therefore not displayed in either rate frame. To replicate the legal data disclosure requirements of standard websites, the interest rate and APR data for each loan was viewable if desired, made accessible by clicking on the '?' icon next to each mortgage choice. The message "click to see individual rates and APR's" displayed at the bottom of the tool was provided to prompt this action. Figure 5.2b *default* and *non-default* views show examples of what happened when the '?' icon was clicked. In this instance, the green bar at the bottom of the tool displayed the individual data for the selected loan.

In the *realistic framing condition 3* shown in figure 5.2c, the informational framing in the optimistic framing condition 2 was repeated, except that the default view was set to total loan costs in *future rates* (hence the 'realistic' framing) and the alternative rate frames were disclosed simultaneously (i.e., side-by-side on the same screen), as opposed to sequentially on separate screens.

Generation of the Trial Stimuli

All trial stimuli were constructed using data from real-world mortgages available on the UK mortgage market during April 2015. Across eight trials involving different duration and amount combinations, participants selected what they believed to be the best loan from five mortgage options with various current rate terms (see table 5.1 for the trial stimuli details).

Table 5.1 Loan Term, Amount and Rate Combinations Per Trial in Exp 3a
The mortgage term, borrowing amount and current interest rate combinations presented per trial in each condition.

Trial 1 and 2		Trial 3 and 4		Trial 5 and 6		Trial 7 and 8	
10 year term		15 year term		20 year term		25 year term	
£110,000	£150,000	£220,000	£400,000	£95,000	£275,000	£185,000	£375,000
1.89% for lifetime	4.45% for lifetime	1.50% for 2 yrs then 4.99%	3.69% for lifetime	1.99% for 2 yrs then 4.75%	2.22% for 2 yrs then 4.99%	1.99% for 2 yrs then 3.94%	3.65% for 3 yrs then 4.95%
1.79% for 2 yrs then 3.99%	3.95% for 3 yrs then 4.99%	1.69% for 2 yrs then 4.75%	1.69% for 2 yrs then 3.99%	1.85% for 2 yrs the 4.99%	2.69% for 2 yrs then 4.99%	1.79% for 2 yrs then 3.99%	2.95% for 3 yrs then 4.99%
1.89% for 2 yrs then 4.49%	3.49% for 2 yrs then 5.64%	1.15% for 2 yrs then 5.54%	3.14% for 5 yrs then 4.99%	1.69% for 2 yrs then 5.95%	1.95% for 2 yrs then 5.69%	3.65% for 5 yrs then 4.15%	2.29% for 2 yrs then 4.99%
1.70% for 2 yrs then 4.99%	4.99% for 2 yrs then 5.49%	1.84% for 2 yrs then 5.79%	5.04% for 2 yrs then 4.74%	5.58% for 5 yrs then 5.73%	4.99% for 5 yrs then 5.95%	4.29% for 5 yrs then 5.79%	5.99% for 3 yrs then 5.73%
5.59% for 2 yrs then 5.69%	5.69% for 5 yrs then 4.99%	4.99% for 2 yrs then 5.49%	5.45% for 4 yrs then 4.49%	6.09% for 2 yrs then 5.73%	5.25% for 5 yrs then 5.99%	6.09% for 5 yrs then 5.73%	5.24% for 5 yrs then 5.79%

The interest rates presented per trial shown in table 5.1 were exactly as advertised by the individual mortgage lenders. When these rates were applied to the borrowing amounts and loan terms selected for each trial, the fixed rate options consistently yielded higher total repayment costs over the full term compared to the standard variable (SV) and discounted variable (DV) rate options. Thus, in current rate frames, the SVR/DVR options always yielded the most favourable choice (i.e., the lowest total costs). However, the optimality of the SVR/DVR choices only remained so if there was no variability in the fixed and subsequent rates of the mortgages over the full term.

Typically, fixed rate mortgages offered a higher initial rate period and subsequent term rate comparted to SVR and DVR alternatives. This creates the perception that fixed rate mortgage represented the worst choice (i.e., was the most expensive in the long-term). Most lenders offered an initial fixed period in which the interest rate would not alter. However, following the fixed period, the SVR/DVR options were subject to rate increases at the individual lenders discretion increases in rates in addition to predicted base rate rises. The fixed rate options conversely, were not subject to future rate fluctuations following the initial period because the subsequent rate was fixed for the remainder of the term. Thus, when computing the effects of projected base rate increases combined with the probability of lender rate increases, fixed rate options can represent the most optimal choice over the full term. The effects of compound interest rates mean that even slight increases can result in significantly greater monthly and total costs over longer horizons.

Thus, to facilitate optimal choice by creating realistic repayment scenarios, projected base rate growth was built into the future cost framings in condition 2 and 3. This was achieved by increasing the SVR/DVR options by +1.5% in 0.5% increments over three consecutive years beginning in the second year of the mortgage term. The experiment was designed in 2015, which meant that the simulated rate increase took effect from 2016 (hypothetically year two of the loan) through to 2018 (year four) which was in accordance with predicted increases at the time of writing. It is relevant to note that the following future cost estimates were particularly conservative as they were based on the smallest predicted base rate rises in 2015 and did not include any additional rate increases applied by lenders.

For example, figure 5.3 shows that in current rates (with no rate manipulations), a SVR mortgage for £220,00 over 15 years at a fixed rate of 1.69% for 2 years then 4.75% yielded a total repayment cost of £301,951. In future rates however, the total cost increased to £332,219. Following the lenders fixed period of two years at 1.69%, this was generated by adding 1.0% (0.5% + 0.5%) to the subsequent rate of 4.75% in year three of the loan to generate a rate of 5.75% for one year, followed by another increase of 0.5% in year 4 to generate a rate of 6.25% for the remainder of the term. In contrast, a fixed rate mortgage alternative at 1.84% for 2 years, then 5.79% yielded a total cost of £323,973 regardless of rate fluctuations. As this example shows therefore, when considered over the short term, a fixed rate loan with a higher initial monthly cost is suboptimal compared to a SVR with a lower initial fixed rate period. However, when framed over the full term the reverse is shown, with the higher initial rate fixed version being the most effective long-term choice.

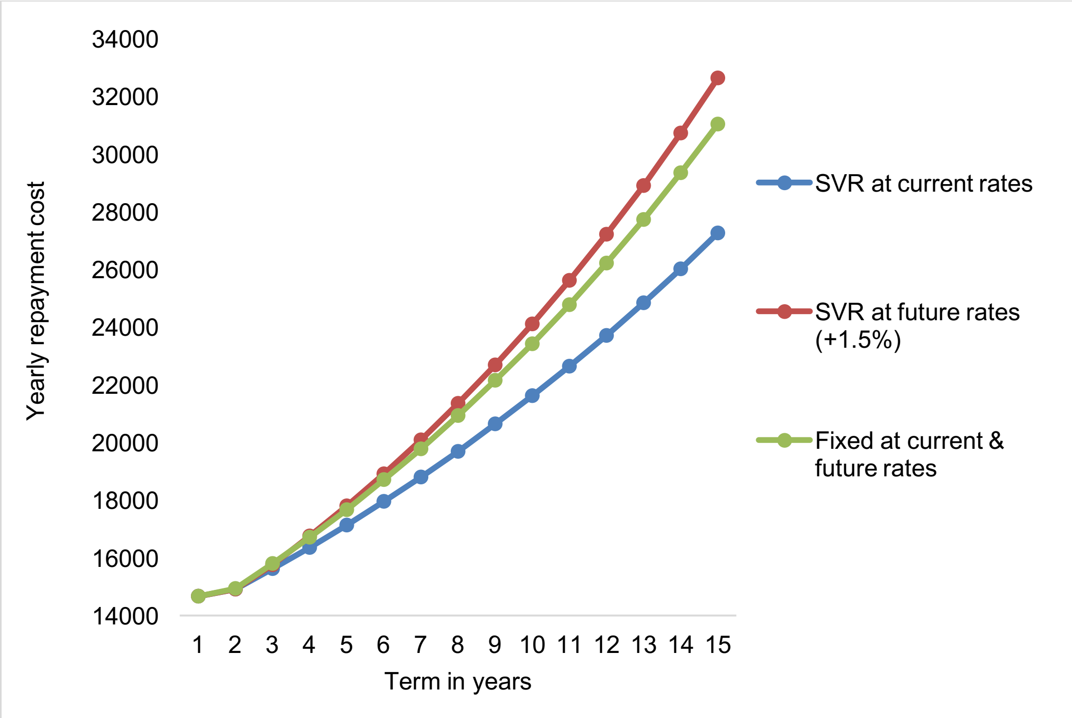


Figure 5.3 Costs for Different Mortgages in Current Vs Future Rates

This example shows the yearly repayment costs for a £220,000 mortgage over a 15-year term at current versus future interest rates for a standard variable (SVR) versus fixed rate mortgage option. In current rates, the SVR option represents the best choice (lowest total repayment cost). However, when accounting for the most realistic future scenario of a minimum base rate increase of +1.5%, the fixed rate options represents the optimal decision, coming in at a lower total repayment cost compared to the SVR.

Two of the trials were designed as ‘attention checkers’ to measure participants level of focus on the task. In these trials, the total loan costs for each of the five mortgage choices remained the same across the current and future rate frames in condition 2 and 3. I.e., the SVR and DVR options were the cheapest, and the fixed rate options were the most expensive in both the current and future rate frames. Thus, the best loan choice did not change making it very easy to detect. Any participant who failed to answer these trials correctly were removed from the dataset.

Procedure

After providing demographic information for age, education and gender, participants were randomly assigned to one of the three conditions in which they underwent the eight unique trials (two of which were attention checkers designed to filter out non-attentive respondents). On the landing page of the experiment participants were given the following instruction:

"In this experiment you will be asked to select the best mortgage you can for a given property price and loan term. There are 8 trials in total and some brief questionnaires at the end. In each trial you will be shown a mortgage price comparison tool with different loan choices. You can toggle the mortgages by different features and screens to help you choose."

In each trial, the following instruction was provided at the top of the screen with the relevant loan amount and term information inserted:

"Imagine you want to buy a house that costs £340,000. You need to get a mortgage for £220,000 over 15 years. Please use the loan calculator below to find the best mortgage you can. To help you make your choice, you can use the toggle keys in the header to switch the order of the mortgages according to each attribute.

You can click on 'current rates' and 'future rates' to compare the repayment costs over time. By clicking on the '?' you can see the APR and interest rate information for each individual mortgage. When you are ready, click on your chosen mortgage in the 'SELECT MORTGAGE' column on the right-hand side to submit your choice."

The instruction remained at the top of the screen for the full duration of each trial with the price comparison tool positioned in the lower half of the screen and no time limit was applied to participants' responses. After submitting a choice, participants were asked to cancel or confirm their choice in order to move to the next trial. Each new trial followed exactly the same format, with the loan amount and term differing in the instruction section at the top of the screen and the information changing within the tool.

After completing all eight trials, a new screen was shown asking participants to indicate which information they used to make their loan choices. Using drop-down menus, participants then indicated whether they used "current rates only", "future rates only" or "both current and future rates" to make their loan choices, and whether "total repayment cost" or "monthly cost" was more important to their choice. After submitting responses to each question, participants were thanked for their participation and provided with instructions for remuneration.

5.2.2 Results

Analysis of Choice Effectiveness Per Condition

Thirteen participants who did not complete all eight trials were removed from the data set, leaving 166 remaining (54, 57 and 55 in condition 1, 2 and 3 respectively). Of the remaining 166 participants, each correctly answered both of the attention checker trials. For the purposes of the experiment, judgments were categorized as either ‘correct’ or ‘incorrect’ based on the minimization of total repayment costs over the full loan term. Participants choices in each condition were coded as ‘correct’ (1) if the target loan was selected in each trial (i.e., the choice which yielded the minimum total repayment cost over the full term) and ‘incorrect’ (0) if any other choice was made.

Performance was thus measured from the standpoint of economic ‘rationality’, whereby choices which acted to maximize utility based on the minimization of total repayment costs, were considered ‘correct’. Judgments which yielded repayment costs anywhere above the minimum failed to maximise utility are were thus regarded as ‘incorrect’ in this context. However, an additional, less stringent measure of judgment performance was also computed to account for the fact that a binary ‘correct/incorrect’ classification may not be entirely relevant in the context of the real-world factors which impact peoples’ financial judgment and behaviour.

To gain an indication of the degree of choice optimality (as opposed to correct vs. incorrect), an interval scale was formed by individually scoring the best choice per trial as 1.0 and the worst was as 0.2 according to the rank order in the future rate frame. This created a monotonic scale of choice effectiveness ranging from 0.2 to 1.0 across both frames. Figure 5.4 shows both the proportions of correct choices and the mean choice scores per condition. This interval scale of judgment performance showed differences in the degrees of choice optimization per condition. For the purposes of analysis, this provided more information regarding the effectiveness of the information manipulations in relation to the control.

It is necessary to note however, that ‘rationality’ in the context of financial choices and behaviour are likely to be influenced by individual circumstance factors which were unaccounted for in the experiment. For example, it may be considered a ‘rational’ judgment for a person to avoid higher costs in the present (and choose to pay more over the long term) in the knowledge that (although currently unable to meet the higher financial demands), they would be financially better off in the future. Thus, in situations such as this, delay discounting may not always be irrational. This may be the case for someone due to acquire income through future inheritance, or a pending job contract or career move, etc. In this sense, more complex mechanisms are likely to underpin the relation between intertemporal choice and the ‘rationality’ of real-life financial decisions than factors related to informational format alone.

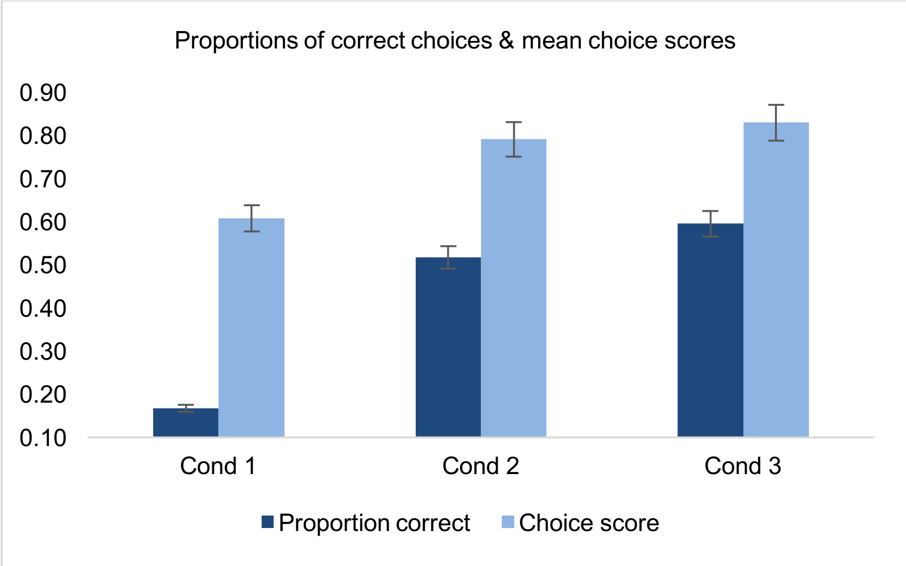


Figure 5.4 Correct Choices and Choice Scores Per Condition in Exp 3a

Proportions of correct choices and choice scores per condition with standard error bars. As expected, the proportions of correct (best) choices and the mean choice scores were significantly greater in the realistic framing condition 3 compared to the control condition 1 and the optimistic framing condition 2. The similarities in the mean choice scores and the proportion of 100% correct choices in condition 2 and 3, and the degree of difference between these two measures indicated that condition 2 and 3 were similarly effective in yielding both fully optimized choices and increasing choice effectiveness in relation to the traditional disclosure condition.

A logistic regression analysis was conducted to predict the proportions of correct choices (1/0) per condition. As expected, condition level (1 to 3) significantly predicted the proportion of correct choices made, (B = 0.94, z = 12.28, p<.001, 95% CI [0.78, 1.08]). Post hoc analyses with Bonferroni adjusted p values showed an increase in the proportions of correct responses from condition 1 through condition 3. Proportions of correct responses in condition 1 (M=0.17) were significantly lower than those in condition and 2 (M=0.52), p<.001, which were significantly lower than those in condition 3 (M=0.6), p<.05.

A one-way ANOVA conducted on the choice scores per condition reflected the results of the logistic regression, showing a significant main effect for condition, $F(1,1325) = 179.9$, p<.001; 95% CI = [.0948, 0.1273]. Post hoc analysis confirmed that choice scores in the realistic framing condition (M=0.83) were significantly higher than those in the optimistic framing condition (M=0.79), p<.05., which were significantly greater than those in the control condition (M=0.61), p<.001. Figure 5.5 shows the proportion of choice scores per condition to illustrate the distribution of total choice effectiveness across the range from 0.2 (worst) to 1.0 (best). In addition to the proportion of correct choices, the choice score dependent variable is useful in showing the dissemination or distribution of choice effectiveness per condition compared to the binary correct/incorrect outcome. This measure of performance is important because it shows the degree of choice effectiveness which is not indicated when responses are assessed in terms of 'correct' (fully optimized) versus 'incorrect' (i.e., all other options). Assessing choice performance on a continuum in this way therefore reveals more information regarding the effects of each manipulation.

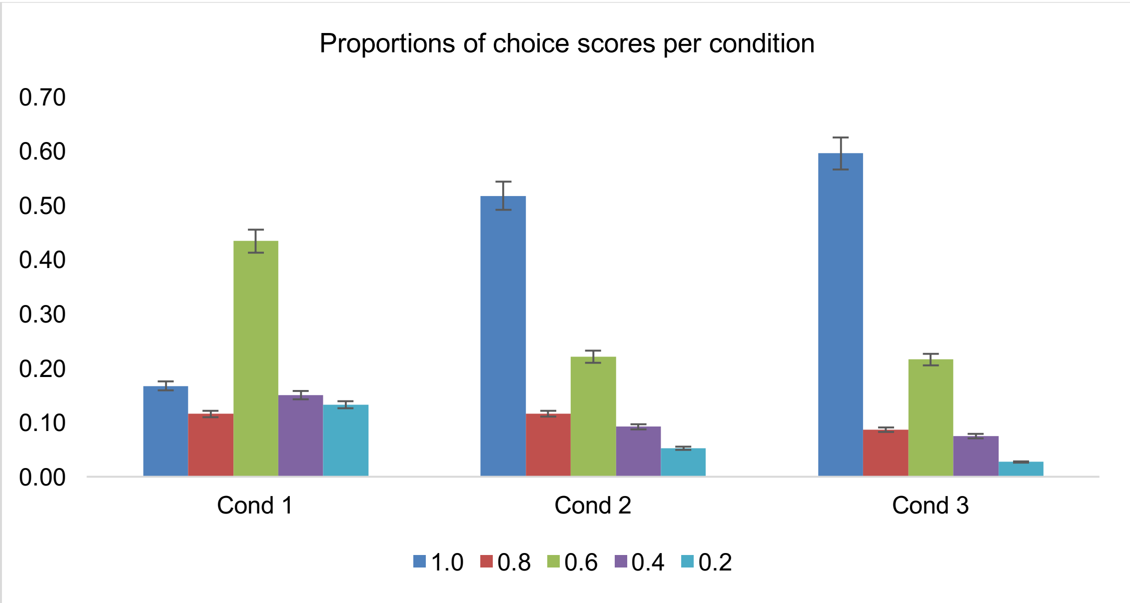


Figure 5.5 Proportion of Choices Per Score and Condition in Exp 3a

The proportions of choices displayed per score in each condition with standard error bars. The majority of choices in the control condition achieved a score of 0.6. This indicates that when presented with the standard industry data format, participants tended to make choices which were at a 50% level of effectiveness. However, when presented with rate frames in the optimistic framing condition 2 and the realistic framing condition 3, the majority of choices were made at a 100% level of effectiveness, reflected by a choice score of 1.0. Overall, the realistic framing condition (3) yielded the highest proportion of fully effective choices within and between conditions.

The significantly higher choice scores in condition 2 and 3 therefore suggest that the framing manipulation positively impacted choice optimality. To explore the difference in choice effectiveness between the current and future rate frames in condition 2 and 3, a logistic regression was conducted firstly on the proportions of choices made per frame per condition (results are shown in figure 5.6).

Current Vs Future Frame Choice Proportions

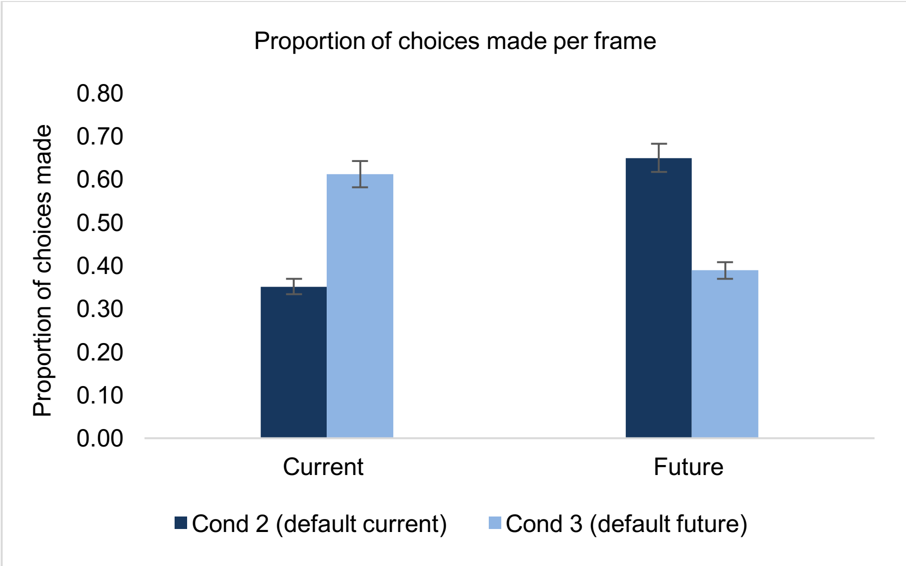


Figure 5.6 Proportion of Choices in Current Vs Future Frame in Exp 3a

The proportion of choices made in the current and future frame in the optimistic framing condition 2 and the realistic framing condition 3 with standard error bars. Significantly more choices were made in the opposite frame to the default in both conditions, indicating that participants in each condition tended to click through to the alternative frame before making a choice. In condition 2, 30% more choices were made in the future frame compared to the default current frame, and in condition 3, 22% more choices were made in the current frame compared to the default future frame.

Results of a logistic regression to predict proportions of choices made in the current (0) and future frame (1) per condition showed that the proportion of choices made in the future frame significantly differed between conditions, (B = -1.06, z = -7.71, p<.001, 95% CI [-1.34, -0.79]). Post hoc analysis confirmed that a significantly higher proportion of choices were made in the future frame in the optimistic framing condition (2) (M=0.65) compared to in the future frame in the realistic framing condition (3) (M=0.39), t(891.11) = 8.07, p<.001, and a significantly higher proportion of choices were made in the current frame in condition 3 (M=0.61) compared to in the current frame in condition 2 (M=0.35), t(891.75) = -4.63, p<.001.

Choice Scores Per Rate Frame

Figure 5.7 shows the mean choice scores per frame and condition. A linear regression model with choice score as the dependent variable and condition and the frame in which the choice was made as factors in the model showed that choice scores were significantly higher in the realistic framing condition 3, $R^2 = .05$, $F(3,892) = 16.47$, $p<.001$, $B = -0.061$, $t(892) = -2.58$, $p<.01$; 95% CI = [-0.10, -0.01] and significantly differed between the chosen frame per condition, with higher scores yielded in the current frame across conditions, $B = -0.461$, $t(892) = -5.52$, $p<.001$; 95% CI = [-0.6254, -0.2973]. The interaction term was found to be significant due to an overall higher mean choice score in the current rate frame in the optimistic framing condition 2, $B = 0.15$, $t(892) = 4.71$, $p<.001$.

Planned t-tests showed that mean choice scores were significantly higher in the realistic framing condition 3 (M=0.83 vs M=0.79), $t(891.94) = -2.40$, $p<.05$. Between conditions, future frame choice scores were significantly higher in condition 3 (M=0.83) compared to the optimistic framing condition 2 (M=0.74), $t(366.44) = -3.8375$, $p<.001$, whereas current frame choice scores were significantly higher in condition 2 (M=0.90) compared to condition 3 (M=0.83), $t(355.53) = 2.9194$, $p<.01$. Within conditions, choice scores were significantly greater in the current frame in condition 2 (M=0.89 vs M=0.74), $t(396.93) = 6.94$, $p<.001$, in condition 3 however, no difference in choice scores was found between the future (M=0.83) and current frame (M=0.83), $t(328.53) = -0.06$, $p=0.95$. This indicated that despite there being more choices made in the non-default frame in condition 3, there was no significant difference in the effectiveness of the choices made in either frame. In contrast, condition 2 yielded a higher proportion of more effective choices in the non-default frame, underpinning the significant interaction term in the regression model.

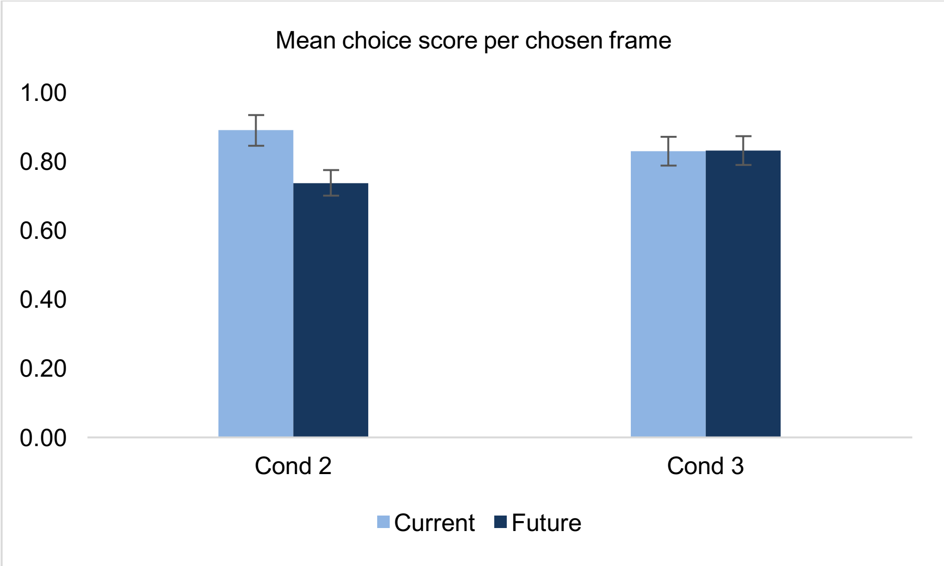


Figure 5.7 Mean Choice Score Per Frame in Exp 3a

The mean choice score per frame for each condition with standard error bars. Although there was a larger proportion of choices made in the alternative frame to the default in both conditions (figure 5.6), the effectiveness of the choices in the optimistic framing condition 2 was greater in the default (current) frame. This suggests that when two comparisons of the rate frames were made (i.e., by clicking to view rates in the non-default frame, then back again), the resulting choices were more effective compared to when only one comparison between the current and future rates was made (i.e., clicking to view the non-default then making a choice). In the realistic framing condition 3, choice scores were significantly higher than those in the optimistic framing condition 2, however, within condition 3, there was no significant difference between choice scores in either frame. This indicates that the number of comparisons made in condition 3 might have been less important to choice effectiveness than in condition 2.

Correct Choices Per Rate Frame

A logistic regression was conducted to predict the proportions of correct (1) and incorrect (0) choices made per chosen frame in each condition. Results (displayed in figure 5.9) showed a significant effect of condition (B = -0.64, z = -2.99, p<.01, 95% CI [-1.07, -0.22]) and frame (B = -4.39, z = -5.82, p<.001, 95% CI [-5.88, -2.92]) on the proportion of correct choices made, and a significant interaction term, (B = 1.51, z = 5.19, p<.001, 95% CI [0.95, 2.09]). As expected, these results reflect the pattern of choice scores shown in figure 5.7, with post hoc analysis indicating that a significantly higher proportion of correct choices were made in condition 3 (M=0.60) compared to condition 2 (M=0.52), t(893.72) = -2.35, p<.05, with more correct choices made in the current frame (M=0.63) compared to the future frame (M=0.48) across conditions, *t*(891.9) = 4.57, *p*<.001.

Between conditions, there were significantly more correct choices made in the future frame in the realistic framing condition 3 (M=0.62) compared to the optimistic framing condition 2 (M=0.41), *t*(357.85) = -4.57, *p*<.001, and significantly more correct choices in the current frame in condition 2 (M=0.73) compared to condition 3 (M=0.58), *t*(360.64) = 3.12, *p*<.01. Moreover, current frame correct choice proportions in condition 2 (M=0.73) significantly exceed future frame proportions in condition 3 (M=0.62), *t*(328.91) = 2.05, *p*<.05, and vice versa for condition 3 current frame proportions (M=0.58) versus condition 2 future frame (M=0.41), *t*(557.28) = 4.20, *p*<.001. Within conditions, correct choice proportions were significantly higher in the current frame in condition 2 (M=0.73 vs. M=0.41), *t*(352.96) = 7.0, *p*<.001, however, no difference was found in the proportions of correct choices between the current (M=0.58) and future frame (M=0.62) in condition 3, *t*(366.19) = -0.83, *p*=.40.

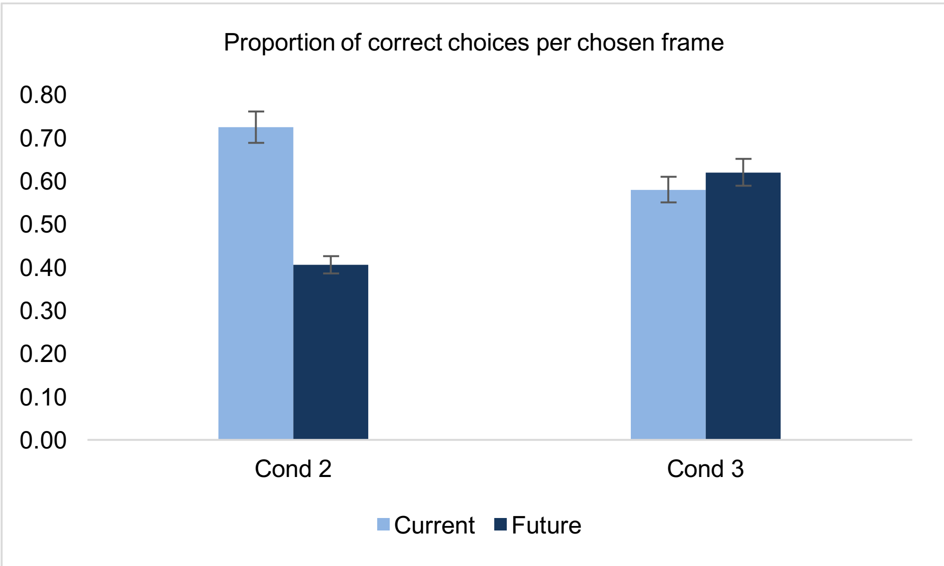


Figure 5.8 Proportion of Correct Choices Per Frame in Exp 3a

The proportion of correct choices made per chosen frame in each condition with standard error bars. Overall, a higher proportion of correct choices were made in the realistic framing condition 3 and there was no difference in choice effectiveness between the frames in which the choices were made in condition 3. In the optimistic framing condition 2, a higher proportion of 100% optimal choices were made in the default frame, suggesting that making two (or more) comparisons of the data by clicking back to the default frame again before making a decision (i.e., current-future-current viewings), increased the propensity to make 100% optimal choices. The difference between the proportions of 100% optimal choices in the default and non-default frame in condition 2 was also greater than the difference in choice scores between the frames. This indicated that making more than one comparison of rates increased the propensity to make a choice that was 100% optimal compared to a choice which was more effective only to some extent. In condition 3, the lack of difference in proportions of correct choices between frames indicated that both frames were equally effective, thus choice optimality was less dependent on making multiple alternations between screen rate frames.

Analysis of Rate Frame and Repayment Data Usages

Analysis of the information participants provided at the end of the trials indicated that there was no difference between their use of "total repayment cost" or "monthly cost" when making loan choices in the optimistic framing condition 2, $t(57) = 0.52$, $p=.60$, or in the realistic framing condition 3, $t(55) = 0.79$, $p=.43$. Moreover, the use of "both current and future rates" was significantly greater compared to use of "current rates only" in condition 2, $t(57) = 10.30$, $p<.001$ and 3, $t(55) = 13.70$, $p<.001$, and use of "future rates only" in condition 2, $t(57) = 6.55$, $p<.001$ and 3, $t(55) = 9.89$, $p<.001$.

These results indicated that the majority of participants in both conditions viewed the loan information in both the current and future frame before making a choice. Combined with the above results, this supports the view that the most effective choices per condition were made when participants made two or more comparisons of the data before making a selection. I.e., after clicking to view the alternative frame to the default view in each condition, they then clicked back to the default frame before making a decision.

This process of comparison between data frames might have occurred multiple times (the data relating to the number of screen transitions was not collected). However, consistent with the responses to the final questions, it seems likely that an increase in the number of screen transitions (i.e., multiple data comparisons) was associated with increased choice effectiveness. If it were the case that the higher proportion of correct choices in the default frames resulted from participants making a choice based on viewing the data in that frame only without clicking to view the alternative frame, we would expect to have found higher proportions of "current rates only" and "future rates only" responses compared to "both current and future rates" when prompted to indicate which data frames people used to form their decisions.

5.2.3 Discussion

As expected, mortgage rates framed in concrete currency costs presented in current versus future rates over the full term significantly increased choice effectiveness compared to mortgage data presented in a standard industry price comparison format. The framing manipulations in both the optimistic framing condition 2 and the realistic framing condition 3 yielded a higher proportion of correct choices and higher choice scores compared to the standard industry disclosure which involved the information presented in APR and rate format.

Overall, choice scores were highest in condition 3 and a larger proportion of fully optimal choices were yielded compared to condition 2. This result supports the predicted effectiveness of condition 3 based on the simultaneous presentation of the rate alternatives combined with the future rate frame default view. In both conditions, a higher proportion of choices were made in the non-default frames which suggests that the data formats encouraged participants to view the alternative frames once or multiple times before making a choice. In condition 3, there was no difference between the current and future rate frame in either choice scores or the proportions of correct choices, whereas in condition 2, choice scores and proportions of correct choices were greater in the default (current rate) frame.

Whilst the number of transitions participants made between screens was not recorded, the results suggest that, rather than simply selecting the first (lowest) total cost viewed when clicking to see the future rates (in the non-default view) in condition 2, the best choices were made by people when they returned to the current rate frame (the default view) before making a selection. Although it is likely that the best decisions were based on multiple current versus future rate comparisons, it is clear that returning to the default frame was important to choice effectiveness in the optimistic framing condition 2. It is likely that the sequential presentation of the rate alternatives (i.e., the separate screen views) is what underpinned the increase in rate frame comparisons, leading to higher rate of optimal decisions being made in the default view.

The sequential presentation of the data frames in condition 2 meant that less information was provided per screen which potentially simplified the process of data synthesis. However, it also required successive alternations between screens to perform the comparative analysis of the choice alternatives. When using the toggles to rank loans by different attributes and in different frames, it may have been more difficult to identify how the costs changed in relation to the rate frames in the sequential display in condition 2 compared to in condition 3 where the cost and rate changes were viewed simultaneously. Thus, despite less data being shown per frame in condition 2, the effort necessary to place the data in context and effectively evaluate the alternatives (i.e., the necessity to switch screens multiple times) is likely to have increased task demands, leading to increased attentional load and resulting in a reduction in judgment performance.

The increased effectiveness of condition 3 indicated that the ability to view and compare the rate alternatives side-by-side (i.e., in the simultaneous presentation format) was likely to have facilitated evaluative judgment to a greater extent than in condition 2 where the alternatives could only be compared by clicking to access the current and future rates on separate screens. It is also possible that the future rate default set in condition 3 was effective in anchoring participants on the best choice prior to them making any comparisons between rates and product types. In this respect, the default would have acted to enhance the facilitative effect of displaying the rates and product information simultaneous. Within the context of the less effective (current rate) options, presenting people with the highest ranking (most optimal selections) first is likely to have effectively communicated how the costs for fixed versus variable rate product types altered and re-ordered within each rate frame and how this impacted the rank order of product effectiveness. Disclosing the re-ordering of the loans based on rate differences side-by-side is thus likely to have facilitated comprehension of how and why different product types significantly altered in their level of optimality when considered over the full loan term.

In addition to the use of defaults to create anchors, simultaneous data presentations and disclosing percentages and rates in absolute terms, there is also evident to suggest that disclaimers can be effective in improving peoples’ choice of investment options in percentage formats by increasing peoples’ sensitivity to fees (Newall, 2016b). This suggests that a disclaimer could also be effective in improving mortgage product choices when the data is presented in standard industry APR and rate formats. In particular, disclosure manipulations are shown to be more effective in aiding consumer financial judgment when they are delivered in short, simplified formats (e.g., Bhargava & Manoli, 2013), and are aimed specifically at encouraging people to compare different options and individual attributes (e.g., Salisbury, 2011).

From this perspective, it is possible that simply heightening peoples' awareness of rate variability using a disclaimer may be sufficient to improve judgment performance by motivating greater cognitive effort and attentional resources. Increasing peoples' sensitivity to financial risk may therefore activate more complex judgment processing involving a higher cognitive load and the application of attentional resources which could leading to more comprehensive and effective comparative analysis of individual cues and attributes. One possible explanation for the results shown in the standard industry format condition therefore, is that when presented with complex data in percentage formats, people tend to default to frugal judgment strategies which minimize cognitive effort and resources to yield quick decisions. These strategies thus result in format biases, additive processing techniques and the tendency to assume linear relation which can result in poor judgments and choices. However, when provided with a stimulus to evoke increased effort and attentional resources, such as being explicitly warned about the implications of future rate increases, peoples' judgment may improve in complex environments, even where information is framed in percentages.

It may be the case therefore, that format biases and the tendency to arithmetically compute values are not the primary factor underpinning poor financial judgment in all contexts. For example, people may possess sufficient attentional capacities and may be capable of accurately processing and comparing percentage data in complex environments if there is a strong motivating factor for increasing the cognitive effort and time people invest in the judgment process. It is therefore necessary to test whether prompting people to consider future rate costs using a short, informational disclaimer can be effective in increasing judgment performance when data is disclosed in both absolute currency costs (i.e., using the same data framing manipulations tested in the current experiment) and in standard industry formats. This would provide evidence for whether format biases and inherent limitations in attentional capacities underpin poor performance, or whether it may be possible in some contexts to improve judgment effectiveness based on increasing attention and cognitive effort, despite percentage and rate information.

For example, if it is the case that judgment performance is not characterized entirely by format biases, we may see an increase in judgment performance in the standard industry format condition with the addition of a disclaimer which warns people to carefully consider their choices based on the financial implications of future rate increases. Thus, despite percentage and rate formats, drawing peoples' attention to the risks associated with low rate choices may be successful in motivating them to apply more cognitive effort and manage a greater attentional load leading to a more optimal decision than in the absence of a motivational prompt. People may therefore be capable of undertaking a higher cognitive load, but are predisposed to engage in low effort cognitive processes to increase judgment speed and efficiency in complex environments where there are multiple attributes and it is difficult to weight the importance of individual cues in the judgment task.

Before discussing other factors which may shape peoples' financial choice and behaviour, it is important to identify potential limitations in design which could have impacted results. The purpose of the experiment was to examine the impact of different informational manipulations on peoples' propensity to make economically rational financial choices (i.e., those which act to maximize utility over the long-term), based on the particular data and scenario presented. From this perspective, it is probable that people would have processed the data in accordance with the aim of minimizing total repayment costs. However, with respect to real-world interest rate variability, peoples' decisions may have been influenced by the knowledge that (although predicted to rise) the base rate has historically remained low and predictions for increases have not actualized to date. It is therefore possible that people will base financial decisions on the belief that interest rates are unlikely to rise, or to an extent great enough to create negative financial impacts. Although this perspective does reflect an optimism bias, it may represent an adaptive response to historical trends in interest rates.

Similarly, people may have processed the information in accordance with the knowledge that it is possible to switch between mortgages following initial fixed rate periods. This could enable the extension of a lower borrowing rate and would thus lead people to opt for the lowest cost option in current rates (which would represent the least effective choice in the experiment) in the belief that they could change products or providers once the low rate period ceased. This strategy may therefore be perceived as a 'rational' strategy for some individuals in a real-world setting. However, it is still the case that many borrowers do still take out long-term mortgages and are subject to the rate increases following the fixed rate periods.

As emphasized above, the aim of the experiment was to elicit the effects of data manipulations on the economic rationality of choices (i.e., those which act to maximize utility over the long-term), based on the hypothetical base rate increases applied to the lenders rates following the fixed periods. Although there was the potential to yield ineffective choices in the context of the experiment which may be otherwise be considered 'rational' in the context of individual circumstances, knowledge and beliefs etc., it was not possible within the confines of the experiment to capture the impact of individual knowledge etc. Future work focusing on these variables would thus provide useful insights into how such factors influence financial behaviour and how financial 'rationality' may be determined and defined from an individual perspective within a given context or real-life setting.

Other factors which could impact financial choice performance are individual differences in abilities such as financial literacy and numeracy which could interact with the framing manipulations tested in experiment 3a in particular ways. Findings show for example, that downside financial risk judgments positively correlate with financial literacy and numeracy (Newall, 2016a). Numeracy and risk literacy are found to relate to poor statistical inference and incomprehension of probability formats in tasks assessing peoples' judgments of medical risk (e.g., Cokely, Galesic, Schulz, Ghazal & Garcia-Retamero, 2012). High and low numeracy is also associated with over and underestimates of loan repayment amounts (Soll, Keeney & Larrick, 2013) and people higher in financial literacy are shown to be more likely to select less optimal, high fee mutual funds (Newall, 2016a). It is apparent therefore that the findings relating to the effects of financial literacy and numeracy are somewhat mixed. Some evidence suggests that increasing financial literacy is not effective in improving financial behaviour as it is likely that factors such as numeracy mediate the relation between literacy and behaviour (Fernandes, Lynch & Netemeyer, 2014). In light of these findings, it is thus important to include these individual difference measures when examining the possible mechanisms involved in how the framing effect identified in experiment 3a is effective in promoting judgment performance.

Another potentially relevant individual difference variable is temporal preference in the form of optimism. At extreme levels, trait optimism is associated with suboptimal financial judgment and behaviour, however, moderate levels are positively associated with more optimal financial behaviours, mediated by increases in self-control (Puri & Robinson, 2007). It is therefore possible that in this particular financial choice context, a high temporal preference (or lower 'self-control'), exhibited as the propensity to opt for low current costs and thus favor the present based on delaying higher costs into the future, reflects a harmful optimism bias.

In this view, higher optimism may correlate with lower choice performance based not only on the tendency to opt for low current costs, but also on percentage format biases and the tendency to linearly extrapolate rates, leading people to underestimate future costs. Thus, combined with format biases leading to linear judgments, optimism could be a strong factor involved in suboptimal financial choice, underpinning peoples' propensity to downplay the risks associated with future rate increases, and overestimate their ability to meet increased financial demands in the future. It is therefore a possibility that an interaction between the framing effect in experiment 3a and optimism exists. For example, if a significant relation between optimism and choice performance exists in the realistic framing condition 3 whilst the framing effect remains robust (i.e., choice performance remains high in condition 3 compared to the control), this may suggest that the framing manipulation is effective in counteracting harmful optimistic tendencies.

To test possible interactions between the framing manipulations and individual difference factors, and whether choice performance is related to inherent format biases (as opposed to attention and cognitive effort) in this particular context, experiment 3b is conducted as a replication study to test the robustness of the framing effect identified in experiment 3a. Alongside measures of financial literacy, numeracy and trait optimism, experiment 3b involves the repetition of conditions 1, 2 and 3 with the addition of a disclaimer akin to a real-world disclaimer. The financial services industry provides disclaimers on product websites to warn consumers of potential financial risks, presumable with the purpose of guiding people towards more optimal choices. The manipulation in experiment 3b is designed to replicate such disclaimers, warning about the effects of future rate variability and to carefully consider choice alternatives.

As described above, the mechanisms by which the disclaimer may promote more optimal choice could be based on evoking more effective comparative analysis of the individual product rates for the fixed versus variable rate alternatives. If sufficiently motivated to expend more cognitive effort and attentional resources, people may be capable of synthesizing data effectively, even in percentage and rate formats. From this perspective, percentage format biases and the inability to extrapolate compound interest rates might not underpin poor financial judgment in all contexts, or to the extent that the results of experiment 3a indicate. (I.e., it is possible that effort as opposed to format biases underpin poor choice in some domains). Adding the disclaimer to all three conditions in experiment 3b examines this possibility by enabling the assessment of changes in choice effectiveness in the control (standard industry format) compared to the optimistic framing condition 2 and the realistic framing condition 3. Improvements in decision making in the control would thus indicate that increasing the propensity to evaluate choice alternatives by simply reminding people of the risk associated with overly optimistic choices (i.e., those based on the lowest rate options in the present), could be enough to improve financial decision making in online price comparison environments.

5.3 Experiment 3b Behavioural Disclaimer Vs Interest Rate Framing Effects on Financial Choice

Experiment 3b is conducted as a replication study to test the robustness of the framing effect identified in experiment 3a in the context of a disclaimer and measures of relevant individual difference factors. Although experiment 3a showed that the framing manipulation was effective in improving peoples' financial judgment compared to information disclosed in standard industry APR and rate formats, it was unclear what the mechanisms were that underpinned the effectiveness of the framing manipulation. The premise for reframing rate data in concrete (currency) values and displaying them in current versus future rate alternatives was based on the body of evidence relating to percentage formats biases and the strength of the linear judgment heuristic shown to be particularly prevalent throughout financial judgment and choice domains.

This rationale assumes that limited attentional capacity coupled with the tendency to additively process percentages, apply linear functions and possess and an implicit and explicit optimistic bias, are what lead people to form ineffective judgments in financial contexts. It is important however, to determine that these biases and processing strategies are accurately describing peoples' cognitive characteristics and that manipulations designed to mitigate these tendencies are thus promoting human rationality via the most relevant and effective mechanisms. It is possible for example, that other judgment strategies, or decision making motives and drivers may be involved in complex financial judgment contexts which were not explored in experiment 3a. To address these shortcomings, experiment 3b investigates the potential that people may in fact possess the necessary computational and attentional capacities to form effective judgments in complex numerical data settings, if provided with a stimuli sufficient to motivate increased cognitive effort and activation of more comprehensive processing strategies.

From this perspective, it is possible that the nature of informational framings and communications are important to judgment performance by providing the motivational factors necessary to evoke activation of more complex and resource heavy cognitive processing. In this respect, suboptimal judgments which are shown to stem from fast frugal heuristic strategies (e.g., the tendency to assume linear relations) may be associated with limitations in effort based on low motivation as opposed to strict cognitive limitations and computational problems related to numerical formats. Furthering our understanding of the cognitive biases in numerical judgment is thus necessary to determine the mechanisms underpinning human rationality in different groups and settings where different motivational and behavioural factors are involved.

To examine this possibility, experiment 3b repeats the trials from experiment 3a with the addition of a disclaimer designed to increase the attention and effort people apply to the selection task by warning them about the risk of future rate variability and the significant impact that rate increases will have on their repayments. If peoples' judgment effectiveness increases in the standard industry format condition (involving data framed in percentages and rates) when presented with the disclaimer, this would provide evidence that peoples' judgmental abilities may not limited entirely in the way that findings suggest, i.e., shaped by format biases, over-optimism and restricted attentional capacities.

One possibility for example, is that people apply default processing strategies in everyday choice situations which are perceived as having non-critical outcomes. Such default strategies are likely to be based on cognitively frugal heuristics involving arithmetic processing and the application of linear functions as an efficient means of yielding fast, 'satisfactory' judgments. However, behavioural or social comparison prompts (e.g., Newall, 2016b) could be effective in improving judgment performance in contexts involving percentage and rate information. When rationale for deeper level processing is provided (e.g., by heightening peoples' sensitivity to financial risk), the motivation to avoid losses may stimulate engagement in more detailed comparative analysis of the available data, requiring greater activation of attentional resources in an effort to increase choice optimality.

If, on the other hand, no difference in choice performance is identified with the presence of the disclaimer in the standard industry format condition, this would support the rationale for the framing manipulation, indicating that format biases and limited computational and attentional capacities are likely to be the key factors underpinning poor performance in financial choice. The aim is therefore to examine whether choice performance remains constant between experiment 3a and 3b, thus indicating whether judgment effectiveness is related specifically to the framing effect and thus percentage format biases. Finding no difference in choice effectiveness between experiment 3a and 3b would demonstrate the robustness of the framing effect in targeting and mitigating biases which impact human probabilistic inference in complex numerical data fields.

The standard industry disclaimer, “Your home may be repossessed if you do not keep up repayments on your mortgage” found on all mortgage lender websites is presumably designed to draw borrowers attention to the importance of being able to meet monthly repayment costs. The ability to meet future costs is dependent however, on effectively controlling for risk by forming rational financial judgments and choices in the present which account for future eventualities. Findings suggest however, that such disclaimers are no more effective in improving financial judgment than excluding disclaimer information altogether.

In the cases where behavioural disclaimers have proven effective, psychologically informed disclosures conveyed in short, simplified informational formats are shown to facilitate improved judgment and behaviour (Loewenstein, Cass, Sunstein & Golman, 2014). Based on this effect, the following disclaimer information was added to each condition in experiment 3b with the aim of increasing choice effectiveness through promoting the propensity to consider the potential impact of future rate hikes:

“When choosing a mortgage, bear in mind that rates can rise considerably, fall, or remain the same as today. Fluctuations can significantly affect your repayments”.

Comparing the effect of the disclaimer on choice performance in each condition with the results of experiment 3a tests the robustness of the framing effect by examining whether it is necessary for performance enhancement, or if choice effectiveness can be increased behaviourally using a disclaimer to increase consideration of the risk associated with low rate choices. This is of specific interest in the control condition where the data is disclosed in percentage and rate formats. Encouraging more detailed comparative analysis of rates and product types in standard formats could mitigate the optimistic tendency to heuristically select loans based on the lowest current rate, disregarding rate variability over the long term and the costs differences between product types.

Like the framing effect identified in experiment 3a, it is likely that any positive effect of the disclaimer on judgment performance will be associated with minimizing optimistic biases in peoples' financial choice. For example, the framing manipulation in experiment 3a was designed to impact percentage format biases and the inability to extrapolate compound interest rates which leads people to make optimistic (suboptimal) financial choices based underestimation of future repayment costs. The tendency to linearly project rates, therefore indicates an insensitivity to rate variability and the financial consequences of future rate rises. Although the mechanism used by the disclaimer to impact choice performance differs to the framing effect, it is still designed to improve choice by decreasing optimism based on drawing attention to the potential for high future costs, specifically for product options which have the lowest costs in current rates.

Experiment 3b therefore addresses the research question of whether the disclaimer manipulation can also be effective in improving financial choice via mitigating optimistic biases which lead people to downplay the risk of future rate variability and overestimate their capability to make higher repayments in the future. If the disclaimer is found to be effective in increasing choice performance in the control condition for example, this would suggest that poor financial judgment may be related more to optimistic biases in judgments of rate variability, rather than to format biases and the tendency to linearly extrapolate rates. The alternative view therefore, (based on the results from experiment 3a), is that optimism in this particular context is associated with percentage format biases which are successfully mitigated by the framing effect, shown to be robust when compared to the effects of the disclaimer.

Based on the saliency of percentage format biases and the strength of the tendency to linearly extrapolate non-linear trends identified in chapters 3 and 4, it is predicted that the results from experiment 3a will hold. Specifically, there is expected to be no difference in choice effectiveness in the standard industry format condition in experiment 3b with the addition of the disclaimer compared to experiment 3a, and choice performance is predicted to remain the least effective in the standard industry format and the most effective in the realistic framing condition 3 with the addition of the disclaimer.

With respect to individual difference factors, it is expected that there will be a significant association between trait optimism and choice performance, with higher optimism relating to lower performance in accordance with Puri and Robinson's (2007) findings relating to the effects of optimism on financial habits and choice. Optimism is also predicted to interact with frame preference in the framing manipulation conditions. This prediction is derived from the view that higher self-control instantiates a preference for higher future rewards (or a lower temporal preference) which is associated with more prudent financial choices and behaviours leading to increases in financial utility. In this sense, lower optimism may be associated with a tendency to delay gratification and forgo short-term rewards in the form of lower loan costs in the present and opt for higher cost choices based on future rates which represent the most effective options in the long term.

To assess trait optimism in experiment 3b, the Life Orientation Test - Revised (LOT-R) (Scheier, Carver & Bridges, 1994) is used as a means of differentiating between an optimistic bias which can vary from one setting to the next (Weinstein, 1980) and dispositional optimism which characterizes an individual's general propensity to assume positive outcomes. When applied to financial risk judgment, optimism is viewed as an indicator of temporal preference, indicative of the propensity to assume positive future outcomes and discount future costs based on individual differences in self-control (Puri & Robinson, 2007). Temporal preference in financial decision making has also been examined using measures such as the CFC (Consideration of Future Consequences; Strathman, Gleicher, Boninger & Edwards, 1994) designed to assess predisposition toward consideration of future consequences. However, the CFC scale has not been shown to significantly correlate with financial behaviours in studies assessing credit repayment decisions (e.g., Navarro-martinez et al., 2011; Salisbury, 2014). In accordance with Puri and Robinson (2007), the LOT-R was therefore applied in experiment 3b to address the question of whether higher levels of optimism relate to lower choice performance and how optimism may interact with the disclaimer and framing manipulations.

To consolidate the rationale for experiment 3b, it is useful to recap the findings thus far. The previous chapters have shown that the linear prediction bias underpins poor financial decision making (e.g., misjudgement of borrowing costs) based on underestimating (i.e., arithmetically processing) non-linear compound interest rates (Stango & Zinman, 2009). The human propensity to apply fast, efficient heuristic strategies when forming numerical decisions results in the simultaneous (additive) mental processing of percentage and rate information. This generates the bias to treat all numerical data as absolute values and to assume linear relations in the world around us which can lead to probabilistic judgments based on linearly projecting trends. In the case of compound interest rates and other behavioural outcomes which follow non-linear functions (e.g., smoking cessation or exercise behaviour), the propensity to form linear judgments and predictions is therefore likely to limit rationality, creating a negative impact on choices and behaviours based on people making erroneous underestimations of future costs and rewards.

In the case of financial judgement, this tendency to process additively and assume linearity is expressed as a bias to make optimistic choices which underestimate interest costs and do not effectively account for the risk associated with future rate rises and variable rate loan product choices. It is therefore important to assess optimistic tendencies in this context because it will indicate whether the biases in financial judgment are associated with explicit, general optimistic biases, or whether erroneous financial judgments are implicitly optimistic, i.e., stemming from numerical format processing errors as opposed to formal trait optimistic biases. Whether optimism effects judgment directly as a behavioural propensity, or whether it impacts choice performance indirectly via percentage format biases, the influence on financial behaviour is likely to be negative. Thus, in both cases, the framing of rates in concrete values, disclosed in current versus future alternatives is likely to remain a robust corrective measure, benefitting judgement rationality.

In sum, experiment 3b is conducted to evaluate the replicability of the framing effect identified in experiment 3b to assess the judgmental mechanisms underpinning performance in financial choice domains. By adding a stimulus in the form of a disclaimer to motivate cognitive effort and increase attentional resources, experiment 3b assesses whether limitations in performance where percentages and rates are concerned are associated directly with format biases, or whether people do possess the necessary computational abilities to formulate effective judgments where rates are concerned, if sufficiently motivated. It is expected that the framing effect will remain robust based on no difference in choice performance with the disclaimer, thus supporting the rationale for format biases, judgment linearity and the inability to synthesize multiple cues as the key mechanisms underpinning poor performance in complex financial choice contexts.

5.3.1 Method

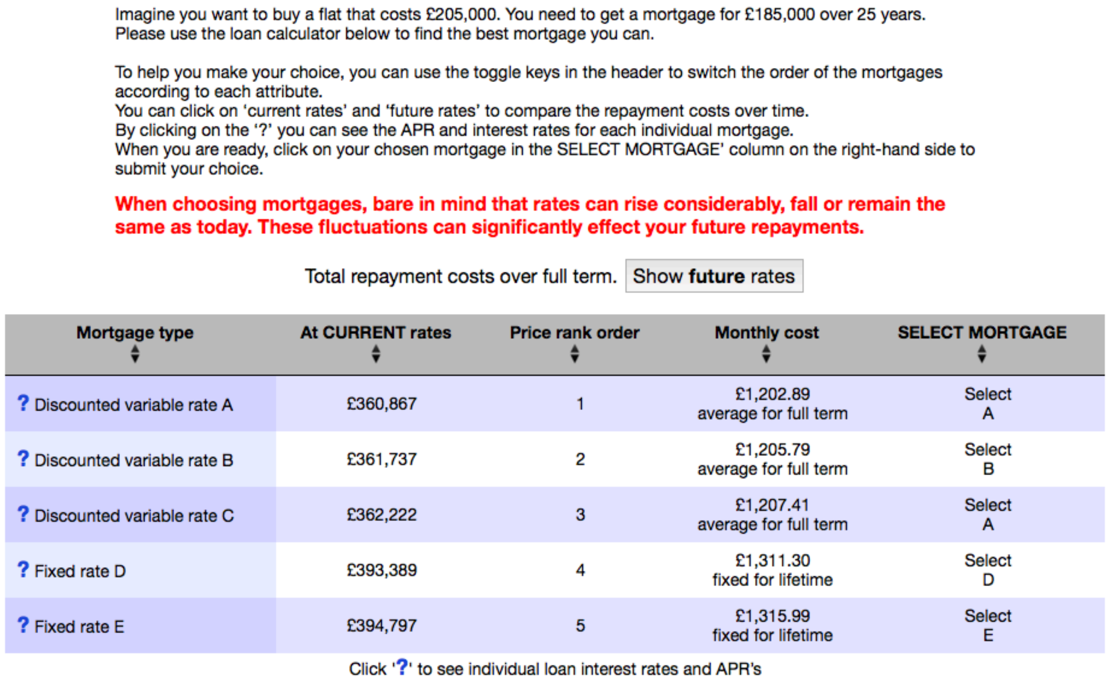
Participants and Design

One-hundred and eighty-three participants aged over 18 years were recruited again via Amazon Mechanical Turk and paid $2.00. All participants were unique in each experiment, no respondent from experiment 3a repeated the trials in experiment 3b. The average age was 36.05 years (SD = 10.36, range = 2 to 62 years), 46.4% were female and 34% were educated to a minimum level of a Bachelor's degree. The same randomized controlled trial design was employed as in experiment 3a, with participants randomly assigned to either the standard disclosure condition (control), interactive/optimistic framing (condition 2), or interactive/realistic framing (condition 3) and trials were randomized per participant in each condition.

Materials and Procedure

After providing demographic information for age, gender and education, participants were then directed to the eight mortgage choice trials. In each trial in all three conditions, the instructions remained the same as they were in experiment 3a, except with the addition of the following behavioural disclaimer manipulation to encourage closer consideration of rate changes over time. As shown in figure 5.15, the following disclaimer information was shown in red, bolded 12- point font and positioned onscreen just below the trial instruction and just above the mortgage calculator tool in every trial:

"When choosing mortgages, bear in mind that rates can rise considerably, fall or remain the same as today. These fluctuations can significantly affect your future repayments."



Imagine you want to buy a flat that costs £205,000. You need to get a mortgage for £185,000 over 25 years.
Please use the loan calculator below to find the best mortgage you can.

To help you make your choice, you can use the toggle keys in the header to switch the order of the mortgages according to each attribute.
You can click on 'current rates' and 'future rates' to compare the repayment costs over time.
By clicking on the '?' you can see the APR and interest rates for each individual mortgage.
When you are ready, click on your chosen mortgage in the SELECT MORTGAGE' column on the right-hand side to submit your choice.

When choosing mortgages, bare in mind that rates can rise considerably, fall or remain the same as today. These fluctuations can significantly effect your future repayments.

Total repayment costs over full term. Show **future** rates

Mortgage type	At CURRENT rates	Price rank order	Monthly cost	SELECT MORTGAGE
? Discounted variable rate A	£360,867	1	£1,202.89 average for full term	Select A
? Discounted variable rate B	£361,737	2	£1,205.79 average for full term	Select B
? Discounted variable rate C	£362,222	3	£1,207.41 average for full term	Select A
? Fixed rate D	£393,389	4	£1,311.30 fixed for lifetime	Select D
? Fixed rate E	£394,797	5	£1,315.99 fixed for lifetime	Select E

Click '?' to see individual loan interest rates and APR's

Figure 5.9 The Behavioural Disclaimer Manipulation in Exp 3b

An example of the disclaimer manipulation, demonstrated in condition 2. The same disclaimer information was applied to all three conditions, and the positioning onscreen remained constant and visible for the full duration of each trial. There was no time restriction applied to responses so participnats were free to analyse the loan data points in as much depth as desired by toggling individual attributues and comparing current and future rate frames in conditions 2 and 3.

Following completion of all eight trials, participants were then directed to the individual difference measures. Firstly, they answered the 10-item Life Orientation Test - Revised (LOT-R) (Scheier, Carver & Bridges, 1994) which is a robust and valid psychometric indictor of optimism, used extensively throughout psychology literature[1] (e.g., Puri & Robinson, 2007). This was followed by a 13-part financial literacy scale (Fernandes, Lynch, & Netemeyer, 2014) and the multiple-choice format of the Berlin numeracy test (Cokely, Galesic, Schulz, Ghazal & Garcia-Retamero, 2012), each of which have robust psychometric features. The final experiment questions relating to use of repayment and rate information were then answered using the drop-down menus as displayed previously in experiment 3a. Participants were then provided with the reimbursement information.

[1] For a list of references to articles in which the LOT-R has been used, see http://www.psy.miami.edu/faculty/ccarver/sclLOT-R.html

5.3.2 Results

Analysis of Choice Effectiveness Per Condition

Choice scores and the proportions of correct choices made per condition with the addition of the disclaimer were almost identical to those found in experiment 3a with the framing manipulations only (see figure 5.11). Table 5.2 displays the differences in the mean choice scores, proportions of correct choices, and choice frames between the experiment 3a (non-disclaimer) trials and those with the addition of the disclaimer (experiment 3b).

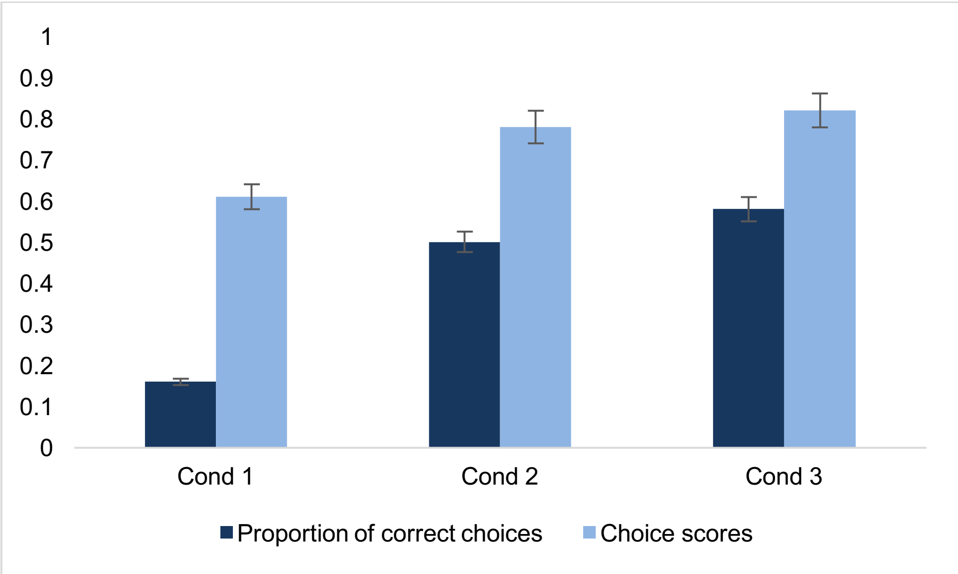


Figure 5.10 Choice Scores and Correct Choices Per Condition in Exp 3b

Mean choice scores and proportions of correct choices per condition with the addition of the disclaimer manipulation (experiment 3b) with standard error bars. As shown in table 5.4, these values did not significantly differ to the results in experiment 3a, indicating that the disclaimer had no additive impact on judgmental performance in any condition. The ineffectiveness of the disclaimer in the control condition thus indicates that the higher performance in condition 2 and 3 is associated with the framing manipulations and not simply encouraging people to more carefully consider their decision based on future rate variability.

Experiment 3a and 3b Comparison of Choice Performance

Table 5.2 Mean Choice Scores and Correct Choices in Exp 3a Vs Exp 3b

Mean differences in choice scores and proportions of correct choices made per condition with the addition of the disclaimer manipulation (experiment 3b) and without the disclaimer (experiment 3a).

	Framing manipulation only (exp 3a)	Framing manipulation with disclaimer added (exp 3b)
Choice score	Cond 1 0.61 Cond 2 0.79 Cond 3 0.83	Cond 1 0.61 Cond 2 0.78 Cond 3 0.82
Proportion of correct choices	Cond 1 17% Cond 2 52% Cond 3 60%	Cond 1 16% Cond 2 50% Cond 3 58%
Proportion of choices in current frame	Cond 2 35% Cond 3 61%	Cond 2 63% Cond 3 38%
Proportion of choices in future frame	Cond 2 65% Cond 3 39%	Cond 2 37% Cond 3 62%
Choice score in current frame	Cond 2 0.89 Cond 3 0.83	Cond 2 0.73 Cond 3 0.88
Choice score in future frame	Cond 2 0.74 Cond 3 0.83	Cond 2 0.87 Cond 3 0.78
Proportion of correct choices in current frame	Cond 2 73% Cond 3 58%	Cond 2 40% Cond 3 65%
Proportion of correct choices in future frame	Cond 2 41% Cond 3 62%	Cond 2 66% Cond 3 53%

The results displayed in table 5.2 show a reversal in the pattern of responses found in experiment 3a for the proportion of choices made per frame. In experiment 3a, a higher proportion of choices were made in the non-default frame in both the optimistic framing condition 2 and the realistic framing condition 3. With the addition of the disclaimer manipulation however, a significantly higher proportion of choices were made in the current (default) frame in condition 2 compared to with the framing manipulation only (M=0.63 vs M=0.35), $t(892.19) = 8.83$, $p<.001$ and in the future (default) frame in condition 3 (M=0.62 vs M=0.39), $t(830.87) = 6.80$, $p<.001$.

The distribution of choice scores and proportions of correct choices also altered by frame with the disclaimer information. While the overall means across frames per condition were almost identical between experiments, choice scores in condition 2 were significantly higher with the disclaimer in the future frame compared to without the disclaimer (M=0.87 vs M=0.74), $t(400.8) = 6.15$, $p<.001$ and lower in the current frame

(M=0.73) compared to without (M=0.89), $t(395.57) = -7.17$, $p<.001$. The same pattern was observed in condition 3 with significantly higher choice scores in the current frame with the disclaimer (M=0.88 vs M=0.83), $t(344.43) = 2.31$, $p<.001$, however, the choice scores in the future frame were not significantly lower than those with the framing manipulation only (M=0.78 vs M=0.83), $t(380.65) = -1.83$, $p=.07$. Previously, no difference was found between choice scores per frame in condition 3, however, the disclaimer yielded a significant increase in choice scores in the current frame compared to the future frame in condition 3 (M=0.88 vs M=78), $t(385.27) = 4.01$, $p<.001$

Congruent with the results of experiment 3a, experiment 3b also yielded an overall higher proportion of correct choices in the realistic framing condition 3 compared to the optimistic framing condition 2 (M=0.58 vs M=0.50), $t(832.09) = 2.31$, $p<.05$. Moreover, the proportions of correct choices per frame reflected the pattern shown with the choice scores, with significantly larger proportions in the non-default compared to the default frame in condition 2 (M=0.66 vs M=0.40), $t(343.51) = 5.62$, $p<.001$ and in condition 3 (M=0.65 vs M=0.53), $t(332.57) = 2.33$, $p<.05$. Again, this was the reverse to what was found in experiment 3a without the disclaimer manipulation. In condition 2 and 3, 80% and 84% of participants reported using both current and future rates to make their choices indicating that choices were made by making one or more comparisons between rate frames in both conditions (as opposed to using current or future rate information). This suggests that choice performance was influenced by the comparisons being made between the different rate frames (i.e., the degree of comparative evaluation of choice alternatives undertaken) and not simply selections based on the default frame with no comparisons made.

Analysis of Individual Differences and Choice Performance

With regard to the measures of individual differences, the original results were shown to hold when these variables were factored in which supports the robustness of the framing effect in facilitating financial judgment across individual differences. Participants on average possessed moderate levels of optimism, financial literacy and numeracy. Table 5.3 displays the mean individual difference scores for each condition. A significant positive correlation between choice scores across current and future frames and financial literacy was found in condition 2, $R^2 = .11$, $F(1,53) = 7.66$, $p<.01$, $B = .043$, $t(53) = 2.77$, $p<.001$. Aside from this result, no other significant relation was identified between the individual difference variables in any of the conditions. The inclusion of optimism, financial literacy, numeracy, education and gender as covariates in regression models to predict choices scores and proportions of correct choices across and between conditions showed no significant main effects or interactions terms. Regressing each of the individual difference variables on choice scores and proportions of correct choices made in current versus future frames in the optimistic framing condition 2 and the realistic framing condition 3 also showed no significant impact of these factors on choice effectiveness.

Table 5.3 Mean Individual Difference Scores Per Condition in Exp 3b
Mean individual difference scores per condition.

Condition	Optimism Max score 24	Financial literacy Max score 13	Numeracy Max score 4
1 (control)	14.05	9.59	1.69
2	16.45	8.85	1.71
3	15.52	8.74	1.62

The variation in the individual difference scores between conditions is not specifically of interest because participants were randomly assigned to each condition in an independent groups design. However, table 5.3 is given for illustrative purposes to show that participants in the control condition possessed the same financial decision making capabilities based on financial literacy and numeracy scores as those in the experimental conditions and were thus able to achieve the same levels of choice performance. The fact that the individual difference scores across conditions are almost identical therefore indicates that the differences in choice effectiveness between conditions (see figure 5.10) were associated with the effects of the framing manipulations.

In accordance with Puri and Robinson (2007), moderate optimism scores indicate more optimal financial judgment and behaviour based on heightened levels of self-control and better long-term financial planning capabilities. These behavioural tendencies suggest a ‘future’ temporal preference where delaying gratification to achieve long-term rewards which maximize utility is favoured over short-term payoffs which minimize costs in the present but act to reduce overall utility. The mean scores in table 5.3 show no difference in levels of optimism between the conditions, which suggests that the framing manipulations were effective in increasing judgmental performance when trait optimism (‘temporal preference’ or self-control in this context) was included as a covariate in a model to predict choice performance. However, the lack of significance of optimism prevents any conclusion regarding the importance of optimism to financial choice behaviour in this particular context. Other measures may be more relevance to assessing temporal preference (i.e., the tendency to focus on long-term rewards/maximize utility versus immediate gains/delayed costs) in online financial decision making situations.

5.3.3 Discussion

As expected, the effects of the framing manipulations in the optimistic framing condition 2 and the realistic framing condition 3 were shown to be robust, with experiment 3b yielding almost identical choice scores and proportions of correct choices to those in experiment 3a. The effectiveness of the framing was consistent when the individual difference measures of financial literacy, numeracy and trait optimism were included as covariates in the model to predict choice performance. Although this result is unexpected with regard to the measure of trait optimism, the effect of the other individual factors is congruent with previous research which shows that financial decision making performance is likely to be independent of the impact of individual differences in numeracy and financial literacy (Fernandes et al., 2014).

It is therefore concluded that the manipulation applied in the realistic framing condition 3 (i.e., current versus future rates framed simultaneously in absolute currency costs with a default set to future rates) is more effective in optimizing financial choice compared to both standard industry formats and a sequential display of rate alternatives with a default set to current rates.

To summarise the details of the format manipulation in experiment 3b, a disclaimer was added to increase sensitivity to rate variability in an attempt to stimulate increased attentional resources in the process of comparing product type alternatives. As expected, this manipulation was shown to have no impact on overall choice effectiveness in either the control or the framing manipulation conditions. The lack of difference in choice effectiveness in the control condition indicated that even when provided with a stimulus to evoke more comprehensive decision making strategies, it was not possible to improve peoples' judgment processes when data was presented in standard industry disclosures (i.e., in percentage and rate formats). These results suggest that the judgment irrationalities in complex numerical environments are likely to stem from the tendency to erroneously process and interpret percentages and non-linear relations between variables.

In this respect, the findings support the robustness of the framing effect in mitigating the biases impacting rationality, which in turn provides support for format biases as one of the primary mechanism underpinning poor judgment in complex non-linear environments. It is possible that in some contexts and among some groups, people may possess the attentional capacities, or be capable of applying the computations necessary to produce rationale inferences based on extrapolating and comparing percentage information. It is also possible that particular informational framings and communications may be effective in eliciting better judgments by providing the stimuli necessary to motivate decision makers to invest the time and cognitive resources required for optimal choice. In this scenario for example, the problems associated with numerical format may be overcome if the context or potential consequences of the judgment are perceived as important enough to warrant the time and effort. Based on the current findings however, it is likely that in the majority of contexts and groups, peoples' judgmental processes will be hindered by the inability to interpret and synthesize multiple cues when disclosed in percentage and rate formats.

As with investigations of other measures of temporal preference in financial judgment, we too found no association between choice effectiveness and trait optimism in this particular context. Despite this, it is possible that optimism is important to financial choice in other situations which involve shorter term judgments and choices. It is possible for example, that budgeting behaviour or consumer purchase decision making is impacted more by optimistic biases where impulsivity or impatience play a stronger role, motivated by other factors such as marketing, digital and social media. Whether or not the disclaimer was useful in decreasing optimistic biases in peoples' judgment processes by drawing attention to the financial risks associated with current low rate choices, the lack of change in choice performance in the control condition in experiment 3b clearly suggests that even when prompted to carefully evaluate choices based on rate differences, people are still unable to effectively perform the necessary numerical computations.

Despite all the information being present for effective choice in the standard industry disclosure, the data format and cue complexity of the decision making domain created a strong barrier to judgment rationality. This finding supports the limited attentional capacity perspective, indicating that peoples' cognitive capabilities are not geared for data synthesis in complex non-linear environments which involve comparing and computing rates and percentages. There were however, changes in the proportions of choices made per frame in experiment 3b which suggest that the disclaimer was effective in increasing the propensity to evaluate the choice alternatives more thoroughly by repeatedly comparing rate frames.

Depending on the task objective, promoting comparative analysis in situations where choice effectiveness and satisfaction is based on subjective measures could be beneficial. For example, choosing between houses, cars or restaurants are possible situations in which an effective choice is based on computing data in accordance with individual goals and requirements rather than objective performance parameters.

With respect to the study limitations, the inability to track the screen alternations and time spent viewing each screen per participant and condition are important shortcomings in both experiments 3a and 3b. Replication as eye-tracking experiments would address this problem and provide richer data relating to the effects of the framing manipulations. As opposed to relying on the measure of numeracy employed, it could also be beneficial to devise a bespoke metric for assessing the propensity to additively process percentages, leading to judgments and choices based on linear projections, which would be more applicable to determining the impact of percentage biases and how they relate to 'optimistic' financial judgment in this particular context. Moreover, recruitment of participants specifically seeking mortgages would also provide richer data relating to the motivations and decision making processes of people evaluating financial product attributes in complex, real-world online settings.

The results of experiment 3a and 3b therefore demonstrate the validity of framing data to influence choice behaviour using clearly defined, easily comparable concrete values rather than expecting people to perform more cognitively demanding comparative analysis of rate data. This suggests that regardless of individual differences, lessening the cognitive load in complex decision environments will yield better results compared to encouraging more cognitively involved evaluative analysis of individual attributes and alternatives. The ability to make quick comparisons of the effects of rate changes over time in concrete values (currency costs) is cognitively efficient and 'intuitive'. From the perspectives of bounded rationality and adaptive heuristics, the conversion of numerical data into absolutes values (particularly in complex, multi-cue environments) is logical, because concrete values with non-normalized base rates are congruous with cognitive methods of sampling, hence the effectiveness of frequency data formats in facilitating statistical reasoning.

In sum, the replication experiment 3b verifies the robustness of the framing effect identified in experiment 3a by showing that with the addition of the disclaimer and inclusion of individual difference factors as covariates, the effects of the original framing manipulation on choice performance remained constant. The results thus indicate the potential for numerical data frames which facilitate human decision making based on minimizing the biases associated with percentage and rate formats. An approached focused on delivering numerical values in concrete, absolute terms presented within a contextual framework of minimal data points involving the use of default values to create anchoring effects is likely to improve judgment rationality to a greater extent than focusing on behavioural manipulations aimed at motivating people to engage in more complex cognitive processing strategies. The results therefore suggest that the affective and behavioural factors motivating people in real-world numerical choice contexts may be less important to judgmental performance than peoples' interpretation of the numerical values within the data environment itself. I.e., the semantics of the situation and how the data assimilates within the wider context may be outweighed by the focus on the data in a minimal capacity, assessing a limited number of points in relation to one another rather than synthesizing the data with other cues and knowledge to construct meaning and deduce the wider implications.

It is likely that in most cases, the tendency to use numerical information in this way to form probabilistic judgments stems from limited attentional capacities and biases in the computation of percentage points and extrapolation of and non-linear trends, as opposed to a lack of motivation to apply greater cognitive effort and resources. In this sense, peoples' judgmental strategies reflect their processing and computational constraints which cannot be overcome, even when the judgement context is made more salient to motivate greater effort in achieving optimal choice. Although the judgment context may be particularly important or meaningful, and the perceived choice consequences are great, the findings suggest that applying behavioural manipulations to motivate people to make better choices (e.g., by evoking loss aversion in a financial setting) cannot offset the biases generated by the underlying cognitive mechanisms which shape peoples' interpretation of numerical information, and consequently the rationality of human judgment. These results are contrary to previous studies which have shown behaviourally informed disclaimers designed to increase sensitivity to mutual funds investment fees by utilizing social comparison are effective in improving choice when information is disclosed to investors in percentage formats (Newall, 2016b).

Although the robustness of the framing manipulation across experiments 3a and 3b establishes the importance of delivering data in concrete (absolute) terms using minimal cues, is it not clear whether effectiveness was related specifically to the framing of rates in absolute terms or to the use of the future rate default in the realistic framing condition 3 which acted to anchor people on the most optimal choice prior to assessing the rate frame alternatives. Further examination is therefore necessary to distinguish the effects of simultaneously displaying rate alternatives in absolute values to encourage comparative analytic processes from the effects of using a default setting to influence judgment by anchoring people on a specific value within a contextual framework constructed within the data environment.

To address this question, chapter 6 extends the investigation of the framing effect in experiments 3a and 3b by testing the components of the effect in two individual data disclosures. Experiment 4 explores the impact of a default manipulation separately from the effect of a default with the addition of future rate context data on monthly loan repayment judgments. In conditions 1 and 2, a minimum suggested repayment amount for the full loan term in current rates is disclosed with a default repayment figure in the form of a higher monthly repayment option for the loan at a 50% term reduction. In the second condition, the same default is presented, except that repayment information for future rates are presented simultaneously with the current rate information. The addition of the future rates context data thus acts to increase the value of the default and widen the range between the minimum (full term) and the default repayment figures. The aim is to identify whether the saliency of the added context data combined with the increased range between the repayment figures in condition 2 is more effective in heightening peoples' monthly repayment choices above the minimum compared to when viewing the default in the absence of future rate context data.

Chapter 6

Framing Interest Rates and Loan Term Information to Influence Repayment Behaviour

In this chapter, the effectiveness of a default setting is explored in relation to the framing effect which was established in chapter 5 as robust in improving financial judgment based on mitigating percentage data format biases. Experiment 4 tests the additive effects of the framing manipulation in conjunction with an anchoring effect created by providing a higher repayment option in relation to a minimum suggested amount. The default is generated by disclosing the costs to repay a loan over a 20-year vs. a 10-year term. By creating an upper bound in the context of the minimum amount over the 20-year term, the default was shown to increase the inclination to make a monthly loan repayment significantly greater than the minimum suggested amount. This effect was amplified when costs were disclosed simultaneously in current vs. future rate frames based on a larger absolute difference between the upper and lower bound generated by the data for future rates.

Akin to the findings from the retail setting, people were shown to combine the data points additively, forming a judgment based on the arithmetic mean of the highest and lowest observed values per condition. Repayment judgments were slightly greater than the mean of the upper and lower bound in the combination condition, indicating an anchoring effect based on the size of the absolute difference between the minimum and maximum options. Findings indicate that judgments are formed by additively combining higher and lower values in multiple cue environments, and that the larger the upper bound in relation to the lower bound, the greater the tendency to anchor on the higher figure, thus creating a lesser adjustment towards to mean.

6.1 Background and Rationale

In terms of the findings across all chapters thus far, experiment 1, chapter 3 has shown that when predicting product sales based on individual numeric values, retail forecasters employed only two data points to form probabilistic judgments. The judgment strategy involved treating all numbers as absolute values and applying arithmetic operations to yield the absolute difference between the last two data points per trial. They then added or subtracted the absolute difference from the last observation to predict sales for the fourth week. This method acted to yield judgments which linearly project the trends, even though some data were percentages, displayed with ‘%’ signs, and trended exponentially rather than linearly.

The same strong tendency to form predictions by linearly extending trends was shown again among humanitarian aid workers in experiment 2, chapter 4. When presented with data in time series formats, the tendency to seek and apply linear functions increased in noise and correlated with performance decline. Noisier data (i.e., increases in the number of cues) was shown to impact judgments by promoting the inclination to extend the direction of the trends when congruity was observed. This ‘trend effect’ occurred regardless of the non-causality of the additional cues or the fact that the real-world data trended cyclically and not linearly. Thus, whether data was framed in numeric or time series formats, both retailers and humanitarians applied methods which consistently yielded linear judgments. Moreover, the linear tendency persisted with the addition of context data in the time series format, with people showing a strong bias to visually incorporate the cues in a process of guiding judgment formation based on the same additive/linear methods.

Examining the tendency to interpret numeric data in concrete terms and apply additive methods in the context of financial choice in experiment 3a and 3b in chapter 5 confirmed the robustness of the format bias, suggesting these tendencies characterize decision processes across contexts, levels of expertise and environmental complexity. Whether making a formal probabilistic judgment (i.e., in a retail or humanitarian aid forecasting context), or implicitly forming probabilistic inferences when choosing between financial product alternatives, people are shown consistently to utilize minimal data and arithmetically process numbers resulting in linear inferences. Reframing interest rates in current versus future alternatives in absolute costs, combined with a future rate default demonstrated the robustness of the framing effect in facilitating performance by removing the barriers created by these biases in percentage information domains.

The framing effect was shown to hold when individual difference covariates were included in models to predict judgment performance, and when the framing manipulation was tested in relation to a behavioural disclaimer to increase activation of attentional resources in standard percentage formats. It is a possibility that the saliency of the context was not sufficiently heightened by the disclaimer to motivate people to increase cognitive effort. However, based on the widespread evidence of percentage format biases in financial judgment contexts, it is concluded that the consistently poor performance in the control condition in experiment 3b was related to limitations in the processing of data in percentage and rate formats. This therefore indicates that rather than a lack of cognitive effort, format biases are the primary mechanism underpinning erroneous judgment in complex percentage format domains. For example, even in real-world contexts where there is motivation for high performance, the findings across all chapters indicate that people do not possess the cognitive capacity to synthesize cues and execute the appropriate operations to compute and extrapolate percentages and rates in complex settings.

The results of experiment 3b indicate the framing effect to be highly robust which in turn supports the rationale for the linear prediction heuristic as underpinning percentage format biases. However, the mechanism via which the framing effect facilitates performance may be related more to the use of the default setting in the realistic framing condition 3 which anchored people on the most optimal choice than to the side-by-side disclosure of current versus future rate cost alternatives There is substantial evidence both for the use of defaults in creating anchoring effects which lead to improvements in judgment and choice as well as for the presentation of choice information in simultaneous displays. It would therefore be useful to follow up the results from experiments 3a and 3b with an experiment to test whether the default may be the key factor contributing to the judgment effect, or whether the additive effects of combining the default and future rate frame manipulations creates improvements over the default alone.

To further the investigation of the effects of defaults and future event context information on financial judgment, experiment 4 examines the impact of the two mechanisms on monthly mortgage repayment judgments by separating the informational content between two experimental conditions. Experiments 3a and 3b involved the testing of a framing effect in which the information presented in each condition remained the same, with the presentational format altering between them. However, in experiment 4 the mechanisms involved in the framing effect are further explored in separate data manipulations which involve two conditions disclosing different information. By examining the components of the framing effect in isolation, it may be possible to identify how the mechanisms involved in financial data frames and formats influence peoples' interpretations and judgments.

Akin to selecting a mortgage, making a monthly loan repayment decision is also an example of an important real-world financial judgment in which an individuals' interpretation of the information can be significantly influenced by the way in which the data is framed and disclosed. There are various advantages to testing the effects of defaults and future rate context information in a mortgage loan monthly repayment situation. Based on the findings, it may be possible to make recommendations for increasing the amount people opt to repay on a long-term loan per month. This can be financially advantageous by both reducing the loan term and thus the total repayment cost, and also by helping people to restructure peoples' budgeting behaviours in preparedness for rate increases. In the context of experiment 4, a higher repayment choice thus reflects a lower discounting rate which in turn indicates a less optimistic and more effective financial judgment.

It is necessary to note however, that there are complexities involved in the assessment of the rationality of financial judgment which extend beyond the scope of experiment 4. For the purpose of the experiment, judgment 'rationality' is measured from the economic perspective, in terms of the inclination to make higher monthly repayments in the present which act to reduce total repayment costs when factoring in interest rates over the course of the loan terms. As discussed in chapter 5, however, there are cases in which delay discounting, or the decision to make a smaller monthly repayment which leads to a larger cost overall, may not be irrational when considering a persons' individual circumstances. For example, showing a tendency for delay discounting by opting for low monthly costs in the present despite higher overall costs in the future could be a rational strategy in the context of an expected inheritance, or future change in employment circumstances. In fact, delay discounting in such circumstances could actually indicate increased judgment rationality, based on more effective long-term financial planning and management skills.

Previous studies which have tested financial data manipulations similar to real-world credit card statements have yielded positive effects on consumer financial judgments. For example, based on the format of the Credit Card Accountability and Responsibility Disclosure (CARD) Act of 2009, Salisbury (2014) for example, showed that proving costs and loan repayment duration information for an additional higher amount in conjunction with the minimum costs increased monthly credit card repayments from the minimum required. The CARD Act of 2009 introduced the "minimum payment warning" which involves the following text prompt: "Minimum Payment Warning: If you make only the minimum payment each period, you will pay more in interest and it will take you longer to pay off your balance". This is combined with information about how much borrowers need to pay per month to clear the balance in three years in relation to the costs associated with making the minimum repayment only. To test the effects of this disclosure format, Salisbury (2014) presented consumers with the credit card balance, APR, time it would take to repay the loan based on the minimum repayment, and also the interest costs and time it would take to clear the balance over a three-year period.

Results showed that minimum repayment cost and time information had no effect on propensity to repay any amount. However, providing additional three-year information for either time or cost, or both time and cost was effective in increasing the propensity to repay between \$60-\$70 per month rather than the minimum amount of \$38.74. The probability of repaying less than \$60-\$70 was found to decrease with the presence of the three-year cost information. However, the probability of actually paying more than the three-year repayment amount (\$66.21) decreased with presence of the three-year time information, particularly among those with a low knowledge of compound interest. Moderation analyses further indicated that the three-year time information increased the probability of repaying less than the three-year amount among consumers with six or more credit cards (16% of the sample).

These results have important implications for regulatory financial policies and lender practices, suggesting that nudges in the form of alterative repayment choices are necessary for improving repayment behaviours. Rather than simply indicating that slower repayment incurs greater costs, it is necessary to understand the psychological factors underpinning the propensity to repay among different consumer groups. As demonstrated, cost and time information effected consumers differentially. Cost information was shown to increase repayment amount in some cases, whereas time information decreased repayment in others and had a stronger effect on consumers with multiple credit cards or less knowledge of interest compounding. It is likely that the three-year time information negatively impacted the propensity to pay more than the three-year interest amount because presentation of the two values created the perception of an upper and lower bound (\$38.74 and \$66.21). This may have led consumers to adjust downwards from the \$66.21 towards the midway point (e.g., Slovic, Finucane, Peters & MacGregor, 2002; Roller, 2011).

Experiment 4 involves a similar informational format to that used by Salisbury (2014) with a default manipulation in the form of cost information for the mortgage at a reduced term of 10-years. In both conditions 1 and 2 this default information is applied in the presence of the full term, 20-year repayment information. In condition 1, both these suggested monthly repayment figures are presented in current interest rates only. The presentation of the minimum monthly repayment cost for the half versus the full term thus acts to create a numeric range in the data environment within which a judgment can fall. Based on the saliency of the default information (i.e., it being associated with lower interest and overall loan costs) and peoples' propensity to derive meaning through comparison against a referent point within an upper and lower numerical bound (Roller, 2011), it is expected that the default will create an anchoring effect. This will lead people to focus more on the 10-year repayment figure than the 20-year figure and adjust downwards from 10-year value towards the minimum suggested amount.

In condition 2, the same data points are disclosed as above, except that the data presentation also includes the information for the full and reduced term costs in future interest rates. This future rate context data thus increases the absolute difference between the default (i.e., the 10-year term data which is now observable in future rates) and the minimum suggested amount (i.e., the 20-year data which is still observable in current rates). It is expected that by increasing the value of the default repayment figure in the context of the future rate information, the saliency of the default will heighten and increase its anchoring effect by increasing peoples' awareness of risk and activating greater loss aversion.

Similar judgment mechanisms to those shown in experiment 1 with the retail forecasters are likely to be observed among people making loan repayment decisions. It is expected that people will make judgments based on two numeric referents (i.e., the minimum and maximum suggested repayment values) and additively combine the values in each condition to yield a decision which reflects the arithmetic mean of the two figures. The anchoring effect of the default is expected to impact this process, leading people to make further adjustments to the additive computation based on the size of the absolute numeric difference between the minimum and maximum values. In this view, the larger the data range, the more the default will influence peoples' repayment judgments based on the increased importance of the default which is inferred when positioned in the future rate context.

The proposed mechanism therefore involves levering additive processing propensities applied in the presence of a numeric continuum, combined with creating an anchoring effect using a numeric default which is made more effective when disclosed in highly salient context information. In condition 1, it is therefore predicted that people will show a propensity to make repayment decisions which are significantly above the minimum suggested 20-year term value, and that judgments will fall around the mean of the 10-year default figure and the 20-year amount. In condition 2 however, it is expected that positioning the default in the added future rate context information will increase the anchoring effect of the default and lead people to make lesser adjustments away from it, thus generating more economically optimal judgments. The effect is based on the increase sensitivity to the default which is created by the saliency of the future rate context, and the magnitude of the absolute difference perceived between the default and minimum suggested repayment figures. Thus, compared to condition 1, repayment decisions in condition 2 are expected to be significantly higher and to fall within in a range which is higher above the mean of the 20-year and 10-year figures.

In sum, the informational manipulations tested in experiment 4 further explore the mechanisms involved in the data framing which was shown to facilitate judgment performance in experiment 3a and 3b. Specifically, experiment 4 assess the impact of a default value in the form of a reduced term repayment suggestion versus a full term repayment suggestion on peoples' loan repayment judgments in condition 1. In condition 2, future rate context is added to the data disclosure which acts to increase the value of the default figure and increase the absolute difference between the minimum and maximum values in the data environment. The aim is to examine whether the addition of highly salient future rate information is more effective in promoting optimal loan repayment judgments and behaviours compared to applying a default without the added saliency generated by providing future event information. As with experiments 3a and 3b, judgment 'rationality' in experiment 4 is determined from an economic perspective, whereby judgments which act to minimize total repayment costs are regarded as optimal.

Greater choice effectiveness in condition 2 would thus indicate the importance of future event information to peoples' judgment processes where compound interest is involved. From this perspective, the use of salient future event information to amplify the anchoring effects of defaults and highlight intertemporal cost implications, may be an effective means of counteracting biases and optimistic tendencies involved in ineffective financial judgments and other judgment situations which involve non-linear relations between variables such as health behaviours.

6.2 Experiment 4

Experiment 4 further examines the mechanisms underpinning the effectiveness of the framing manipulation in experiment 3a and 3b by testing the anchoring effects of a default value compared to the effects of the default when disclosed in the context of future rate information on mortgage repayment judgments. In condition 1, the default (in the form of the monthly repayment cost for the loan over a reduced term) is presented simultaneously with the minimum monthly repayment figure for the loan over the full term in current rates. In condition 2, the same data is presented but with the addition of the repayment figures in the context of future interest rates. Using a design similar to that employed by Salisbury (2014), participants in both conditions are presented with a 'mortgage repayment statement' in which the monthly mortgage repayment costs are broken down into interest and total value amounts akin to a monthly credit card statement.

In the *term default only framing condition 1*, the total balance, interest rate and term information is given, together with the current balance, monthly repayment amount and interest costs and for clearing the balance over the full term of the loan (20 years) in current rates. The same information is also shown for the loan at a reduced term (10 years) which creates the default repayment figure. In the *term default plus future rate framing condition 2*, the information and repayment options disclosed in condition 1 are repeated but with the addition of the full and reduced term data (i.e., the default figure) framed in future interest rates. The data manipulation in condition 2 thus involves the repayment information framed in both current versus future rates, plus the default figure which acts to further increase the size of the default in relation to the full term monthly repayment figure compared to condition 1. By disclosing different repayment information in separate data manipulations it is possible to compare the effects of the default to the effects of the added future rate information between the conditions.

The heightened default value combined with simultaneous current and future rate disclosure in condition 2 is expected to be particularly effective in increasing repayments above the minimum based on past findings which have shown that warnings about higher costs and repayment horizons associated with minimum repayments were effective in increasing willingness to pay more via the increased saliency of the consequences that decisions can have on future outcomes (Haws, Bearden & Nenkov, 2012). Simultaneously framing three-year and minimum repayment data is also shown to be effective in increasing the salience of the trade-off between current and future costs and time horizons, facilitating borrowers in evaluating the differences between making minimum versus larger repayments (Soil, Keeney & Larrick, 2013). Moreover, Soil, Keeney and Larrick also found that the dual presentation of minimum and three-year repayment data significantly reduced the tendency to overestimate and underestimate the three-year repayment costs which was associated with high and low levels of numeracy, respectively.

These findings suggest that the effects of defaults may depend on the absolute difference between the default value and the data available for comparison (e.g., a minimum suggested repayment amount). Thus, the theoretical question relates to the perceived size of the default in relation to other relevant data points in the judgment context. As explained above, it is likely that the larger the difference between the default and the minimum suggested amount (and thus the perceived range between the upper and lower bound), the more the default is likely to influence the judgement process. I.e., the greater the absolute difference, the stronger the anchoring effect is likely to be, leading people to adjust less from the upper bound towards the mean.

These effects may also be context dependent. For example, in this particular scenario, the use of the default acts to increase the saliency of loan costs in future rates, making people aware of the optimality of a higher repayment choice compared to the minimum suggested amount, which over the long term, represents a far less effective decision. Optimistic tendencies are therefore likely to come into play when making decisions about how much to repay per month. However, presenting a default value of a reduced term in a future rate frame is likely to heighten the saliency of compound interest rates and communicate the costs associated with rate variability, thus counteracting optimism by anchoring people on the most effective option. From this perspective, downwards adjustments towards the mean are likely to be influenced by the size of the default in relation to minimum presented values, with larger absolute differences dictating lesser adjustments, leading to more effective (less optimistic) repayment decisions.

In sum, experiment 4 involves the further investigation of the framing manipulation applied in experiments 3a and 3b in the arena of loan repayment decision making. The aim is to assess whether the propensity to make monthly mortgage repayments above a minimum suggested figure can be impacted by using a default to elicit anchoring effects, and whether adding future event context increases the effectiveness of the default. Increasing monthly loan repayments has the benefit of minimizing total mortgage costs by reducing the loan term and interest payments. There is also the added advantage of making higher monthly repayments in facilitating financial preparedness ahead of interest rate rises.

It is expected that the default in the form of the cost to pay off the balance over half the loan term (in condition 1) will be moderately effective in increasing the tendency to repay an amount above the minimum suggested amount (i.e., the figure necessary to clear the loan over the full term). The dual effects of presenting the default figure in the context of future interest rates (in condition 2) is expected to be even more effective in yielding repayments closer the optimal amount. The increased effectiveness of the combined manipulation is based on creating a larger absolute difference between the default and the minimum suggested amount which works advantageously with the propensity to additively process the two values (thus yielding a higher mean value compared to condition 1). Increasing the range between the minimum and maximum suggested amounts raises data saliency in the context of loan costs and interest rates which is likely anchor people more firmly on the default and increase its anchoring effect. People in condition 2 are therefore expected to make lesser adjustments away from the default figure and towards the minimum suggested amount. This manipulation has the potential to be particularly effective in counteracting the harmful effects of optimistic or present biases, which involve the tendency to minimise present costs and discount the impact of rate rises in the future.

6.2.1 Method

Participants

Another 160 participants aged over 18 years were recruited via Amazon Mechanical Turk and paid $2.00 (again, no participants from experiment 3b repeated the trials in experiment 4). The average age was 33.82 years (SD = 10.75, range = 20 to 69 years), 40% were female and 36% were educated to a minimum level of a Bachelor's degree.

Design

A single factor independent groups design was used in which participants were randomly assigned to the repayment *term default only* condition 1 which involved a default setting of costs for the full versus half the loan term, and the repayment *term* *default plus rate framing* condition 2 which involved the costs framed in current versus future rates for the loan term in full versus half.

Materials and Procedure

After providing demographic information for age, gender and education, participants were then directed to the main choice task which involved answering one question relating to their repayment decision. On a blank screen, they were first provided with the following information:

"Imagine you are taking out a mortgage for a value of $275,000 over a term of 20 years for an apartment costing $325,000. On the next screen you will see information relating to your repayment options for clearing the balance. You will be asked to choose how much you would like to repay per month."

After clicking to go to the main task, they were provided with the following instruction:

"Please use the information below to carefully consider how much you would prefer to repay per month on your mortgage of $275,000 and enter a figure into the box at the bottom. The information is there as a guide you can choose any amount to repay. Please treat your payment decision as you would in your everyday life."

This text was presented at the top of the screen with the mortgage information displayed in a tabulated format in the lower section of the screen, as shown in figure 6.1. The above instruction remained onscreen for the full duration of the trial and there was no restriction placed on response time. After examining the information provided in the 'Mortgage repayment statement' in each condition, participants then entered a monetary amount into the response box at the bottom of the statement. Values of less than $1.00 were disallowed, and the following prompt was used to ask participants to confirm their decision before being able to submit their answer to encourage more careful consideration of responses:

"You have chosen to repay $2000. To confirm your answer and move to the next section please click OK."

Mortgage repayment statement

Mortgage type: Mortgage value: Interest rate: Mortgage term:	Standard variable rate $275,000 2.22% for 2 years then 4.99% 20 years
If you repay this amount per month, it will take 20 years to clear the balance:	$1,841.83
Total interest paid over 20 years:	$167,038.41
Total mortgage cost over 20 years:	$442,038.41
If you repay this amount per month, it will take 10 years to clear the balance:	$2,811.12
Total interest paid over 10 years:	$62,334.61
Total mortgage cost over 10 years:	$337,334.61

How much will you pay per month?

$ 0 **next**

Figure 6.1 (A) The term default only framing condition 1

Mortgage repayment statement

Mortgage type: Mortgage value: Interest rate: Mortgage term:	Standard variable rate $275,000 2.22% for 2 years then 4.99% 20 years	
	At current rates	**At future rate increases**
If you repay this amount per month, it will take 20 years to clear the balance:	$1,841.83	£2,122.48
Total mortgage cost over 20 years:	$442,038.41	£509,396.10
Total interest paid over 20 years:	$167,038.41	£234,396.10
If you repay this amount per month, it will take 10 years to clear the balance:	$2,811.12	£2,964.46
Total mortgage cost over 10 years:	$337,334.61	£355,735.65
Total interest paid over 10 years:	$62,334.61	£80,735.65

How much will you pay per month?

$ 0 **next**

Figure 6.1 (B) The term default plus and rate framing condition 2

Figure 6.1 Screenshot of the stimuli presented in condition 1 and 2

Based on the disclosure format for credit card statements introduced by the CARD Act of 2009, both conditions provided the basic mortgage information relating to the cost of the loan, the current interest rates and the loan term (presented in the upper section of the mortgage repayment statement). In both conditions, this information was then presented in the context of the total cost of the mortgage over the full term, further broken down into the interest cost and the monthly repayment amount necessary to clear the balance over the full term (shown in the mid-section of the statement in both conditions). This repayment data was then placed in the context of the total, monthly and interest cost alternatives for repaying the full balance over half the loan term, i.e., in 10 years as opposed to in 20 which was displayed in the bottom section of the statement in both conditions.

In condition 2, another layer of contextual information was provided for evaluative purposes by framing both the 20 and 10-year cost alternatives in current versus future interest rates. The future rate costs were computed using the same method applied in experiment 3a, in which +0.5% was added to the rate each year for three consecutive years following the 2-year fixed period. Over the 20-year term for example, this generated a rate of 5.99% in year three and 6.49% in year four which continued through to year 20. This additional future rate information was predicted to be more effective in increasing the propensity to repay more than the minimum suggested amount to clear the balance over the full term ($1,841.83) compared to the disclosure of costs over the 20 versus 10-year term only in condition 1. The aim therefore, is to examine the effectiveness of these framing alternatives in counteracting over-optimistic tendencies associated with choosing to delay costs/repay smaller amounts in the short term.

After submitting a monthly mortgage repayment decision, participants were directed to a new screen where they answered questions relating to potentially relevant individual differences in repayment behaviour, again using the LOT-R measure of dispositional optimism (Scheier, Carver & Bridges, 1994), the financial literacy scale (Fernandes, Lynch, & Netemeyer, 2014) and the Berlin numeracy test (Cokely, Galesic, Schulz, Ghazal & Garcia-Retamero, 2012). Prior to completing the inventories, an additional question was also included as an alternative means of assessing temporal preference for gains. Using a dropdown menu, participants were asked to respond to the following question:

"Would you rather win US$10,000 now or win US$13,000 a year from now?"

If a participant chose to win US$10,000 today, they would be categorized as being impatient (or possessing a high temporal preference for gains/low temporal preference for losses), based on an implicit discount rate of more than 30% (e.g., Loewenstein, Read & Baumeister, 2003). This measure therefore indicates a preference for immediate gains (indicating a preference for delayed higher repayment costs) versus delayed gains (indicating a preference for higher immediate costs), and was thus included as an additional measure to the LOT-R to examine decision performance based on underlying behavioural propensities related to self-control (Puri & Robinson, 2007). Following completion of the inventories, participants were provided with the reimbursement information.

6.2.2 Results

Of the 160 participants who completed the experiment, four were removed from the data set due to incomplete responses. This yielded 79 participants in condition 1 and 77 in condition 2.

Figure 6.2 shows participants mean repayment judgments per condition in relation to the minimum suggested repayment amount in both conditions (i.e., the necessary amount to clear the loan balance over the full term based on current interest rates with no rate variability over the full term), the maximum suggested amount (i.e., the default), and the average of these two values. As expected, one-sample t-tests conducted on the mean repayment amount per condition in relation to the minimum suggested amount ($1,841.83) showed that repayment decisions were significantly greater than the minimum in condition 1, $t(78.96) = 45.77$, $p<.001$ and in condition 2, $t(76.96) = 61.28$, $p<.001$. A one-way ANOVA conducted on the repayment amount per condition confirmed hypotheses, showing that participants decided to repay a significantly larger amount per month on average in condition 2 (M=$2,492.18) with the loan term default manipulation in the context of the future rate data, compared to in condition 1 (M=$2,297.67) with the loan term default manipulation only, $F(526.29) = 178.9$, $p<.001$.

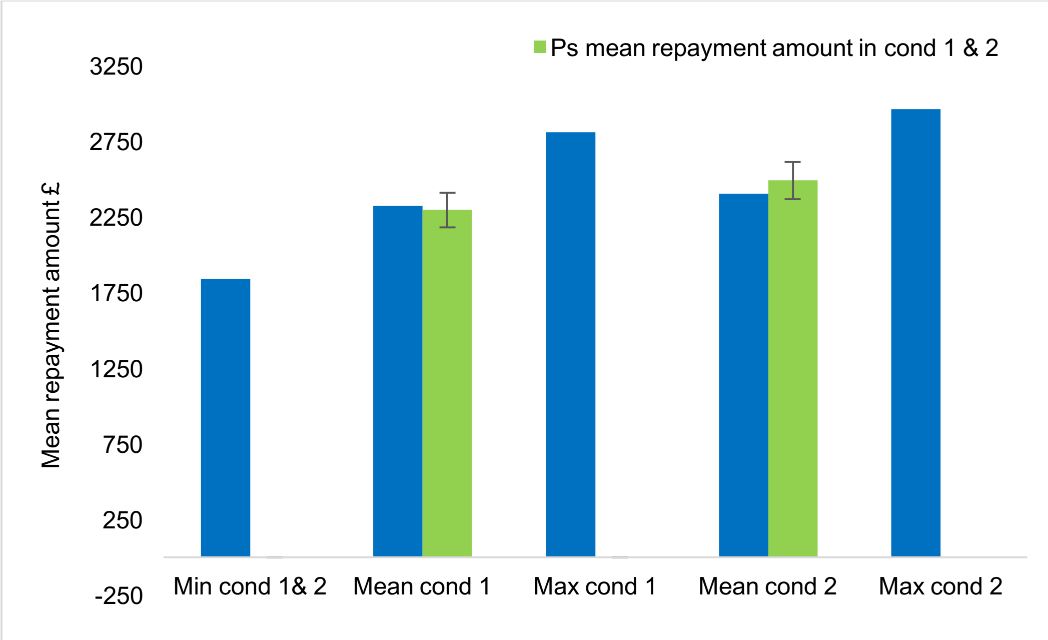


Figure 6.2 Mean Monthly Repayment Decisions Per Condition in Exp 4

The mean monthly repayment judgments with standard error bars in the reduced term default only framing condition 1 and the reduced term default plus future rate data in condition 2. Repayment judgments in both conditions are represented by the green bars and shown in relation to the minimum suggested repayment amount ($1,841.83), the maximum suggested repayment amount (default) in condition 1 and 2, and the average of these upper and lower bound values. The participants' judgment is positioned next to the mean of the minimum and maximum values in condition 1 and 2 to illustrate the proximity to the arithmetic average each judgment fell.

This finding confirmed expectations regarding the effects of the default, showing that disclosing repayment costs in 10-year terms in the context of the full term repayment data (condition 1) resulted in repayment amounts which were greater than the minimum suggested amount. In the absence of a baseline condition, the possibility cannot be excluded that people may have made repayments which exceeded the minimum suggested amount without the presence of the default in condition 1. Based on past findings however (Salisbury, 2014), it is very probable the observed effect was related to the disclosure of the default repayment figure. In future extensions of the research, comparisons to baseline measures are necessary to fully isolate the effects of default settings.

Disclosing the same information in condition 2 with the addition of costs frames in future rates (thus enabling comparison with the minimum suggested amount in current rates) significantly increased the mean repayment judgment in condition 2 further above the minimum suggested amount compared to condition 1.

On average, participants' repayment judgments in condition 1 ($2,297.67) represented an 124.7% increase from the minimum suggested amount, which equated to 80.2% of the maximum suggested amount (i.e., the default of $2,811.12 based on the cost to clear the balance in current rates over the reduced 10-year term). This resulted in a mean repayment choice which fell marginally below the average ($2,326.47) of the minimum and maximum suggested repayment values presented in condition 1. In condition 2, the mean repayment judgment ($2,492.18) rose to 135.3% of the minimum suggested amount, equating to 84% of the maximum suggested amount which resulted in repayment judgments falling marginally above the mean of the minimum and maximum values ($2,403.145) presented in condition 2. This reflected in an absolute difference of $28.80 in condition 1, and $89.03 in condition 2 between the mean of the upper and lower bound and participants' repayment judgments.

This indicates an anchoring effect in combination with additive processing of the upper and lower bound values. The fact that judgments in both conditions were so close to the average of the minimum and maximum suggested amounts indicates that people combined the values using the arithmetic mean as the computational strategy for yielding a judgment in this context. The fact that the judgment was in excess of the mean in condition 2 corresponds with the increase in absolute difference between the minimum and maximum suggested amounts (i.e., the extent of the range between the upper and lower bounds). In this respect, the greater the absolute difference between the maximum and minimum values presented, the greater the anchoring influence of the upper bound on peoples' decisions.

It is necessary to further test the possible effect of the size differentials between upper and lower bounds in other financial decision making contexts as well as domains such as health behaviour change consumer purchasing behaviours. It is necessary to examine the extent to which judgments deviate from the mean, and the direction in which they increase and decrease based on smaller and larger absolute differences between minimum and maximum values. It is also necessary to assess the impact of increases and decreases in the numbers of informational cues provided in decision environments such as these to determine parameters for the tendency to arithmetically combine only two data points under condition of varying complexity. It is possible that the greater the degree of noise (i.e., the number of non-causal or low causally-related cues) in the environment, the more people are inclined to use minimal cues in the attempt to simplify the judgment task. This may hold for example, in online consumer contexts where multiple conflicting promotional data create high variance which can impact cue selection.

In sum, it is evident that the anchoring effect of the default setting used in this setting had a stronger impact on repayment decision making when applied in the context of the current versus future rate repayment data. In relation to chapter 5, this suggests that anchors in the form of higher repayment options created by future rate information can increase peoples' propensity to make financial choices which act to offset the optimistic tendency to delay costs and select products or make repayments based on the lowest possible amounts in the present. Thus defaults applied in presence of relevant, complex context information (future rate data) may be useful in facilitating comparisons which guide judgments towards more optimal, less optimistic financial choice and behaviour.

Table 6.1 displays a breakdown of data relating to the repayment decisions made in each condition. Descriptive statistics are provided including the percentages of repayment amounts which fell below, above and in between the minimum and maximum suggested amounts. The minimum ($1,841.83) is the least effective repayment decision, based on the amount necessary to clear the loan balance over the full term in the most optimistic (unrealistic) circumstance (i.e., current interest rates with no variability over the full term). Whereas the maximum ($2,964.46) is the most effective repayment decision, based on halving the loan repayment period to minimise the impact of interest rate variability over the full term. A moderate incline in rates is the most realistic scenario, thus this repayment choice is the least optimistic and the most likely to maximise utility over the full time horizon.

Table 6.1 Descriptive Statistics for Repayment Choices in Exp 4
Descriptive statistics for the repayment choices in each condition and percentages in relation to the minimum and maximum suggested amounts.

	Condition 1 term default only framing	Condition 2 term default plus rate framing
Mean repayment amount	\$2,297.67	\$2,492.18
Range	\$1,000 - \$3,400	\$1,150 - \$3,500
Mode	\$2,811	\$3,000
Median	\$2,200	\$2,700
Low range: Percentage of repayments below the min (\$1,841.83)	10.13%	5.19%
Mid-range: Percentage in between the min and max (\$2,964.46)	86.08%	75.32%
High range: Percentage above the max	3.80%	19.48%

As shown in table 6.1 above, the overall mean repayment amount in condition 2 was higher compared to condition 1 which, as expected, indicates that the addition of the rate frame manipulation in condition 2 was more effective in increasing the propensity to make a higher monthly mortgage repayment compared to the term only manipulation in condition 1. Three individual logistic regressions were conducted in which the proportions of repayments in the low, mid and high range were entered as the dependent variable and condition was entered as the predictor (displayed in figure 6.3 below). The results indicated that the framing manipulation in condition 2 was significantly predictive of both the propensity to make less repayment choices below the minimum (B = -0.72, z = -5.903, p<.001, 95% CI [-0.96, -0.48]) and more repayment choices above the maximum suggested amount (B = 1.81, z = 14.38, p<.001, 95% CI [1.57, 2.06]) and condition 1 was significantly predictive of the propensity to make a repayment which fell in between the minimum and maximum suggested amounts (B = -0.70, z = -8.752, p<.001, 95% CI [-0.86, -0.55]).

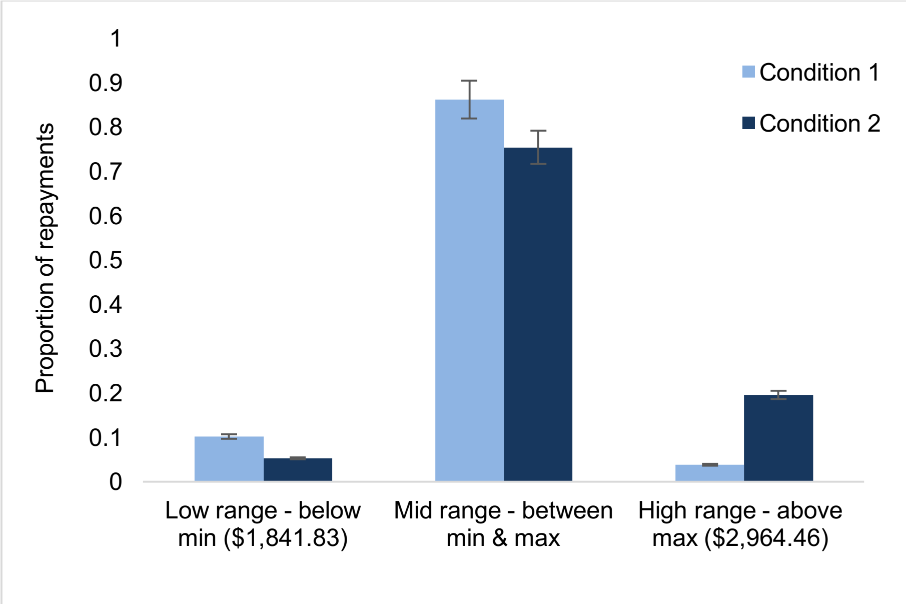


Figure 6.3 Proportions of Decisions Per Repayment Range in Exp 4

The proportions of repayment decisions which fell into each range with standard error bars. This information is the same as the percentages breakdown presented in table 6.1, but displayed graphically for illustrative purposes. As is shown, the largest proportion of repayment choices were in the mid-range (i.e., above the minimum but below the maximum suggested amounts) in both conditions. Although condition 1 yielded a significantly higher proportion of repayment decision in the mid-range where the majority of repayment amounts fell in both conditions, condition 2 was still more effective overall, generating a larger proportion of decisions across both the mid and high range categories combined.

The distribution of repayment decisions across the ranges indicates that condition 1 was more effective than condition 2 in nudging a larger proportion of people towards a repayment amount higher than the minimum (i.e., creating a moderate improvement in loan repayment behaviour). However, condition 2 was more effective in moving peoples' repayment decisions into the high range, i.e., the most effective bracket above the maximum suggested amount. When assessing the overall proportions of repayment decisions which exceeded the minimum (including both mid and high range repayment decisions), condition 2 yielded 95% whereas condition 1 generated a 90% rate of heightened decision performance. Thus, it can be concluded that condition 2 was the more effective manipulation in terms of overall improved repayment choice behaviour.

One possibility is that the variation in repayment decisions between conditions could be due to individual differences interacting with the framing manipulations in each condition. Table 6.2 shows the demographic information and the mean scores for each of the individual difference inventories per condition. As shown, participants in both conditions exhibited similar, moderate levels of financial literacy, numeracy and trait optimism. A larger proportion of participants in condition 2 possessed a temporal preference for gains, based on choosing to receive “$10,000 now” as opposed to “$13,000 in one year.” This preference for smaller immediate gains indicates a higher level of financial optimism which could be associated with choosing to repay less per month, based on the bias to believe that financial adversity will not occur, or that one will have the capability to cope with future financial adversity.

Table 6.2 Individual Difference Scores and Demographics in Exp 4

Mean individual difference scores and demographic information per condition.

	Condition 1 term default only framing	Condition 2 term default plus rate framing
Mean age	33.77 years (range 20-69, SD=9.91)	33.86 years (range 20-69, SD=11.55)
Gender	40.51% female	38.96% female
Education (Bachelor’s degree or higher)	58.23%	38.96%
Optimism (Max score 24)	16.54	17.30
High temporal preference for gains (10k now)	45.57%	57.14%
Financial literacy (Max score 13)	9.35	8.74
Numeracy (Max score 4)	1.70	1.66

Additional Analysis of Relevant Individual Difference Variables

This subsection of the results involves the analysis of individual difference variables which are shown in previous studies to impact of financial judgment and behaviour. Findings regarding the effects of these variables are mixed, and have often yielded unexpected and contradictory outcomes. Thus the purpose of their inclusion in this experiment is to assess the potential relevance of such factors in the context of this particular framing manipulation and task objective. Identifying significant relations between performance and individual differences, or interactions between the framing conditions and individual factors, would provide useful data relating to underlying mechanisms involved in the effectiveness of the manipulations. Increasing the understanding of how difference variables may interact with data disclosures is therefore important to further studies aimed at improving financial behaviour based on data framings.

To test for the impact of individual differences on repayment decisions and their potential interactions with the framing manipulations, a regression model was constructed with repayment amount as the dependent variable. The responses to the surveys for financial literacy, numeracy and optimism were converted into standard normal variables and entered into the model as predictors along with the temporal preference indicator, age, education and condition (framing manipulation).

The inclusion of all these factors in the model indicated that the framing manipulation in condition 2 significantly predicted the propensity to make higher monthly repayments, $R^2 = .122$, $F(8,147) = 3.714$, $p<.001$, $B = 235.806$, $t(147) = 3.169$, $p<.01$. Being female, $B = 172.406$ $t(147) = 2.289$, $p<.05$ and possessing higher financial literacy, $B = 89.152$, $t(147) = 2.098$, $p<.05$ were also significant predictors of higher repayment decisions across conditions. As expected, a significant negative relationship was also found between the temporal preference indicator and repayment amount, indicating that increases in the preference for smaller short term gains (i.e., the tendency to delay costs into the future) was predictive of smaller monthly repayment choices, $B = -171.974$ $t(147) = -2.179$, $p<.05$.

To examine the effects of these individual differences between the framing manipulations, separate regression models were then conducted per condition. In each model, the dependent variable was repayment amount and the individual difference parameters were entered as predictors one by one. In condition 1, significant positive associations were found between chosen repayment amount and financial literacy, $R^2 = .036$, $F(1,77) = 3.955$, $p=.05$, $B = 102.29$, $t(77) = 1.989$, $p=.05$, and education, $R^2 = .11$, $F(1,77) = 11.49$, $p<.01$, $B = 335.99$, $t(77) = 3.39$, $p<.01$, and a significant negative relationship was identified between repayment amount and temporal preference, $R^2 = .11$, $F(1,77) = 10.8$, $p<.01$, $B = -323.87$, $t(77) = -3.287$, $p<.01$. In condition 2, financial literacy was also found to positively correlate with repayment amount, $R^2 = .04$, $F(1,75) = 4.864$, $p<.05$, $B = 120.17$, $t(75) = 2.206$, $p<.05$. When regressing repayment amount on each of the individual difference variables simultaneously per condition, the main effect of financial literacy held for condition two. In condition 1, temporal preference and education also remained significant, however financial literacy was no longer found to be predictive of repayment amount, and instead gender presented as having a significant effect, $R^2 = .21$, $F(1,71) = 4.00$, $p<.001$, $B = 239.291$, $t(71) = 2.458$, $p<.05$.

Thus, although variations in the effects of individual differences were shown between participants in each condition, the key variables of importance to repayment amount when regressing across participants in both conditions were shown to be gender, financial literacy and temporal preference for gains.

To assess the effectiveness of the framing manipulations across participants on the basis of these individual difference measures, separate regression models were conducted with repayment amount across groups as the dependent variable and the main effect of each individual difference parameter as the predictor variable plus an interaction term between the predictor and the framing manipulation (condition). This revealed a main effect for condition, $R^2 = .08$, $F(3,152) = 5.673$, $p<.01$, $B = 389.2$, $t(152) = 3.703$, $p<.001$, with repayment amount increasing in condition 2, a main effect for education $B = 669.6$, $t(152) = 2.822$, $p<.01$, indicating that higher educational level was predictive of higher repayment choices, and a significant interaction term between condition and education, $B = -333.6$, $t(152) = -2.207$, $p<.05$, indicating that the lower[higher] the educational level, the higher[lower] the impact of the framing manipulation on chosen repayment amount across conditions.

6.2.3 Discussion

Disclosing repayment information using a default manipulation in the form of loan repayment costs for the loan at a 50% reduced term was shown to positively impact the propensity to make monthly mortgage repayments above the minimum necessary to clear the loan over the full term in current rates. This manipulation was therefore effective in improving judgments by reducing the propensity to make financial decisions based on optimistic estimates of interest rate variability. However, with the addition of future rate context information to the data environment in condition 2, the effectiveness of peoples' repayment judgments improved further.

The positioning of the default figure in the future rate context data acted to raise the saliency of the default which heighted the anchoring effect, and increased the range between the minimum and maximum values. Based on the tendency to apply arithmetic operations to key referents and seek upper and lower bounds in numeric judgment environments, the mechanisms in condition 2 combined to yield more effective, less optimistic repayment judgments which were closer in value to the upper bound of the judgment continuum. The findings have implications for data manipulations which may be effective in counteracting format biases and optimistic tendencies which are shown frequently to limit financial choice and behaviour across situations and groups.

The results reflected those found with the retail forecasters in experiment 1 of chapter 3, in that people in both conditions formed judgments by applying additive operations to the two relevant data points to yield the arithmetic mean of the values. In the case of the retailers, forming predictions using the arithmetic mean led to judgments akin to a two-point linear regression model. In the current experiment, the propensity to additively combine two salient values to make a choice yielded repayment judgments which reflected the arithmetic mean of the most and least optimal repayment options. Based on this tendency, the framing of the data in both conditions 1 and 2 functioned to yield repayment judgments which significantly exceeded the minimum (least optimal) amount. In this respect, both conditions were successful in mitigating optimistic biases in financial choice, with condition 2 generating a larger effect based on the larger absolute difference between the minimum and maximum values presented in the data environment.

The greater the value of the repayment judgments shown in condition 2 was not only associated with the tendency to arithmetically combine two points to generate a higher mean value in the context of a larger minimum and maximum range. The effect of disclosing the default value in the context of future rates further impacted the propensity to make higher repayment judgments by increasing the saliency of the rate information and the financial effects. Thus, combined with the propensity to form a judgment based on the arithmetic mean, increasing the saliency of potential costs by delivering future event information acted to highlight the size of the cost differences between the minimum and maximum suggested amounts and increase the relevance of the default in the data environment. When viewing the default, it is likely that disclosing the effects of future rate increases on loan repayments in condition 2 was effective in delivering a strong rationale for making a higher monthly repayment. This acted to emphasize the importance of the upper bound repayment figure in the judgment process to a greater extent than in condition 1. Heightening the relevance of the upper bound with the addition of the future rate context thus elicited a stronger anchoring effect from the default which increased peoples' inclination to make judgments which were closer in value to the upper bound.

From this perspective, the strength of the anchoring effect of the default is likely to be associated with the saliency of the context information. In condition 1 (without the added context data) participants' mean repayment judgment fell marginally below the average of the minimum and maximum suggested amounts, indicating a greater adjustment away from the maximum and towards the more optimistic minimum repayment option. In condition 2 however, the mean repayment fell above the average of the minimum and maximum values, indicating that the default manipulation was more effective in counteracting optimistic financial tendencies when positioned within the context of the future event data. It is therefore possible, that in the context of less salient additional information (e.g., context information which may be unrelated to rate rises and the cost implications), the default figure may have been less influential in the judgment process. This would likely result in an attenuation of the anchoring effect, leading to repayment decisions which would be closer in value to the lower end of the continuum and more akin to the less optimal minimum suggested amount.

With respect to individual difference factors, lower levels of financial literacy and higher temporal preference for short term gains were associated with the tendency to make smaller (less optimal) monthly repayment decisions across conditions. When controlling for the effects of individual differences on the effectiveness of the framing manipulations, educational level was shown to interact with the framing manipulations across conditions. This interaction revealed that each framing manipulation was more effective among people with a higher educational level. A larger proportion of participants in condition 1 possessed a higher educational level (Bachelors or Graduate degree) compared to those in condition 2 (see table 6.2), yet the increases in repayment amounts above the minimum in condition 1 were only moderate compared to the improvements in condition 2.

In conclusion to the supplementary analysis of individual factors, the significant interactions between the conditions and individual difference variables suggests that the effect of the framing manipulations in each condition may be associated with particular individual difference factors. However, when including all factors as covariates in the model to predict performance, the results remained robust which indicated the effectiveness of the framing manipulation in overriding any effects of individual differences. The increased effectiveness of condition 2 suggests that the disclosure of current versus future rate costs was a particularly advantageous component in addition to the framing of costs over different term horizons in overcoming limitations in financial choice related to lower educational level, financial literacy and optimistic tendencies associated with a higher temporal preference. The combined informational disclosure of a reduced term default positioned within future rate context data is therefore concluded as robust in heightening financial judgment performance despite individual variations. This finding points strongly to potential applications for supporting effective financial judgment and behaviour across socio-economic and educational groups and real-world choice contexts.

With respect to potential limitations, the study design involved repayment data disclosed in a format resembling a credit card statement, extracted from mortgage products available in the UK loan market in April 2015. In this sense, the framing was hypothetical, designed to impact monthly repayment decision making and behaviour in the same way that credit card statements are designed to increase the propensity to make higher monthly repayments. It must be noted therefore, that although the data presented in the experiment was representative of real-world products available in the market during April 2015, the participant group consisted of US citizens who were likely to be less familiar with the nature of the decision task compared to UK populations based on the American loan market which typically involves acquiring loans at fixed rates. Thus, rather than payments being subject to variable rates, US borrowers are able to make top up payments should and when they choose to.

Chapter 7

General Discussion

In this chapter, a summary of the main findings and scientific contributions is provided. This is followed by an overall review and evaluation of the key findings including alternative theoretical perspectives and limitations. This is proceeded by further research requirements and suggestions for practical applications and design, followed by final concluding comments.

7.1 Summary and Contributions

The aim of this thesis was to examine the biases and limitations in human probabilistic judgment which are associated with frames and formats of numerical information in different situations. By assessing performance in real-world judgment contexts, it was possible to contrast the performance of consumer and novice groups with that of professional populations using data frames and formats akin to those found in the real world.

The findings revealed robust similarities in judgment processes across groups which provide critical evidence increasing our understanding of the role of linear tendencies in human rationality. Independent of skill and experience, the propensity to perceive linear relations and apply judgment strategies based on assuming concrete values was highly robust across contexts varying in complexity and format. Whether predicting sales, selecting between financial products, or forecasting refugee camp data, people used minimal cues and systematically applied additive methods to percentage data and exponential trends to yield linear estimates in both rich and sparse informational contexts. It is concluded that the linear bias is a key defining characteristic of human probabilistic inference, stemming from the notion of cognitive algorithms designed to compute frequency data as underlying judgment processes. From the ecological viewpoint, cognitive algorithms developed in accordance with peoples' natural environment in a pre-mathematical world, thus creating the propensity to interpret and handle information in absolute terms or natural frequency formats.

This conclusion is supported by the strong facilitative effects of frequency formats, showing that communicating base rate information by representing relations between data using whole numbers, human rationality can be significantly improved. The use of icon arrays further increases peoples' ability to contextualize data when communicated in absolute frequencies by facilitating correspondences between the data and the real-world. This heightened data saliency aids the conceptualization of judgment and choice implications in meaningful, real-world terms which enables people to better understand the impact of judgments in a broader sense and over longer time horizons.

The findings deliver both theoretical and practical contributions. Across groups and individual differences, the effects of informational format and the tendency to linearly extrapolate are connected by the bias to perceive values in concrete terms and make sense of data by seeking simple reference points. People compare and combine referents using additive methods when inappropriate and adhere strongly to defaults when applied in simultaneous data frames in complex environments.

The practical contribution involves a framing manipulation which shows that format biases (i.e., additive processing) and optimism (i.e., the propensity to delay costs and downplay risks) can be counteracted in judgments involving percentages and exponential growth rates by using absolute formats and framing data using defaults positioned within future event context information. This framing manipulation was highly effective in improving loan choice and repayment judgments compared to information in standard finance industry formats. There is a strong potential to increase rationality using this framing manipulation in other financial settings, as well as domains such as health behaviour judgment and choice in which peoples' erroneous interpretation of percentages and non-linear relations between variables can negatively impact rationality, leading to maladaptive choices and behaviours.

7.2 Overview and Evaluation of the Main Findings

The findings from across the domains of retail, humanitarian aid and financial choice indicate that there were commonalities in the processes involved in the formation of judgments in each of these contexts which connect informational format with the linear bias. These processes which characterized judgments across settings involved the tendency to assume absolute values, seek relevant referents in data environments, and compare and combine the referents using arithmetic operations in numeric data. In graphical formats, the tendency to seek patterns and linearly project trends based on congruencies in noise was shown to characterize judgments (particularly for those with specialist domain knowledge). In prediction domains, the extrapolation of judgments based on arithmetic computations results in linear forecasts. Thus, in the context of non-linear growth trends and percentage data, arithmetic operations yield predictions which either under or overestimate outcomes, and 'optimistic' choices which reflect a high temporal preference. Results showed that in rich context environments involving absolute formats, adherence to defaults was strengthened by highly salient 'future event' information when framed simultaneously with information relating to events in the present. In this respect, optimistic tendencies relating to higher temporal discounting were shown to be minimised by this framing manipulation. The following section discusses how the judgment processes underpinning the linear bias are related in each setting.

In sparse data environments, people in both retail and humanitarian aid contexts made judgments based on linearly extrapolating the cues in the absence of additional information. Akin to a two-point linear regression model, the retail forecasters focused on the last two observations per trial and computed forecasts based on the absolute difference between these two values, regardless of the format or whether sequences trended linearly or exponentially. The fact that error was greater when observed percentages trended linearly and vice versa (i.e., for mismatches), supports the viewpoint that processing errors (i.e., adding and subtracting values rather than processing points multiplicatively) underpinned the linear bias in this context.

These findings indicate that when data is disclosed in numerical formats (as opposed to time series formats) people seek key reference points which they then combine using additive methods to generate linear judgments. Thus, in sparse contexts, where there is no additional data for comparison, people make sense of the environment by seeking relevant referents which can be easily and efficiently processed to make a fast, cognitively efficient judgment based on combining the referents arithmetically. A heuristic strategy such as this is logical, in that it minimizes effort and attentional resources where minimal information is available.

The same processes and methodology were applied when viewing sparse data in time series format, except that this environment was less conducive to the identification of key referents. As in the retail context, both humanitarians and non-humanitarians with no prior forecasting experience formed consistently linear forecasts amidst cyclically trending data, which reflected a linear model based on all three cues (as opposed the last two points). This suggests that in time series formats, the tendency to linearly extrapolate may be based more upon the visual comparison and combination of all available cues rather than a computation based on only two key data points. It is possible that the graphical display facilitated visual analysis to a greater extent than the numeric format which increased the inclination to additively compute estimates based on the arithmetic mean of two relevant points. Either way, the processes employed in each environment yielded consistently linear predictions based on the assumption of absolute values and the arithmetic processing of cues.

When examining performance in rich data environments involving numerical formats, people making financial repayment judgments showed the same propensity to seek relevant reference points and combined them additively just as the retailers did in sparse numerical contexts. In this setting however, a default in the form of loan costs for a reduced mortgage term were displayed alongside costs for the full term in condition 1. In condition 2, the default was positioned in rich contextual data relating to loan costs in future rates. This future rate data for both the full and default (reduced) term information was disclosed simultaneously with the current rate data.

When seeking meaning in more complex data environments such as those in experiment 4 which involved a wide range of cues (i.e., the reduced and full term cost information), people sought key referents in the form of the minimum and maximum repayment amounts. These two values were then additively combined to generate a decision based on the arithmetic mean. This indicates that when additional information is present, people make sense of the information in the environment by locating referents which represent an upper and lower bound. In the context of a repayment decision, this is a logical strategy in that it provides a summary of the most relevant information to the judgment and generates a range within which a repayment judgment can be made.

In this setting, it is thus apparent that the framing mechanisms involving a numeric format and the simultaneous disclosure of default information in concrete terms worked in congruence with the judgment processes underpinning linear tendencies. The effect of the default thus interacted positively with the tendency to seek and combine referents based on minimum and maximum values which made this mechanism effective in operationalizing the linear bias to improve decision performance rather than to hinder it.

When presented with rich context information in time series formats, the propensity to make linear judgments based on additively combining cues remained robust. However, in this setting, non-causal cues as opposed to relevant context information was added to the data environment to create noise, which resulted in an increase in judgment error. In this situation, the same propensity to compare and combine data when simultaneously viewing target and added context cues was evident. However, it is possible that the time series format made the identification of useful numerical referents more difficult. Compounded by the lack of relevance of the context information, people thus sought meaning in the data environment based on visually comparing cues to identify similarities in the directions of the trends. Rather than seeking a clear upper and lower bound, this visual method appeared to characterize judgments, enabling people to derive a rationale for linearly projecting the target data in a particular direction.

As in the sparse time series format, the target data cues were again combined additively to generate linear projections, however the direction of the linear extrapolations were guided by the direction of the trends in the noise. This shows that in complex environments where additional data is unrelated (and thus of no use), people are still prone to utilize the data to help guide and rationalize their judgments. Seeking patterns and similarities between data is therefore likely to be a strategy for making sense in incomprehensible environments when a clear judgment range is not numerically identifiable. This suggests that the increased judgment error in noise shown among humanitarians and novices is likely to stem not simply from linearly extending the target series, but from actively seeking patterns in noise to inform judgment directions.

The key factors which influenced the linear bias in these two complex data environments are therefore the format of the data and context information, and the causality of the additional cues. Both situations involved rich information, however performance in the financial setting was facilitated whereas it was hindered in the humanitarian domain. It is probable that the time series format increased the difficulty of identifying referents, leading people to compare cues using a visual heuristic strategy based on similarities. Target data cues were then combined additively to produce linear extrapolations which followed the same direction of the other cues.

However, in the financial scenario, the additional context information was delivered in numerical formats and the added information was highly relevant to the task objective. Viewing the data in absolute formats in this situation made it easier for people to locate referents and established a clear meaning from the data, based on an upper and lower bound. In the context of the highly relevant additional information (i.e., the reduced term data), the propensity to seek an upper and lower bound and apply arithmetic operations becomes adaptive, based on people comparing and combining values across all observable cues. Therefore, the default (i.e., the reduced term data) in this particular setting represents an important mechanism impacting human judgment by harnessing the propensity to seek and compare key referents in complexity, and apply arithmetic operations. When presented simultaneously with the target information (i.e., the full term data) in concrete terms, the default utilized the processes involved in the linear bias to improve rationality by converting human propensities from erroneous processes into adaptive strategies within the given data environment. Rather than minimizing the linear bias therefore, the default increased performance by working in accordance with the propensities which promote it.

The default further amplified judgment performance when the anchoring effect was amplified by positioning it in the context of future rate information. For example, the findings showed that loan repayment judgments were more optimal when people viewed the default (i.e., the reduced term data) alongside the full term data in 'future rate' context compared to when viewing the default information alongside the full term data in 'current rates' only. By disclosing the default in a future rate frame simultaneously with the target data in current rates, the absolute difference between the minimum and maximum suggested repayment amounts increased. This increase between the upper and lower bound improved judgment effectiveness (i.e., the amount people decided to repay) not only by increasing the value of the arithmetic mean, but also by heightening the saliency of costs associated with future interest rates, which within a financial choice scenario is extremely important.

The strong relevance of the future rate information to the judgment task raised the significance of the increased costs, drawing greater attention to the default information containing the upper bound (i.e., the reduced term costs in future rates). As a result, the default had a stronger anchoring effect compared to when it conveyed reduced term costs in current rates. Under these conditions, people made lesser adjustments away from the default to form repayment judgments which fell above the arithmetic mean of the upper and lower bound. In contrast, when the default was positioned in a current rates frame, judgments fell slightly below the mean.

In sum, when observing numerical values without surrounding contextual information (i.e., in the retail environment), people utilize minimal data points (i.e., the two most relevant observations) and apply fast and frugal arithmetic operations. Where multiple cues exist, (i.e., in the humanitarian setting), the same linear prediction bias characterizes judgments, except that people use the surrounding information to guide their estimates as a means of deriving rationale for judgments based on patterns in the noise. Together, these findings point to the importance of the available contextual information in judgment situations. In the absence of contextual cues, people resort to basic linear strategies. When context is present, they combine the additional information to either guide linear projections, or to derive the arithmetic mean of two referent points. Depending on the causality and format, the propensity to combine context information can be either harmful or adaptive. Where irrelevant (i.e., in the context of non-causal context cues), the comparison and combination of cues leads to erroneous judgments which are linear in nature. Conversely, in the context of relevant, causal data cues and absolute numeric formats, the linear bias can be adaptive when default values are positioned within the context of future event data frames.

The findings indicate that processes underpinning the linear bias characterize judgments across contexts and groups, and that data format impacts the linear bias via the human propensity to think in concrete terms, look for useful numerical benchmarks and add and subtract values to make choices and predictions. Depending on the format and relevance of contextual information in complex environments, data can be framed to counteract the negative effects of the linear bias in complex judgment domains involving non-linear relations. The robustness of the linear bias shown throughout the findings suggests that interventions aimed at minimizing it are unlikely to be effective. The results relating to the effectiveness of 'future event' defaults therefore deliver important theoretical and practical implications, indicating that it may be possible to facilitate more accurate estimates of the effects of exponential functions by displaying outcomes on a temporal continuum.

Future event defaults delivered in concrete terms and disclosed simultaneously with current event data are effective in creating highly salient context which improves judgment by increasing the relevance of future events, and thus the anchoring effect of the default. It is possible therefore, that simultaneously framing 'current' versus 'future' outcomes at different points in time will facilitate more detailed comprehension of non-linear relations between variables, and help to anchor people even more on 'future' framed defaults. As described above, future event defaults may therefore provide a framing mechanism for improving human rationality by conveying information to work in accordance with linear biases. This approach is likely to be more effective than applying interventions aimed at correcting or retraining cognitive biases and irrational judgment processes.

7.2.1 Alternative Theoretical Perspectives and Limitations in Design

The remainder of this section evaluates the findings discussed above, delivering possible alternative perspectives and explanations for the results, including experimental effects and shortcomings in design which may have contributed to the results.

When considering the linear bias in the retail forecasters for example, it is possible that more ‘conservative’ estimates based on forecasting towards the mean are less erroneous when considered in relation to wider supply chain factors and longer time horizons. When testing performance in real world settings, the tendency to under-forecast and over-forecast may be adaptive in the context of other, un-modelled domain specific information. As a result, tendencies to project trends linearly could reflect tacit knowledge and beliefs about the data which are used in shaping judgments beyond the observable cues. However, data relating to peoples’ individual beliefs and knowledge about the domain and the data were not captured in the experiment. Further investigation using rich context is necessary to indicate how the linear bias may be impacted by beliefs and knowledge, based on how people synthesize data with beliefs and use additional context to guide judgments. Based on the sparse numerical cues used in the experiment, it is not possible to draw conclusions regarding how knowledge and beliefs are integrated in the judgment processes to yield the linear bias. However, it is clear that data format and framing is important to interpretation of cues and how people integrate additional data in the formation of linear judgments.

This suggests that beliefs and knowledge will also be impacted by format. It is probable that they will shape judgments by feeding back into the interpretation of additional context data, and applied as a means of confirming initial interpretations of the data, based upon the format. For example, Hohle and Teigen (2015) showed that participants predicted experts’ climate change forecasts by endorsing trends which corresponded with their individual beliefs about the consequences of climatic changes. Predominantly, people continued the extension of upward trends for sea level and temperature rises and downward trends for grain productivity. This propensity was particularly marked among people with stronger beliefs in anthropogenic climate changes.

It is also possible that the linear bias may have arisen based on experience of the effects of trend-damping. For example, damping statistical model trends in the retail forecasting domain may have proven to be an effective strategy in the past, particularly in the context of sparse data, hence the strong propensity to linearly extend trends in the experiment. If the retailers were to show evidence of accurate non-linear extrapolation in contexts other than supply chain, this would indicate that the linear bias may be more associated with beliefs based on domain-specific experience, rather than global judgmental processes. Testing this hypothesis is warranted, because there may be a stronger element of domain specificity involved in the linear bias than these results indicate. Depending on the nature of the data, the linear bias may be more 'rational' in certain environments, and the impact of peoples' knowledge and beliefs on probability estimates may vary greatly between settings (Lawrence, Goodwin, O'Connor, & Önkal, 2006).

Making conservative estimates or under-adjustments to statistical forecasts could also be an expression of the 'golden rule' of forecasting (Armstrong, Green & Graefe, 2015), which states that 'forecasters should be conservative by making proper use of cumulative knowledge and not go beyond that knowledge'. It is feasible that the conservativeness of the retailers' estimates (i.e., the tendency to forecast towards the mean) reflected this concept, based on the assumption that their knowledge is not sufficiently greater than that of the systems or other experts involved in generating model data. A similar explanation to this is the 'asymmetric loss function' which accounts for operational logistics, marketing and sales forecasters' propensity to over-predict in an effort to minimize losses. This is based on the concept that over-ordering stock has less negative financial impact in the long term compared to under-ordering which risks empty stores. The retailers may therefore have underestimated increases as a mechanism for moderating risk associated with over-estimating returns (Lawrence, O'Connor, & Edmundson, 2000). Alternatively, the over-forecasting of decreases may reflect an optimistic belief that sales will not decline to extent the data suggests.

In the case of both the retailers and the humanitarian aid workers in experiment 2, it is possible that the artificiality of the situation and the experimental presentation of the data encouraged participants to seek and extract trends leading to prediction anomalies or atypical strategies which would not occur in the real-world setting. In unfamiliar situations for example, people are shown to make judgments based on strategies which are better suited to other circumstances which they are more familiar with (Oskarsson, Van Boven, McClelland & Hastie, 2009). An absence of real-world background information may have increased the propensity to seek patterns and regularities in the case where the underlying mechanisms and causal factors where unknown (Elliman, 2006).

Throughout JDM research, biases which are identified are often recognised as arising from participants assuming that all available data is there to be used, and thus incorporating it in their judgments (Wanke, 2007). This could account for aid workers and novices' tendency to linearly project trends in noise based on a perceived 'consensus' between the directions of all observed cues. This finding may therefore have reflected an expectation bias, whereby people assumed that the trends were to be projected in the same direction, without considering what the task required or represented in reality. These possibilities point to some of the well-recognised difficulties and limitations associated with experimentally replicating field environments (Lipshitz, 1993; Lipshitz, Klein, Orasanu & Salas, 2001).

When viewed as an experimental context effect, the tendency to trend damp and anti-damp (i.e., under-estimate increases and over-estimate decreases) may be considered an adaptive strategy, despite that linearly extrapolating the data created error in the particular data environments tested. For example, people are shown to regress towards the mean trend encountered when forming forecasts in multiple trials (Stevens & Greenbaum, 1966; Warren, 1985). In this respect, linear regression towards the mean may be a rational strategy for minimizing expected error by forecasting according to the estimated mean trend, or by taking a weighted average of the estimated trend on a particular trial and the estimated mean trend (Harvey & Reimers, 2013). A similar perspective to this is the adaptation hypothesis which also explains trend damping as an adaptive strategy, suggesting that people are predisposed to forecast in accordance with degrees of growth and decay which are representative of natural environments.

Patterns in ecological growth and decline, found widely throughout natural time series, tend to follow sigmoidal functions resulting from an initial positive acceleration followed by trend damping due to depletion of resources. This leads to a levelling off which reflects the carrying capacity of the environment (Tsoularis & Wallace, 2002). Trend-damping may be considered as the tendency to predict in accordance with these natural functions, hence the apparent effectiveness of trend-damp in various contexts (Collopy & Armstrong, 1992; Gardner & McKenzie, 1985).

Interestingly, Harvey and Reimers (2013) found that the tendency to trend damp in accordance with representative patterns of growth and decline in the natural environment occurs independently of context effects. In a single-shot between-participants experiment involving a single time-series, damping was shown to occur with positively accelerated series and anti-damping with negatively accelerated series. In the absence of repeated trials, these results suggest that rather than a bias created by context effects, trend damping may be the result of long-term adaptation to the natural environment from which trends occur.

In terms of the prediction patterns shown among retail forecasters, the propensity to extrapolate linearly could thus be an expression of the tendency to trend damp and anti-damp in accordance with patterns of growth and decline in the environment. These patterns may be associated directly with ecological factors such as seasonal variance, as well as specific supply chain factors, such as manufacturing lead times. For example, naturally occurring growth and decay is reflected in weather changes which shape patterns in supply and demand. Ecological factors are also likely to impact factors such as product lead times, and may thus have an indirect influence on supply and demand.

Factors such as lead times also mean that estimates have to be formed for events far into the future, which suggests that both retailers and aid workers' may be better at forecasting over longer time horizons, using more cues compared to the short three-point sequences used in the experiments. The inclination to moderate increasing and decreasing trends may therefore be an effective 'rational' strategy over longer horizons in complex environments. Findings have shown for example, that novices with no prior experience of a given field tend to employ strategies which are more adapted to shorter time horizons compared to those with domain specific knowledge (Thomson, Pollock, Henriksen, & Macaulay, 2004). In this view, the linear bias (i.e., the propensity to trend damp) may arise from people being more adapted to forecast over longer time horizons based on the effectiveness of the strategy in terms of real-world supply chain or refugee camp morbidity indicator outcomes. Depending on whether the variance in these factors reflects natural patterns of growth and decline, it is possible that the linear bias in this context represents peoples' inherent awareness of such variance, thus supporting the ecological perspective.

Although this view does not effectively explain the importance of beliefs and domain knowledge to judgment performance, it posits the idea that the linear bias could be related to judgment processes which are attuned to naturally occurring variance and that human judgment processes yield effective estimates when applied in the context of time lines which more accurately reflect those found in the real-world environment. It is possible therefore that people with more experience may potentially be more sensitive to environmental patterns of growth and decay, and thus possess a greater ability to form more effective judgments in the context of more data over longer time spans. Further studies are necessary to test this possibility in the retail and aid environments, and other complex domains involving people with and without specific experience in the field.

In the domain of financial choice, the ecological viewpoint could also account for the linear bias and its harmful effects. For example, a tendency to trend damp based on judgment processes being geared for natural growth and decline in the context of compound interest (exponential growth) will yield underestimates of future outcomes. Moreover, the effects of compounding mean that inaccuracies increase (as opposed to decrease) over longer time spans. The linear bias may therefore be particularly non-adaptive in environments where trends follow 'un-natural' (non-sigmoidal) patterns, such as the exponential growth of interest rates. This notion of the linear bias as a product of adaptation to the environment fits with the concept of 'natural algorithms' forming the basis of human rationality, which involve processing data in terms of absolute counts, and sampling from the environment in natural frequencies.

In the context of exponential growth trends, optimism is another factor contributing to judgment inaccuracy. In domains such as financial decision making, consumer choice, or health, judgments which involve computing percentages and modelling exponential trends often include other factors such as marketing, or public health messages which can impact peoples' motives and behaviours. For example, choosing between credit cards, or whether to give up smoking, or to take up exercise are all judgments which involve extrapolating compound effects (i.e., exponential functions) and are subject to powerful individual motives and drivers. In these contexts, there is a strong propensity to think optimistically and underestimate future outcomes. The linear bias is thus behaviourally amplified in such environments which acts to increase its detrimental impact on choice and behaviour.

The results of the framing manipulation tested in the financial repayment judgment context demonstrate a means of improving judgments by counteracting optimism where compounding is concerned. From the ecological viewpoint, the framing manipulation may be increasing peoples' ability to control for the effects of 'un-natural' exponential trends, by increasing rationality in accordance with statistical models rather than naturally occurring variance. The framing effects also held when factoring in other important individual differences. Participants with lower levels of education, financial literacy and a preference for short-terms gains (i.e., financial optimism) made significantly higher (more optimal) repayment judgments compared to a minimum suggested amount in current rates over the full term. Previously, these individual differences have been associated with poor financial judgment and behaviour. However, the robustness of the framing effect across groups indicates that the framing of rates in the future is likely to be effective in overcoming such effects. Future event frames may therefore be a useful manipulation for promoting behaviour and choice in other exponential growth contexts where individual differences and behavioural motives are likely to have a strong impact.

It is important to note however, that the association between financial literacy and behaviour is unclear, with some studies yielding negative associations between financial literacy and performance (e.g., Newall, 2016b), and others showing that raising financial literacy does not lead to improvements in judgments (e.g., Fernandes et al., 2014). The current findings thus provide a strong basis for further examination of the framing effect among individual differences in complex non-linear judgment domains to identify mechanisms which may be mediating performance effects in these groups, and whether or not they hold across settings.

In summary, various explanations for the linear bias have been posited, which include experimental effects and the impacts of domain specific knowledge and experience, and trend damping as a reflection of the propensity to predict in accordance with natural growth and decline. From these perspectives, the linear bias is potentially adaptive in the context of more data and longer time horizons which could be more representative of the temporality of events in the real-world (i.e., in 'nature'). In the context of exponential growth (e.g., compound interest rates) which are less reflective of patterns found in nature, the linear bias is detrimental to performance, particularly so over longer time horizons. Presenting data in absolute terms with defaults positioned in 'future outcome' information could be a useful mechanism for optimizing judgment tendencies in accordance with the linear bias to yield effective choices and behaviours in the context of exponential trends and complex, varied cues.

Although performance may be difficult to determine based on only three data points (particularly in the context of noise), the robustness of the linear bias shown across groups and contexts throughout the prior chapters suggests that barriers to computation based on the tendency to think in concrete terms and apply simultaneous processing to numerical data are key factors underpinning the widespread tendency to make linear judgments amidst non-linear trends and percentages.

Differences in how the linear bias impacts performance is dependent on data frame and format, and the richness and relevance of context information. These components then interact with judgment processes to shape rationality depending on individual motives, knowledge, experience and beliefs in the context of the task and specific judgment objectives. In both sparse and complex environments, it is likely that experience and beliefs etc., are applied in the judgment process to aid in the interpretation and synthesis of cues. Based on mechanisms identified in the 'summary and contributions' above, it is posited that people linearly extrapolate data (or adjust additively computed estimates) in ways which are confirmatory of beliefs and experience. In time series formats for example, this can result in following trends and seeking patterns in data which fit with expectations and are congruent with beliefs. In numeric formats, it is likely that people look for relevant reference points to compute estimates which are then adjusted in accordance with beliefs.

7.3 Further Research

There are possible modifications and extensions which could be made to the experiments in each chapter to further examine the processes underpinning the linear bias in the settings investigated. Firstly, it would be useful to deliver a formal measure of numerical data processing in each context to verify the propensity to additively process percentage data in the settings investigated. This would strengthen the evidence for this characteristic as key human trait underpinning the linear bias. For example, this could involve (as previous studies have done) assessing retailers, aid workers and consumers' ability to determine whether the final return value of a stock rising and falling in percentage points is smaller or larger than the initial value.

Overestimating investment returns in this task indicates the tendency to simultaneously process information (Newall, 2016a), which would confirm format confusion and failure to multiplicatively processes as a key mechanism contributing to the linear bias when viewing percentages and exponential trends in each of the settings investigated. A formal measure of numeracy could also provide support for the robustness of percentage processing errors across groups and settings. For example, numeracy is shown to correlate not only with numerical skills (Tubau, 2008), but also with performance in Bayesian inference tasks when data is presented in both probability (percentage) and frequency formats (Chapman & Liu, 2009; Johnson & Tubau, 2015).

Another important set of variables warranting further analysis is the impact of peoples' individual beliefs based on experience, knowledge and motives on judgment formation and how they affect the linear bias. In conjunction with how format influences peoples' perception of numerical information, it is likely that beliefs about the data and its meaning in the real-world are imported into the judgment process to influence perceptions of context information and the interpretation of patterns and relevant reference points in the data environment. There is likely to be a tendency to then seek features and meanings in the data with support beliefs, and make judgments based on additively processing the data to form linear estimates which reflect those beliefs.

It is necessary to examine how these factors interact with data format, frames and context information to impact judgment processes in different settings. In domains such as retail and humanitarian aid for example, methodologies used in naturalistic decision making research could be usefully applied to collect data in investigative field experiments. Methods such as these would deliver insights into how experience and beliefs are combined with data and incorporated into the linear judgment processes. As outlined in the above, it is probable that people form linear judgments in accordance with beliefs, experience and motives based on seeking patterns in data which align with these factors. In this sense, individual expectations and experiences feedback into judgmental processes to inform the interpretation of data and guide judgment adjustments via linear perceptual processes. (These processes may differ therefore, depending on data format and the nature of context information).

A closer analysis of the behavioural and motivational factors impacting judgment formation in consumer choice and financial decision making is also required. In these areas, it is important to understand how individual motives and factors such as temporal discounting can influence peoples' interpretation of information and contribute to the linear bias. In situations such as online shopping, lottery playing, or selecting between credit cards for example, tendencies such as optimism and impulsiveness are likely to play a key role in mediating or amplifying the linear bias. These tendencies are connected with high temporal preference (or the strong desire to maximize short term gains). This may be reflected in irrational, non-cost effective purchase decisions, and the propensity to select options with the lowest repayment costs, despite the fact that these choices incur higher costs overall. It is necessary to tease apart the effects of motives and beliefs in these situations to understand how irrational choices and behaviours (associated with the linear bias) are connected to informational formats and framings steering the judgmental processes.

The findings from across all domains investigated indicate that format is an important factor determining the judgment processes which contribute to the linear bias. To verify the effects of format, it is necessary to test whether the judgment processes are impacted in the same way when alternating data formats between each of the environments assessed. Drawing parallels between performance across fields based on format is important to verify the proposed mechanisms for how data format is connected to the linear bias. For example, by presenting the retail forecasters with sales data in sparse time series format, we would expect to see the same tendency to linearly project trends as was shown by the aid workers. Likewise, if aid workers were shown numeric formats in sparse context trials, we would expect the same propensity to additively process percentages and extrapolate trends based on the last two observations as shown among retailers.

Evidence suggests that people tend to interpret statistical information (specifically percentage data) as 'benign', which can lead to less accurate assessments of risk compared to when information is viewed in frequency formats, which more effectively promote the meaningfulness of the information (Slovic, Monahan & MacGregor, 2000). In the financial choice setting, it may therefore be possible to further optimize rationality by using graphical displays to communicate the temporal effects of compounding in terms of absolute costs when considering different rates, product types, and loan terms. A visual representation of the exponential relations could be more effective in countering optimism (or high discounting) than simultaneously displaying the current versus future rate information in numeric formats.

It would be interesting to test if displaying current versus future rate data and repayment information graphically (either in line or bar format) would further optimize the framing effect to facilitate loan choice and repayment decision making. For instance, a dynamic line or bar format to communicate loan options and monthly costs (in which people manipulate variables to see physical effects in terms of size/shape changes) could increase visualization of the data. Being able to see the physical effects of compounding in dynamic formats is likely to heighten the saliency of future rates, thus drawing greater attention to the real-word implications. Thus, graphical formats may more effectively communicate the concepts and relations between variables involved in complex non-linear environments (particularly in low educational groups).

Another means of increasing the meaningfulness and correspondence of the data with the real-world could be to use icons (i.e., graphical representations of the variables). When applied in frequency format, the use of dynamic icon arrays is shown to significantly improve statistical rationality, particularly among lower numeracy populations (Garcia-Retamero & Galesic, 2009; Okan, Garcia-Retamero, Cokely & Maldonado, 2012). The mechanism is based on heightening the data saliency by using cues which reflect real-world populations, items or events. The frequency display thus conveys the set relations between the variables (i.e., the base rate information) to facilitate the comprehension of a risk factor as a proportion of a population whole. This is shown to greatly improve judgment by making transparent the implications of statistical data in terms of the real-world impacts on individuals and groups.

It is possible that frequency formats using icon arrays may also be useful in the financial judgment setting to communicate cost differences between loan types, and between full versus reduced term information in current versus future rate frames. This could be effective by facilitating the comprehension of the temporal effects of compounding in relation to the continuum of absolute costs for different product types and loan terms over time more effectively than simply numeric formats involving a default in future rate context. Based on the results among lower educational populations, it is also possible that using icons in frequency format may be more effective in promoting judgmental rationality across groups compared to line or bar graph displays used to convey the framing effect.

Other ways of facilitating judgment among low educational, or socio-economic groups could involve framing the data in terms of ‘losses’ versus ‘gains’. For example, in addition to disclosing costs for the reduced term, it might increase the effect to also show the saving in relation to the full term alternative using the following manipulation:

"You will save $104,703.80 over 10 years if you repay $2,811.12 per month" as opposed to "If you repay $2,811.12 per month, it will take 10 years to clear the balance", and vice versa for the full term option;

"You will lose $104,703.80 over 20 years if you repay $1,841.83 per month" rather than "If you repay $1,841.83 per month, it will take 20 years to clear the balance".

Loss/gain framing could also be applied to the interest costs in the same way, for example;

"You will save $6,007.93 in interest over 10 years" versus "You will lose $6,007.93 in interest over 20 years".

One interesting possibility is that framing costs associated with time spans as gains versus losses might interact with temporal preference. It would be worth investigating for example, whether people with higher temporal preference are more sensitive to higher long term costs when framed as gains and less sensitive to lower short term gains when framed as losses. Based on the ambiguity of the relation between financial literacy and financial judgment and behaviour (e.g., Fernandes et al., 2014), examining loss versus gain framings as a potential mediator of financial judgment among groups varying in educational abilities and temporal preferences could be beneficial. As outlined above, loss versus gain disclosures will act to increase the saliency of cost differences in easily relatable terms, thus potentially having a strong impact on rationality among lower educational groups.

Increasing the effectiveness of informational formats and frames in financial judgment settings could also increase the quantity of data people are capable of processing. Past findings have shown that increasing the number of choice alternatives in various decision environments leads to a decline in judgment effectiveness (Benartzi & Thaler, 2002; Sethi-Iyengar, Huberman & Jiang, 2004), due to increases in noise which accompany more informational cues (e.g., Harvey, Bolger & McClelland, 1994). However, it is possible that graphical formats may increase the number of choice alternatives people are able to process onscreen at any one time. In a domain which involves hundreds of choice alternatives, exploring graphical data formats to increase the proportion of data people can simultaneously process could greatly aid judgment effectiveness.

The way in which people process complex cues when disclosed in simultaneous versus sequential frames requires closer analysis to understand how the process of comparative analysis contributes to judgment performance. Reconstructing the loan choice and repayment decision making experiments as eye-tracking studies would show how data was compared in each framing. The time spent viewing each rate frame and the number of comparisons made between each frame would indicate which data points were most important to the judgment process. Recording eye tracking information in combination with screen click data in the loan choice experiments would also show how people choose to reorder loans, providing data relating to which individual loan product attributes people deemed most important to the judgment process.

Together, these results could provide further evidence of factors underpinning the linear bias in complex online choice environments. If tested in relation to standard financial data formats for example, we may see that people are inclined to focus more on percentage information despite miscomprehension of percentage formats, or that people pay particular attention to certain cues, or make particular data comparisons which are ineffective when analysing financial product attributes.

In addition to testing alternative data formats, it is also necessary to assess how the linear bias may be effected by data presented over longer horizons and in different types of context information. It is possible that the three-point time horizons used to test performance in the retail and aid environments were not representative of real-world data, and thus generated unreliable reflections of peoples' judgment processes. Examining performance over longer horizons in sparse and rich contexts is required to assess the robustness of the judgmental processes yielding the linear bias. For example, assessing judgment performance in relation to real-world outcomes (i.e., actual events) in the retail and aid environment may be useful in determining whether the linear bias (i.e., the propensity to trend damp) is effective over the longer time spans (i.e., leads to increased performance in terms of actual outcomes).

As postulated in the 'alternative theoretical perspectives' section, we may find that people in complex environments are more capable of making effective predictions over longer time horizons. As well as reflecting adaptation to the variance in domain specific factors (e.g., product lead times), it could also represent a sensitivity to patterns of variance in the natural world, thus supporting the adaptation perspective. This is possible, considering that naturally occurring patterns of growth and decline are echoed in domain specific factors (e.g., seasonally determined morbidity indicators).

Based on the effects of sparse versus rich context on judgment processes in numeric and time series formats, it is likely that adding noise or relevant information to each environment will result in the same processing mechanisms. For example, we would expect that adding causal cues in numeric format to the retail environment would result in people seeking an upper and lower bound across the whole data environment, just as they did when viewing loan repayment data in current versus future rates. Thus, rather than computing the mean of the last two observations in each sequence, people would additively combine the last point in the target sequence with another point (or the mean of multiple points) in the context data. Depending on nature (and saliency) of the context information, people are likely to then adjust estimates above or below the mean, based on individual experience and beliefs.

In the context of noise however, there is likely to be parallels with the aid workers, whereby people viewing noisy data numeric formats create meaning amidst the complexity by using simple heuristic strategies to identify patterns and similarities between all the observable cues. On strategy people might employ, could involve comparing the absolute differences between points two and three across all cues to identify similarities in trend directions. Akin to the aid worker, they may then incorporate noise into the judgment by predicting increases when all cues are rising from points two to three (based on the arithmetic mean of the last two points), and decreases when all cues are falling. Another possibility is that people will seek to match the final observations in each sequence based on the values in absolute terms, then additively combine points which have a similar absolute value to yield estimates based on the arithmetic mean of these referents.

With respect to the ‘trend effect’ identified in the aid setting, it is necessary to examine its robustness in more varied time series data to see whether there is a limit in terms of complexity in which data becomes too noisy for trends to be visually compared and weighted. At this point, people may resort to different strategies to determine the direction in which to linearly project the target data. However, in the context of causal cues in time series formats, it is likely that the ‘trend effect’ will diminish because of the increase in meaningfulness of the additional information. This will decrease the inclination to look for ‘meaningful’ patterns in the data based on similarities in cue directions, and increase the propensity to evaluate the data in the context of individual knowledge, experience and beliefs. The tendency to linearly project trends is likely to remain robust, based on the same tendency to additively combine cues in both noise and relevant context. However, the more relevant or salient the context, the more likely it will be to activate peoples' individual experiences and beliefs, leading to judgments which are shaped and adjusted to a larger extent by these human factors than judgments formed in situations involving non-causal, noisy data.

7.3.1 Practical Applications for Improving Judgment and Behaviour

The findings offer several implications for design which may be effective in improving human rationality and judgment in the context of consumer choice and behaviour change, and also in professional settings which involve the formation of formal probabilistic judgments in complex, non-linear environments.

Based on the findings in the retail and humanitarian aid environments, it may be considered that the linear bias renders human judgment relatively ineffective in noisy real-world domains. However, results also suggest that there are many un-modelled factors involved in complex environments which contribute to wider forecast performance, beyond the observable data. It is likely that peoples' inherent knowledge and experience of a particular context provides an overall awareness and ability to conceptually synthesize the factors involved. This overarching understanding of the environment is likely to indirectly impact forecast performance via judgment processes and applications which are distinct from statistical knowledge or data handling skills. It is important therefore to facilitate human domain knowledge in complex judgment environments to maximize the benefits of integrating peoples' experience and awareness of un-modelled factors, whilst minimizing the potentially harmful effects of the linear bias.

This may be achieved by applying people in the development and monitoring of statistical models, rather than relying on people to form unaided judgments in noisy real-world settings. By using model forecasts as a basis, human knowledge and experience may be integrating into the data environment via the monitoring and controlling of statistical model predictions. This approach may be effective in extracting the benefits of peoples' additional knowledge by changing the processes underpinning how knowledge and experience is applied to the data. This change in method may thus facilitate people in making informed judgmental adjustments to unbiased statistical models without the risks of forecasts becoming linearly biased by the effects of the judgment mechanisms shown in this thesis to be involved when people form judgments unaided in noisy environments. In this view therefore, the advantages of human experience may be utilized and isolated to some extent from the tendency to erroneously seek trends and interpret congruencies in noise which do not exist.

There is recent evidence which suggests this method of applying human experience and judgment processes is likely to be effective in the aid domain. In this instance, people are employed in the model building, monitoring and interpretation of systems designed to forecast humanitarian aid demands ahead of climate-related disasters (Cousin, 2015). People are also shown to be highly effective in the field of weather forecasting when applied to identify important variables in simulations (e.g., barometric pressure) which statistical models then weight to form weather forecasts. In this area, peoples' domain experience and ability to perform sanity checks (Silver, 2013) are shown to increase the predictive accuracy of weather simulations by up to twenty-five percent (Lynch, 2006, 2008). Combined with examples such as these, the current findings imply significant benefits from assimilating human experience and tacit knowledge with the statistical modelling necessary for analysing, interpreting and utilizing ever-growing data sets in increasingly complex environments.

With regard to real-world probabilistic decision making in non-professional settings, the effectiveness of the future rate default framing manipulation in the financial choice and repayment judgment contexts, indicates there is a strong potential to apply this framing to increase rationality, choice and behaviour in saving, investment and insurance decision making, as well as other settings such as health. Judgments relating to health behaviours are commonly subject to optimistic biases which can lead people to downplay the negative effects of certain actions and choices and delay making decisions which effect behavioural changes. Optimistic tendencies such as these are compounded by problems in interpreting percentage information and estimating the effects of non-linear relations between variables (or actions) and outcomes.

Health outcomes are commonly related to actions and judgments in exponential relations (e.g., health exponentially improves with smoking cessation, reduced alcohol or fast food consumption, and increased exercise and healthy eating). The difficulties people experience in modelling exponential trends are therefore likely to impact the perceived benefits of making certain decisions and engaging in particular health behaviours. This may be reflected for example, in the tendency to regard exercise as an ineffective use of time, or perceive the adversity involved in giving up smoking as outweighing the perceived benefits, based on an inaccurate estimate of the time necessary to achieve 'worth-while' health advantages.

These perceptions are related to the tendency to assume linear relations between actions and events which result in underestimates of the benefits of implementing behavioural changes sooner, or the perception that it will take far longer to experience the benefits than is actually the case in reality. Thus, peoples' failure to action change is likely to be associated with biased perceptions of costs (e.g., time, expense, discomfort) as outweighing the perceived advantages over given time horizons. However, the advantages of stopping smoking, exercising (particularly high intensity interval training, for instance), drinking less alcohol, eating healthier food, or saving money are all examples of actions with compound effects. I.e., the benefits are larger and accrued more quickly over a given time span than people may predict.

It could be possible therefore, to motivate more optimal health related judgments and choices (thus reducing optimistic tendencies which are associated with delaying judgments and actions) by framing present behaviour (and the health/financial implications) in relation to target behaviour (and the health/financial implications) using a 'current' versus 'future' health status/financial advantage frame for a specific time horizon. Akin to the framing manipulation in the monthly loan repayment judgment scenario, this could involve presenting a smokers' health status (e.g., in relation to someone who has never smoked) for the number of cigarettes they currently smoke per day (e.g., 20) versus a target amount (e.g., zero) in the present versus the future (e.g., two years from now). This would convey the relation between the effects of current versus alternative behaviours over the specified time horizon, illustrating the exponential change in health status associated with the reduction (or cessation) in cigarette smoking for the given period of time.

Observing the effects of the target behaviour (e.g., reduced smoking, or total cessation) in the future frame compared to the effects of the current behaviour (e.g., 20 cigarettes per day), is likely to anchor people on the optimal decision (i.e., to reduce or stop smoking now) based on the visualization of the extent of health benefits accrued by the target behaviour in the time frame. This may be effective in providing the motivation necessary to prompt judgment processes underpinning an effective rationale for supporting sustained behavioural adjustments. Thus, the future health status default creates an anchoring effect, increasing peoples' inclination to make adjustments towards this optimal state when the positive effects of certain choices or judgments are made salient. The ability to view increases in health or financial status over different time horizons is likely to increase the inclination to action changes sooner rather than delaying the benefits into the future. This framing therefore has the potential to help correct optimistic biases (associated with temporal discounting) which are shown widely to negatively affect peoples' health and finances.

Using dynamic graphical displays, or icon arrays in frequency formats to convey the current versus future health/financial outcomes could further facilitate the effect by helping people conceptualize the absolute different in the effectiveness of choice alternatives over different time horizons. For example, sliding an icon from left to right across a display could show absolute increases in health status compared to an icon relating to heath status for current choices and behaviours. Using concrete representations to convey the effects of choices over time will also increase data saliency and heighten the perceived impact of certain judgments and behaviours in real-world terms.

As is the case with saving for example, people are prone to underestimate the benefits of small, but consistent behaviours when dealing with non-linear growth functions and tend to believe that only large scale changes, or extreme efforts over extended time periods will lead to positive results. This erroneous propensity to linearly estimate the benefits of certain actions thus impacts motivation and creates barriers to more adaptive choices and behaviours. Observing the behavioural benefits in terms of the levels of effort or costs involved over time necessary to achieve a particular outcome could be useful in conveying the achievability of goals which is important in providing a strong underpinning rationale necessary for motivating sustained behavioural changes.

The findings relating to human performance across domains all point to the importance of appropriate goal setting and the frequent monitoring of goal behaviour which is necessary for the self-regulation underpinning behaviour change (Harkin et al., 2016). Using future event defaults in dynamic graphic displays could therefore provide an effective mechanism for improving rationality underpinning behaviour change by impacting the linear judgment processes which lead people to underestimate the positive impacts that small, consistent changes can have on future health and wealth. By working in accordance with the judgment processes underpinning the linear bias, the framing of future events in concrete terms is a potentially useful tool in assisting behaviour change, which could be more effective than interventions aimed at correcting biases or retraining cognitive processes associated with maladaptive choices and behaviours.

This approach is likely to be more effective than current visual frames used to communicate the negative physical impacts of cigarette smoking for example, which involve images of lung damage etc., disclosed on cigarette packets. It is possible that 'shock-tactics' such as this can lead to informational desensitization when people are repeatedly exposed to such stimuli. The severity of the images may also be acting to demotivate and disengage people rather than inspire (or scare) them into reducing smoking. For example, it may be assumed that the extent of internal damage incurred from smoking (as inferred by the images) is too great or potentially irreversible and thus effort to change behaviour is unwarranted. From this perspective, the time taken to repair the damage (i.e., years of smoking cessation), may be perceived as too great to warrant the effort involved in giving up, or even reducing cigarette smoking.

Although it may be highly salient, emotive imagery and wording is potentially less effective than more pragmatic data disclosures in public health communications which involve conveying the positive temporal effects that behavioural changes can have. For example, the current approach fails to address the attitude that 'just one more' cigarette will create no significant effect, or that delaying reduction or cessation will have no overall impact on health or finances in the long term. It is likely that extreme images alone do not provide enough information to activate the relevant judgment processes.

Thus, more information is necessary, but in a simplified format which delivers data appropriate for forming a rationale to motivate engagement in judgment processes which lead to self-generated conclusions and formulations of choice and behavioural intentions. Disclosing the positive effects of behaviours in future frames could therefore provide a motivational mechanism for impacting health behaviours, as opposed to using negative imagery which utilizes fear and avoidance as the motivating factors. By modelling the non-linearity of the relation between choices and future outcomes, people are able to form more accurate perceptions of the effects of their efforts. This is likely to prove highly motivational when perceived efforts are shown to yield greater benefits over shorter time horizons that previously judged.

Another context in which optimistic tendencies interact with format biases to limit judgment effectiveness is when making food choices based on product packaging and industry information. Effective judgment and choice in this area is increasingly difficult to achieve, based on food industry marketing and promotions delivering conflicting information regarding health versus desirability, combined with ambiguous product data and unsupported health claims regarding the benefits of particular products or dietary choices. As a result, people are confronted with increasing amounts of data regarding the benefits and harms of certain food choices or diets which increases the complexity of the shopping and food choice environment. Judgment processes are further impacted by marketing designed to activate behavioural factors such as optimism, locus of control and self-determination which are levered to work in opposition to optimal choice from both health and financial perspectives. Amidst an obesity epidemic and increases in the associated problems, it is important to therefore consider informational frames and formats applied in food packaging as a tool for impacting the rationality underpinning food and dietary choices and behaviours.

In the UK, the Food Standards Agency provides a ‘traffic light’ labelling system in which energy and nutrient levels (i.e., fat, saturated fat, sugar and salt) are shown in grams and colour coded as ‘high’ ‘medium’ or ‘low’ per product serving alongside each data point as a percentage of the daily reference intake for energy and each nutrient type. The purpose is to define products as either ‘good’ or ‘bad’ based on a larger proportion of ‘low’ or ‘high’ colour codes per choice. However, this system assumes higher levels of any nutrient as ‘bad’ regardless of its type, which can lead to some more optimal options being erroneously framed as ineffective choices, and vice versa. For example, a product may be low in calories yet high in salt, or high in good fat but low in saturated fat.

This data format and framing may also encourage people to assess items in isolation, when it is necessary to determine overall nutrient and energy intake on a daily basis in relation to referent intakes across all options. It is therefore difficult to compare and combine products to form an overall judgment of choice effectiveness per day, specifically when it involves synthesizing and balancing nutrient and energy values across varying portion sizes. It is also probable that people will interpret the nutrient, energy and reference intake cues as representative of the whole product, when in fact they represent a single portion. When considering the effects of marketing combined with optimistic biases in food choices, this is potentially hazardous, leading people to assume that unhealthy options are not as suboptimal as the data suggests (i.e., that it is possible to consume far more of an option per day whilst remaining within the reference intakes). Confirmation biases also dictate that people are more likely to seek or interpret information in ways which align with expectations or beliefs. Thus, when seeking a rationale for an unhealthy choice, it is possible that the current frame and format of nutrient information could actually promote biased judgments and suboptimal behaviours, rather than act to minimise them.

Based on the findings relating to the financial choice framing manipulation, one way to improve the effectiveness of peoples' food choices would be to present minimal data in absolute formats and use a single metric which normalizes choice effectiveness across options to simplify the task of interpreting and comparing data for different portions. For example, energy per serving delivers an overall indication of choice effectiveness which could be sufficient to determine optimality across nutrient types, and groups of products. By converting the percentage of energy per serving (in relation to the 2000 kcal daily reference intake) into absolute values, a simple score on a scale of 1 to 10 could be provided, where 1 equates to 10% of the daily caloric reference intake and 10 equates to 100%. (Caloric values are strongly associated with sugar and fat content, thus making energy a relatively effective indicator of general choice optimality). The advantage of framing caloric values as the key data point for decision making, is that it provides a clear referent point for facilitating people in forming fast, effective judgments based on summing the points across all choices per day.

To more effectively communicate the data relating to portion size, a graphic (as opposed to numeric) format could be used to convey the relative quantities in absolute terms. For example, product packaging may be visually segmented into individual serving sizes using calibrations or shaded squares to represent the absolute size of each serving in relation to the whole product. Energy scores could then be displayed in a large typeface within each serving segment per packet to communicate the size of the portion relative to the score, thus indicating the effectiveness of options in terms of caloric density (i.e., small sizes with high numbers being less optimal). To maintain clarity in the data environment, the data for fat, saturated fat, sugar and salt (also framed in absolutes, i.e., grams per serving), could remain secondary to the energy score, disclosed in a smaller typeface with no scores attached. This will encourage people to focus predominantly on the energy metric, increasing the likelihood of them using the data as a referent point to combine and compare scores across options.

Providing an alternative frame for the energy score, such as a physical activity metric, could create a more meaningful context for the energy cue which may increase the impact of the data on judgment processes. For example, disclosing the number of minutes of brisk walking necessary to counteract the energy intake per potion (based on the average expenditure of 44 kcal per 10 minutes), creates a highly saliency visualization of the data, conveying information in relatable terms. Given that exercise is a strong predictor of general health across the lifespan (not just weigh loss), this could be a particularly beneficial manipulation in addition to energy scores. Using a graphical icon (e.g., a green 'fast walking man') in conjunction with the absolute minutes of brisk walking per portion could also aid the conceptual connection between caloric and nutritional information and everyday actions and behaviours. Using imagery to promote relations between data and real-life experiences could therefore be effective in evoking sensory and emotional responses which may lead to better visualization of the meaning and implications of choices compared to communicating the data in less vivid, 'benign' numeric formats.

In sum, using visual representations and concrete values to frame the future effects of judgment and choice in relatable, real-world terms could overcome biases limiting rationality and behaviour in finance and health. For example, conveying food portion sizes in relation to exercise requirements, or framing health or financial advantages as the impacts of behaviour modification over time, are likely to promote more effective judgment and choice compared to standard industry informational frames and formats. Overall, the findings indicate that the key to effective frames and formats in consumer choice settings is to create mechanisms which work in accordance with peoples' tendency to form linear inferences based on the mechanisms identified throughout this thesis. By supporting human rationality based on the propensity to seek meaning via referents and apply simple additive methods, it may be possible to facilitate and improve judgment and choice to a greater extent than applying measures designed to correct or modify the biases and limitations in peoples' judgment processes.

The communication of data in formats which are compatible with inherent judgment tendencies also support people in reaching effective conclusions and choices independently, through the self-assimilation of information. Creating mechanisms for learning and self-discovery which improve the depth of informational processing are thus important for the acquisition of knowledge underpinning effective rationales for behaviour change. For example, disclosing an energy score from 1 to 10 provides an initial indicator of ‘good’ and ‘bad’ which can then be assessed within the context of the wider nutrient data and exercise requirement. In this sense, the energy data is made sufficiently accessibility to encourage further analysis, leading people to generate a stronger rationale for more optimal choice and behaviour. An educational mechanism such as this is likely to be more effective in creating long term behaviour change compared to information delivered in a purely instructional format, based on creating a deeper level of engagement which is achieved through the self-accrual of knowledge and personal experience.

7.4 Concluding Comments

The findings in this thesis show that the propensity to make concrete comparisons, perceive linear relations, and interpret data in accordance with knowledge, experience and beliefs are important determinants of human rationality in complex numeric domains. By assessing human judgment in both professional and novice/consumer populations, it was possible to test the robustness of processes underpinning biases across groups and individual differences. The tendency to form linear inferences is shown to be connected to informational format by the systematic bias to perceive values as absolutes and make sense of data by seeking linear trends and additively processing simple numeric referents which indicate the numeric continuum or bounds within which a judgment may be placed.

The direction and magnitude of linear judgments are shown to relate to context information in a given situation. The way in which context effects judgment performance depends on the data frame and format which interacts with judgment processes to either help or hinder rationality. For example, in sparse time series formats, people linearly project cues based on all observable points, and in sparse numeric formats, people combine two key referents using arithmetic operations. In the context of non-linear trends and percentages, both these judgment processes can have detrimental effects. The potential for error is particularly high in varied time series formats, in which people tend to seek meaning in noise by utilizing the additional information as a directional guide for linear trend projections. However, when highly relevant context information (in the form of a future event default) is delivered numeric formats, the tendency to additively combine referents across the informational environment can be levered to generate highly effective judgment and choice.

In contextually richer environments therefore, formatting and framing information to operate in accordance with the judgment processes underpinning how people make sense of numeric environments (i.e., the processes which give rise to the linear bias), can lead to significant improvements in rationality and choice effectiveness. The importance of context data on human rationality suggests that judgment processes involve a far wider, more complex array of factors than those disclosed in the immediate data environment. Richer data settings activate peoples' individual knowledge, experience and beliefs which are then imparted into the judgment process to guide formation of inferences and choices. How knowledge and experience etc., impacts rationality and judgment effectiveness thus depends on informational framings and formats. In both professional and consumer choice domains, the findings point towards individual knowledge and beliefs as having an impact on judgment performance which may be greater than statistical rationality, or the interpretation of cues in the immediate data environment alone.

Although human judgment processes are strongly linearly biased, there is evidence that peoples' inherent judgment abilities underpinned by individual knowledge and experience, can be significantly beneficial when applied in combination with model predictions in complex domains. This suggests there could be an evolutionary basis to human judgment processes which provides people with insights which cannot be captured effectively by statistical models in modern data environments.

From the adaptive heuristics perspective for example, the strong tendency to assume linear relations makes sense when considering cognitive algorithms developed in accordance with sampling and summing natural frequencies in a pre-mathematical environment. It is possible therefore that people are more cognitively attuned to the variances which occur throughout natural environments and to data communicated in concrete terms, than to the variances in complex statistical environments which involve information in probabilistic and non-linear formats. Our evolutionary basis may therefore account for the inability to accurately process the base rate information which is conveyed in percentage formats, thus underpinning the robustness of the propensity to seek and perceive linear functional relations, thus viewing the world in absolute terms based on concrete objects. This also accounts for why multiplicative operations are particularly challenging and unintuitive to perform. Compared to arithmetic methods, the attentional resources involved in performing sequential processing far outweigh the cognitive effort involved in adding and subtracting whole values.

Assessing methods for optimizing human judgment in combination with statistical models is thus an effective use of resources, necessary to preserve the benefits of human intuition whilst safeguarding against the pitfalls of our 'hard-wired' linear cognitive propensities. For example, focusing on developing and expanding practical experience and domain knowledge, and applying people in the development, monitoring and controlling of systems in professional settings could help isolate and utilize the advantages of human intuition and inferences. This approach would optimize peoples' ability to identify important wider, un-modelled factors, and develop a rich, tacit awareness of how those wider factors interact and connect in a real-world sense.

In consumer settings, the harmful effects of the linear bias and optimistic behavioural propensities may be counteracted by re-framing percentages in concrete terms and disclosing the temporal effects of non-linear functional relations between stimulus and response variables. This is particularly relevant in domains such as finance and health, where disclosing future outcomes as a default, employing dynamic visual representations may facilitate the comprehension of ratios and proportional differences in terms of the size of an effect relative to a particular action or event for a given time horizon. Modelling the effects of judgments or behavioural modifications in terms of absolute size differences between icons, could therefore convey the non-linearity of the relation between variables without the need for multiplicative processing or extrapolation of exponential trends. The effect of visual data frames could be further optimized by utilizing highly salient metrics to communicate the meaning of the data in terms of real-world experiences. Thus, in combination, the use of relatable, 'real-world' frames and visual formats could help reduce the barriers to peoples' intuitive judgment processes which are created by standard industry data formats in many highly important choice environments.

In sum, the thorough explication of the relation between judgment processes and data frames and formats in this thesis contributes to our understanding of human rationality in terms of factors underlying the robustness of linear judgment biases across groups and domains. Whether forming predictions in varied statistical data, or making financial or consumer choices, human rationality is dependent on inherent processes which are hard-wired to utilize data in concrete terms and form linear inferences which optimize effort and efficiency. Rather than working to correct or align judgment processes with probabilistic models in complex data environments, the focus should be on preserving the value and strengths of human intuition based on these innate cognitive processes and judgment abilities.

Devising more effective mechanisms of data disclosure which facilitate judgment processes will thus enable us to optimize choice and enhance judgment by extracting the advantages of peoples' knowledge and experience whilst minimizing the negative impacts of the linear bias in complex modern decision making domains. Despite cognitive limitations and individual differences, reducing the barriers to judgment processes will thus heighten our ability to reap the true benefits of human inferences and harness the insights and intuitions which exceed model capabilities in many real-world situations.

References

Agarwal, S., Chomsisengphet, S., Mahoney, N., & Stroebel, J. (2013). *Regulating consumer financial products: evidence from credit cards* (Working Paper No. 19484). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w19484

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), 14–14. http://doi.org/10.1167/7.13.14

Ariely, D. (2009). *Predictably irrational, revised and expanded edition: The hidden forces that shape our decisions.* Harper Collins.

Armstrong, J. S. (1980). The seer-sucker theory: The value of experts in forecasting. *Technology Review, 82*(7), 16-24. Retrieved from http://papers.ssrn.com/abstract=648763

Armstrong, J. S. (1985). Long Range Forecasting: From Crystal Ball to Computer. NY: Wiley. Retrieved from http://papers.ssrn.com/abstract=666990

Armstrong, J. S. (1991). Prediction of consumer behaviour by experts and novices. *Journal of Consumer Research, 18*(2), 251-256. Retrieved from http://papers.ssrn.com/abstract=1232569

Armstrong, J. S. (2008a). *Global warming: Forecasts by scientists versus scientific forecasts* (SSRN Scholarly Paper No. ID 1153120). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1153120

Armstrong, J. S. (2008b). *Prediction of consumer behavior by experts and novices* (SSRN Scholarly Paper No. ID 1232569). Rochester, NY: Social Science Research Network.

Armstrong, J. S., & Green, K. C. (2007). *The ombudsman: Value of expertise for forecasting decisions in conflicts* (SSRN Scholarly Paper No. ID 1657162). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1657162

Armstrong, J. S., Green, K. C., & Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, *68*(8), 1717–1731. http://doi.org/10.1016/j.jbusres.2015.03.031

Bateman, H., Dobrescu, L. I., Newell, B. R., Ortmann, A., & Thorp, S. (2016). As easy as pie: How retirement savers use prescribed investment disclosures. *Journal of Economic Behavior & Organization*, *121*, 60–76. http://doi.org/10.1016/j.jebo.2015.10.020

Bateman, I., Dent, S., Peters, E., Slovic, P., & Starmer, C. (2007). The affect heuristic and the attractiveness of simple gambles. *Journal of Behavioral Decision Making*, *20*(4), 365–380. http://doi.org/10.1002/bdm.558

Becker, O., Leitner, J., & Leopold-Wildburger, U. (2007). Heuristic modeling of expectation formation in a complex experimental information environment. *European Journal of Operational Research*, *176*(2), 975–985. http://doi.org/10.1016/j.ejor.2005.09.003

Becker, O., Leitner, J., & Leopold-Wildburger, U. (2008). Modeling expectation formation involving several sources of information. *German Economic Review*, *9*(1), 96–112. http://doi.org/10.1111/j.1468-0475.2008.00425.x

Benartzi, S., & Thaler, R. H. (2002). How much is investor autonomy worth? *The Journal of Finance*, *57*(4), 1593–1616. http://doi.org/10.1111/1540-6261.00472

Bertrand, M., & Morse, A. (2011). Information disclosure, cognitive biases, and payday borrowing. *The Journal of Finance*, *66*(6), 1865–1893. http://doi.org/10.1111/j.1540-6261.2011.01698.x

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2009). *How does simplified disclosure affect individuals' mutual fund choices?* (Working Paper No. 14859). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14859

Bhargava, S., & Manoli, D. (2013). Why are benefits left on the table? Assessing the role of information, complexity, and stigma on take-up with an IRS field experiment. *Amer. Econ. Rev*. Retrieved from https://www.cmu.edu/dietrich/sds/docs/bhargava/Paper%203%20Bhargava%20 2013.pdf

Bodie, Z., Kane, A., & Marcus, A. J. (2001). *Investments, 5th Edition* (5th edition). New York: McGraw-Hill Education.

Bramwell, R., West, H., & Salmon, P. (2006). Health professionals' and service users' interpretation of screening test results: Experimental study. *BMJ*, *333*(7562), 284. http://doi.org/10.1136/bmj.38884.663102.AE

Brase, G. L. (2002). Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making*, *15*(5), 381–401. http://doi.org/10.1002/bdm.421

Brase, G. L. (2013). Frequency interpretation of ambiguous statistical information facilitates Bayesian reasoning. *Psychonomic Bulletin & Review*, *15*(2), 284–289. http://doi.org/10.3758/PBR.15.2.284

Brehmer, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Performance*, *11*(1), 1–27. http://doi.org/10.1016/0030-5073(74)90002-6

Brehmer, B. (1978). Response consistency in probabilistic inference tasks. *Organizational Behavior and Human Performance*, *22*(1), 103–115. http://doi.org/10.1016/0030-5073(78)90008-9

Broadbent, D. E. (1958). The effects of noise on behaviour. *Perception and communication* (pp. 81–107). Elmsford, NY, US: Pergamon Press.

Busemeyer, J. R., Byun, E., Delosh, E. L., & McDaniel, M. A. (1997a). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In K. Lamberts D. R. Shanks (Ed.), *Knowledge, concepts and categories* (pp. 408–437). Cambridge, MA, US: The MIT Press.

Busemeyer, J. R., Byun, E., Delosh, E. L., & McDaniel, M. A. (1997b). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. *Knowledge, Concepts and Categories*, 408–437.

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, *102*(1), 3–21. http://doi.org/10.1016/j.obhdp.2006.09.007

Carroll, J. D. (1963). Functional learning: The learning of continuous functional mappings relating stimulus and response continua. *ETS Research Bulletin Series*, *1963*(2), i–144. http://doi.org/10.1002/j.2333-8504.1963.tb00958.x

Cesana-Arlotti, N., Téglás, E., & Bonatti, L. L. (2012). The probable and the possible at

12 months: Intuitive reasoning about the uncertain future. *Advances in Child Development and Behavior*, *43*, 1–25.

Chapman, G. B., & Liu, J. (2009). Numeracy, frequency, and Bayesian reasoning. *Judgment and Decision Making*, *4*(1), 34–40.

Chen, H., Marmorstein, H., Tsiros, M., & Rao, A. R. (2012). When more is less: The impact of base value neglect on consumer preferences for bonus packs over price discounts. *Journal of Marketing*, *76*(4), 64–77. http://doi.org/10.1509/jm.10.0443

Chen, H., & Rao, A. R. (2007). When two plus two is not equal to four: Errors in processing multiple percentage changes. *Journal of Consumer Research*, *34*(3), 327–340. http://doi.org/10.1086/518531

Choi, J. J., Laibson, D., & Madrian, B. C. (2010). Why does the law of one price fail? An experiment on index mutual funds. *Review of Financial Studies*, *23*(4), 1405–1432. http://doi.org/10.1093/rfs/hhp097

Clark, R. L., Maki, J. A., & Morrill, M. S. (2014). Can simple informational nudges increase employee participation in a 401(k) plan? *Southern Economic Journal*, *80*(3), 677–701. http://doi.org/10.4284/0038-4038-2012.199

Cleveland, W. S., & McGill, R. (1984a). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554. http://doi.org/10.1080/01621459.1984.10478080

Cleveland, W. S., & McGill, R. (1984b). The many faces of a scatterplot. *Journal of the American Statistical Association*, *79*(388), 807–822. http://doi.org/10.1080/01621459.1984.10477098

Cleveland, W. S., & McGill, R. (1986). An experiment in graphical perception. *International Journal of Man-Machine Studies*, *25*(5), 491–500. http://doi.org/10.1016/S0020-7373(86)80019-0

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*, *7*(1), 25.

Collopy, F., & Armstrong, J. S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, *38*(10), 1394–1414.

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. http://doi.org/10.1016/0010-0277(95)00664-8

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, *121*(3), 463–480. http://doi.org/10.1037/a0037010

Cousin, E. (2015, December 3). Forecast-based finance in humanitarian aid. *The Economist.* Retrieved from http://www.economist.com/news/international/21679547-executive-director-world-food-programme-explains-how-spending-money-ahead-disaster-can

Daston, L. (1988). *Classical probability in the enlightenment*. Princeton, JN: Princeton University Press.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997a). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 968.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997b). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 968–986. http://doi.org/10.1037/0278-7393.23.4.968

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, *66*(5), 819–829. http://doi.org/10.1037/0022-3514.66.5.819

Dux, P. E., Ivanoff, J., Asplund, C. L., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron*, *52*(6), 1109–1120. http://doi.org/10.1016/j.neuron.2006.11.009

Edmundson, B., Lawrence, M., & O'Connor, M. (1988). The use of non-time series information in sales forecasting: A case study. *Journal of Forecasting*, *7*(3), 201–211. http://doi.org/10.1002/for.3980070305

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making.

Organizational Behavior and Human Performance, *13*(2), 171–192. http://doi.org/10.1016/0030-5073(75)90044-6

Eisenstein, E. M., & Hoch, S. J. (2005). Intuitive compounding: Framing, temporal perspective, and expertise. *Manuscript, Temple University*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.321.2641&rep=rep1&type=pdf

Elliman, D. (2006). Pattern recognition and financial time-series. *Intelligent Systems in Accounting, Finance & Management*, *14*(3), 99–115. http://doi.org/10.1002/isaf.279

Erickson, K. I., Colcombe, S. J., Wadhwa, R., Bherer, L., Peterson, M. S., Scalf, P. E., … Kramer, A. F. (2007). Training-induced functional activation changes in dual-task processing: An fMRI study. *Cerebral Cortex*, *17*(1), 192–204. http://doi.org/10.1093/cercor/bhj137

Erta, K., Iscenko, Z., Hunt, S., & Brambley, W. (2013). Applying behavioural economics at the Financial Conduct Authority. Retrieved from http://oro.open.ac.uk/42192/1/occasional-paper-1.pdf

Fernandes, D., Lynch, J. G., & Netemeyer, R. G. (2014). Financial literacy, financial education, and downstream financial behaviors. *Management Science*, *60*(8), 1861–1883. http://doi.org/10.1287/mnsc.2013.1849

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, *50*(2), 123–129. http://doi.org/10.1007/BF00309212

Fiedler, K. (2011). *Social communication*. Psychology Press.

Fiedler, K., Kareev, Y., Avrahami, J., Beier, S., Kutzner, F., & Hütter, M. (2015). Anomalies in the detection of change: When changes in sample size are mistaken for changes in proportions. *Memory & Cognition*, *44*(1), 143–161. http://doi.org/10.3758/s13421-015-0537-z

Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cognitive Processing*, *13*(1), 161–164. http://doi.org/10.1007/s10339-012-0477-5

Fisch, J. E., & Wilkinson-Ryan, T. (2014). *Why do retail investors make costly* *mistakes? An experiment on mutual fund choice* (SSRN Scholarly Paper No. ID 2086766). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2086766

Galesic M, & Garcia-Retamero R. (2010). Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*, *170*(5), 462–468. http://doi.org/10.1001/archinternmed.2009.481

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, *28*(2), 210–216. http://doi.org/10.1037/a0014474

Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science*, *22*(5), 392–399. http://doi.org/10.1177/0963721413491570

Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: A cross-cultural comparison. *American Journal of Public Health*, *99*(12), 2196–2202. http://doi.org/10.2105/AJPH.2009.160234

Garcia-Retamero, R., & Galesic, M. (2010). How to reduce the effect of framing on messages about health. *Journal of General Internal Medicine*, *25*(12), 1323–1329. http://doi.org/10.1007/s11606-010-1484-9

Garcia-Retamero, R., Wicki, B., Cokely, E. T., & Hanson, B. (2014). Factors predicting surgeons' preferred and actual roles in interactions with their patients. *Health Psychology*, *33*(8), 920–928. http://doi.org/10.1037/hea0000061

Gardner, E. S., & McKenzie, E. (1985). Forecasting trends in time series. *Management Science*, *31*(10), 1237–1246.

Gardner, P. H., McMillan, B., Raynor, D. K., Woolf, E., & Knapp, P. (2011). The effect of numeracy on the comprehension of information about medicines in users of a patient information website. *Patient Education and Counseling*, *83*(3), 398–403. http://doi.org/10.1016/j.pec.2011.05.006

Gattis, M. (2003). *Spatial schemas and abstract thought*. MIT Press.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, *4*(1), 1–58.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond ‘heuristics and biases’. *European Review of Social Psychology*, *2*(1), 83–115. http://doi.org/10.1080/14792779143000033

Gigerenzer, G. (1994). *Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa)* [Article]. Retrieved 27 May 2016, from https://www.mpib-berlin.mpg.de/volltexte/institut/dok/full/gg/ggwtdsp__/ggwtdsp__.html

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592–596. http://doi.org/10.1037/0033-295X.103.3.592

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you* (1st Edition). New York: Simon & Schuster.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science, 3*(1), 20–29. http://doi.org/10.1111/j.1745-6916.2008.00058.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482. http://doi.org/10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. http://doi.org/10.1037/0033-295X.102.4.684

Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review*, *106*(2), 425–430. http://doi.org/10.1037/0033-295X.106.2.425

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. *Simple heuristics that make us smart* (pp. 3–34). New York, NY, US: Oxford University Press.

Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making*, *12*(1), 37–53. http://doi.org/10.1002/(SICI)1099-0771(199903)12:1<37::AID-BDM319>3.0.CO;2-8

Green, E. (2014, July 23). Why do Americans stink at math? *The New York Times*. Retrieved from http://www.nytimes.com/2014/07/27/magazine/why-do-americans-stink-at-math.html

Green, K. C., & Armstrong, J. S. (2007). Global warming: Forecasts by scientists versus scientific forecasts. *Energy & Environment*, *18*(7), 997–1021. http://doi.org/10.1260/095830507782616887

Harkin, B., Webb, T. L., Chang, B. P., Prestwich, A., Conner, M., Kellar, I., Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *142*(2), 198.

Harvey, N. (1995). Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, *63*(3), 247–263. http://doi.org/10.1006/obhd.1995.1077

Harvey, N., Bolger, F., & McClelland, A. (1994). On the nature of expectations. *British Journal of Psychology*, *85*(2), 203–229. http://doi.org/10.1111/j.2044-8295.1994.tb02519.x

Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 589–607. http://doi.org/10.1037/a0029179

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*(12), 1372–1388. http://doi.org/10.1037/0003-066X.39.12.1372

Hastings, J. S., & Tejeda-Ashton, L. (2008). *Financial literacy, information, and demand elasticity: Survey and experimental evidence from Mexico* (Working Paper No. 14538). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14538

Hausman, J. A. (2012). *Contingent valuation: A critical assessment*. Elsevier.

Haws, K. L., Bearden, W. O., & Nenkov, G. Y. (2011). Consumer spending self-control effectiveness and outcome elaboration prompts. *Journal of the Academy of Marketing Science*, *40*(5), 695–710. http://doi.org/10.1007/s11747-011-0249-2

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, *138*(2), 211–237. http://doi.org/10.1037/a0025940

Hoch, S. J., & Schkade, D. A. (1996). A psychological approach to decision support systems. *Management Science*, *42*(1), 51–64.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*(5), 538–540. http://doi.org/10.1097/00001888-199805000-00024

Hoffrage, U., Krauss, S., Martignon, L., & Gigerenzer, G. (2015). Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Frontiers in Psychology*, *6*. http://doi.org/10.3389/fpsyg.2015.01473

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*(5500), 2261–2262. http://edoc.unibas.ch/dok/A5265669

Hohle, S. M., & Teigen, K. H. (2015). Forecasting forecasts: the trend effect. *Judgment and Decision Making*, *10*(5), 416–429.

Hsee, C. K. (2006). *Less is better: when low-value options are valued more highly than high-value options* (SSRN Scholarly Paper No. ID 930083). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=930083

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*(5), 576–590. http://doi.org/10.1037/0033-2909.125.5.576

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science*, *5*(4), 343–355. http://doi.org/10.1177/1745691610374586

Hung, A., Heinberg, A., & Yoong, J. (2010). *Do risk disclosures affect investment choice?* (SSRN Scholarly Paper No. ID 1688038). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1688038

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. http://doi.org/10.1016/j.ijforecast.2006.03.001

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*(3), 171–216. http://doi.org/10.1006/cogp.2001.0755

Iyengar, S. S., Huberman, G., & Jiang, W. (2004). How much choice is too much? Contributions to 401 (k) retirement plans. *Pension Design and Structure: New Lessons from Behavioral Finance*, 83–95.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995–1006. http://doi.org/10.1037/0022-3514.79.6.995

Johnson, E. D., & Tubau, E. (2015). Comprehension and computation in Bayesian problem solving. *Frontiers in Psychology*, *6*. http://doi.org/10.3389/fpsyg.2015.00938

Johnson, E. J., & Goldstein, D. G. (2003). *Do defaults save lives?* (SSRN Scholarly Paper No. ID 1324774). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1324774

Joireman, J., Kees, J., & Sprott, D. (2010). Concern with immediate consequences magnifies the impact of compulsive buying tendencies on college students' credit card debt. *Journal of Consumer Affairs*, *44*(1), 155–178. http://doi.org/10.1111/j.1745-6606.2010.01161.x

Joireman, J., Sprott, D. E., & Spangenberg, E. R. (2005). Fiscal responsibility and the consideration of future consequences. *Personality and Individual Differences*, *39*(6), 1159–1168. http://doi.org/10.1016/j.paid.2005.05.002

Juslin, P. (2015). Controlled information integration and Bayesian inference. *Frontiers in Psychology*, *6*. http://doi.org/10.3389/fpsyg.2015.00070

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*(1), 259–298. http://doi.org/10.1016/j.cognition.2007.02.003

Juslin, P., Lindskog, M., & Mayerhofer, B. (2015). Is there something special with probabilities? – Insight vs. computational ability in multiple risk combination. *Cognition*, *136*, 282–303. http://doi.org/10.1016/j.cognition.2014.11.041

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. http://doi.org/10.1037/h0034747

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. http://doi.org/10.2307/1914185

Kalish, M. L. (2013). Learning and extrapolating a periodic function. *Memory & Cognition*, *41*(6), 886–896. http://doi.org/10.3758/s13421-013-0306-9

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: Knowledge partitioning and function learning. *Psychological Review*, *111*(4), 1072–1099. http://doi.org/10.1037/0033-295X.111.4.1072

Kasa, K. (1992). Common stochastic trends in international stock markets. *Journal of Monetary Economics*, *29*(1), 95–124. http://doi.org/10.1016/0304-3932(92)90025-W

Keller, C., & Siegrist, M. (2009). Effect of risk communication formats on risk perception depending on numeracy. *Medical Decision Making*. http://doi.org/10.1177/0272989X09333122

Kheifets, A., & Gallistel, C. R. (2012). Mice take calculated risks. *Proceedings of the National Academy of Sciences*, *109*(22), 8776–8779. http://doi.org/10.1073/pnas.1205131109

Koehler, D. J., & Harvey, N. (2008). *Blackwell handbook of judgment and decision making*. John Wiley & Sons.

Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado Law Review*, *67*, 859.

Koehler, J. J., & Mercer, M. (2009). Selection neglect in mutual fund advertisements. *Management Science*, *55*(7), 1107–1121.

Kogut, T., & Ritov, I. (2005). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes*, *97*(2), 106–116. http://doi.org/10.1016/j.obhdp.2005.02.003

Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory,* *and Cognition*, *17*(5), 811–836. http://doi.org/10.1037/0278-7393.17.5.811

Krauss, S., Martignon, L., Hoffrage, U, & Gigerenzer, G. (2002). *Bayesian reasoning and natural frequencies: A generalization to complex situations.* Manuscript Submitted for Publication.

Kruger, J., & Vargas, P. (2008). Consumer confusion of percent differences. *Journal of Consumer Psychology*, *18*(1), 49–61. http://doi.org/10.1016/j.jcps.2007.10.009

Kurzenhäuser, S., & Hoffrage, U. (2002). Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher*, *24*(5), 516–521. http://doi.org/10.1080/01421590210000012540

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*(2), 443–477.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* University of Chicago Press.

Larrick, R. P., & Soll, J. B. (2008). The MPG illusion. *Science*, *320*(5883), 1593–1594. http://doi.org/10.1126/science.1154983

Lavie, N., & Robertson, I. H. (2001). The role of perceptual load in neglect: Rejection of ipsilesional distractors is facilitated with higher central load. *Journal of Cognitive Neuroscience*, *13*(7), 867–876. http://doi.org/10.1162/089892901753165791

Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, *22*(3), 493–518. http://doi.org/10.1016/j.ijforecast.2006.03.007

Lawrence, M., & O'Connor, M. (1993). Scale, variability, and the calibration of judgmental prediction intervals. *Organizational Behavior and Human Decision Processes*, *56*(3), 441–458. http://doi.org/10.1006/obhd.1993.1063

Lawrence, M., O'Connor, M., & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research*, *122*(1), 151–160. http://doi.org/10.1016/S0377-2217(99)00085-5

Lee, J., & Frank, J. (1992). How quantity judgment changes as the number of cues increases: An analytical framework and review. *Psychological Bulletin*, *112*(2), 363–377. http://doi.org/10.1037/0033-2909.112.2.363

Lee, J., & Hogarth, J. M. (1999). The price of money: Consumers' understanding of APRs and contract interest rates. *Journal of Public Policy & Marketing*, *18*(1), 66–76.

Lim, J. S., & O'Connor, M. (1996). Judgmental forecasting with time series and causal information. *International Journal of Forecasting*, *12*(1), 139–153. http://doi.org/10.1016/0169-2070(95)00635-4

Lindsey, S., Hertwig, R., & Gigerenzer, G. (2003). Communicating statistical DNA evidence. *Jurimetrics*, *43*(2), 147–163.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44. http://doi.org/10.1177/0272989X0102100105

Lipshitz, R. (1993). Converging themes in the study of decision making in realistic settings. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 103–137). Westport, CT, US: Ablex Publishing.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*(5), 331–352. http://doi.org/10.1002/bdm.381

List, J. A. (2002). Preference reversals of a different kind: The 'more is less' phenomenon. *The American Economic Review*, *92*(5), 1636–1643.

Loewenstein, G., Asch, D. A., Friedman, J. Y., Melichar, L. A., & Volpp, K. G. (2012). Can behavioural economics make us healthier? *BMJ (Clinical Research Ed.)*, *344*, e3482.

Loewenstein, G., Read, D., & Baumeister, R. F. (2003). *Time and decision: Economic and psychological perspectives of intertemporal choice*. Russell Sage Foundation.

Loewenstein, G., Sunstein, C. R., & Golman, R. (2014). Disclosure: Psychology changes everything. *Annual Review of Economics*, *6*(1), 391–419. http://doi.org/10.1146/annurev-economics-080213-041341

Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 392–400. http://doi.org/10.1037/0278-7393.13.3.392

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy around the world: An overview. *Journal of Pension Economics and Finance*, *10*(04), 497–508. http://doi.org/10.1017/S1474747211000448

Lynch, P. (2006). *The emergence of numerical weather prediction: Richardson's dream*. Cambridge University Press.

Lynch, P. (2008). The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, *227*(7), 3431–3444. http://doi.org/10.1016/j.jcp.2007.02.034

Makridakis, S., & Taleb, N. (2009). Living in a world of low levels of predictability. *International Journal of Forecasting*, *25*(4), 840–844. http://doi.org/10.1016/j.ijforecast.2009.05.008

Mcdaniel, M. A., & Busemeyer, J. R. (2005). The conceptual basis of function learning and extrapolation: Comparison of rule-based and associative-based models. *Psychonomic Bulletin & Review*, *12*(1), 24–42.

McDaniel, M. A., Dimperio, E., Griego, J. A., & Busemeyer, J. R. (2009). Predicting transfer performance: A comparison of competing function learning models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 173–195. http://doi.org/10.1037/a0013982

Mckenzie, C. R., & Liersch, M. J. (2011). Misunderstanding savings growth: Implications for retirement savings behavior. *Journal of Marketing Research*, *48*(SPL), S1–S13. http://doi.org/10.1509/jmkr.48.SPL.S1

McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, *17*(5), 414–420. http://doi.org/10.1111/j.1467-9280.2006.01721.x

Mecklinger, A., Weber, K., Gunter, T. ., & Engle, R. . (2003). Dissociable brain mechanisms for inhibitory control: Effects of interference content and working memory capacity. *Cognitive Brain Research*, *18*(1), 26–38. http://doi.org/10.1016/j.cogbrainres.2003.08.008

Mercer, M., Palmiter, A. R., & Taha, A. E. (2010). Worthless warnings? Testing the effectiveness of disclaimers in mutual fund advertisements. *Journal of Empirical*

Legal Studies, *7*(3), 429–459. http://doi.org/10.1111/j.1740-1461.2010.01184.x

Mitchell, D. J., & Cusack, R. (2008). Flexible, capacity-limited activity of posterior parietal cortex in perceptual as well as visual short-term memory tasks. *Cerebral Cortex*, *18*(8), 1788–1798. http://doi.org/10.1093/cercor/bhm205

Mullainathan, S., Noeth, M., & Schoar, A. (2012). *The market for financial advice: An audit study* (Working Paper No. 17929). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17929

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *26*(9), 1142–1150. http://doi.org/10.1177/01461672002611010

Navarro-Martinez, D., Salisbury, L. C., Lemon, K. N., Stewart, N., Matthews, W. J., & Harris, A. J. (2011). Minimum required payment and supplemental information disclosure effects on consumer debt repayment decisions. *Journal of Marketing Research*, *48*(SPL), S60–S77. http://doi.org/10.1509/jmkr.48.SPL.S60

Newall, P. W. S. (2016a). Downside financial risk is misunderstood. *In Preparation*.

Newall, P. W. S. (2016b). Psychologically-informed investment disclosure. *In Preparation*.

Newall, P. W. S., & Love, B. C. (2015). Nudging investors big and small toward better decisions. *Decision*, *2*(4), 319–326. http://doi.org/10.1037/dec0000036

Newell, R. G., & Siikamäki, J. V. (2013). *Nudging energy efficiency behavior: The role of information labels* (Working Paper No. 19224). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w19224

Newman, G. E., & Scholl, B. J. (2012). Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review*, *19*(4), 601–607. http://doi.org/10.3758/s13423-012-0247-5

Harvey, N., Ewart, T., & West, R. (1997). Effects of data noise on statistical judgement. *Thinking & Reasoning*, *3*(2), 111–132. http://doi.org/10.1080/135467897394383

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012). Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. *Journal of Behavioral Decision Making*, *25*(4), 390–401.

http://doi.org/10.1002/bdm.751

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2015). Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied*, *21*(2), 178–194. http://doi.org/10.1037/xap0000045

Önkal, D., & Muradoğlu, G. (1994). Financial modelling evaluating probabilistic forecasts of stock prices in a developing stock market. *European Journal of Operational Research*, *74*(2), 350–358. http://doi.org/10.1016/0377-2217(94)90102-3

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin*, *135*(2), 262–285. http://doi.org/10.1037/a0014821

Oxlade, A. (2016, April 15). Latest interest rates predictions: First rise in 'December 2019'. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/personal-banking/savings/latest-interest-rates-predictions-first-rise-in-august-2019/

Love, B. C., & Parker, K. N. (2016). Formats for success: Numerical frames with powerful effects in professional retail forecasting. *In Preparation*.

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, *21*(1), 31–35. http://doi.org/10.1177/0963721411429960

Petrova, D., Garcia-Retamero, R., & Cokely, E. T. (2015). Understanding the harms and benefits of cancer screening: A model of factors that shape informed decision making. *Medical Decision Making*, *35*(7), 847–858. http://doi.org/10.1177/0272989X15587676

Puri, M., & Robinson, D. T. (2007). Optimism and economic choice. *Journal of Financial Economics*, *86*(1), 71–99. http://doi.org/10.1016/j.jfineco.2006.09.003

Remus, W., O'Connor, M., & Griggs, K. (1995). Does reliable information improve the accuracy of judgmental forecasts? *International Journal of Forecasting*, *11*(2), 285–293. http://doi.org/10.1016/0169-2070(94)00578-Z

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*(1), 89–107. http://doi.org/10.1016/j.lindif.2007.03.011

Roller, C. (2011). The impact number formats have on judgment and decision making. Retrieved from http://www.uxmatters.com/mt/archives/2011/06/the-impact-number-formats-have-on-judgment-and-decision-making.php

Salisbury, L. C. (2013). Minimum payment warnings and information disclosure effects on consumer debt repayment decisions. *Journal of Public Policy & Marketing*, *33*(1), 49–64. http://doi.org/10.1509/jppm.11.116

Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustment of statistical forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 405–416). Boston, MA: Springer US. Retrieved from http://dx.doi.org/10.1007/978-0-306-47630-3_18

Scalf, P. E., Banich, M. T., Kramer, A. F., Narechania, K., & Simon, C. D. (2007). Double take: Parallel processing by the cerebral hemispheres reduces attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 298–329. http://doi.org/10.1037/0096-1523.33.2.298

Scalf, P. E., Dux, P. E., & Marois, R. (2011). Working memory encoding delays top–down attention to visual cortex. *Journal of Cognitive Neuroscience*, *23*(9), 2593–2604. http://doi.org/10.1162/jocn.2011.21621

Schapira, M. M., Nattinger, A. B., & McAuliffe, T. L. (2006). The influence of graphic format on breast cancer risk communication. *Journal of Health Communication*, *11*(6), 569–582. http://doi.org/10.1080/10810730600829916

Schapira, M. M., Nattinger, A. B., & McHorney, C. A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making*, *21*(6), 459–467. http://doi.org/10.1177/0272989X0102100604

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, *4*(3), 219–247. http://doi.org/10.1037/0278-6133.4.3.219

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078. http://doi.org/10.1037/0022-3514.67.6.1063

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*(11), 966–972. http://doi.org/10.7326/0003-4819-127-11-199712010-00003

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, *130*(3), 380–400. http://doi.org/10.1037/0096-3445.130.3.380

Seifert, M., & Hadida, A. L. (2013). On the relative importance of linear model and human judge(s) in combined forecasting. *Organizational Behavior and Human Decision Processes*, *120*(1), 24–36. http://doi.org/10.1016/j.obhdp.2012.08.003

Seifert, M., Siemsen, E., Hadida, A. L., & Eisingerich, A. B. (2015). Effective judgmental forecasting in the context of fashion products. *Journal of Operations Management*, *36*, 33–45. http://doi.org/10.1016/j.jom.2015.02.001

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, *53*(1), 491–517. http://doi.org/10.1146/annurev.psych.53.100901.135213

Should I be taking statins? (2014, November 26). *National Institute for Health and Care Excellence.* Retrieved from https://www.nice.org.uk/news/blog/should-i-be-taking-statins

Silver, N. (2012). *The signal and the noise: The art and science of prediction*. Penguin UK.

Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, *82*(398), 454–465. http://doi.org/10.1080/01621459.1987.10478448

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. http://doi.org/10.2307/1884852

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, *91*(2), 296–309. http://doi.org/10.1016/S0749-5978(03)00021-9

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *The Journal of Socio-Economics*, *31*(4), 329–342. http://doi.org/10.1016/S1053-

5357(02)00174-9

Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, *24*(3), 271–296. http://doi.org/10.1023/A:1005595519944

Soll, J. B., Keeney, R. L., & Larrick, R. P. (2013). Consumer misunderstanding of credit card use, payments, and debt: Causes and solutions. *Journal of Public Policy & Marketing*, *32*(1), 66–81. http://doi.org/10.1509/jppm.11.061

Stango, V., & Zinman, J. (2009). Exponential growth bias and household finance. *The Journal of Finance*, *64*(6), 2807–2849.

Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics*, *1*(12), 439–446. http://doi.org/10.3758/BF03215821

Stewart, N. (2009). The cost of anchoring on credit-card minimum repayments. *Psychological Science*, *20*(1), 39–41. http://doi.org/10.1111/j.1467-9280.2008.02255.x

Strathman, A., Gleicher, F., Boninger, D. S., & Scott, C. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology*, *66*(4), 742–752. http://doi.org/10.1037/0022-3514.66.4.742

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, *23*(3), 165–180. http://doi.org/10.2307/2295722

Tetlock, P. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*(3), 199–214.

Thaler, Richard H. and Cass R. Sunstein. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Thomson, M., Pollock, A., Henriksen, K., & Macaulay, A. (2004). The influence of the forecast horizon on judgemental probability forecasts of exchange rate movements. *The European Journal of Finance*, *10*(4), 290–307.

http://doi.org/10.1080/13518470110047620

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751–754. http://doi.org/10.1038/nature02466

Tsoularis, A., & Wallace, J. (2002). Analysis of logistic growth models. *Mathematical Biosciences*, *179*(1), 21–55. http://doi.org/10.1016/S0025-5564(02)00096-2

Tubau, E. (2008). Enhancing probabilistic reasoning: The role of causal graphs, statistical format and numerical skills. *Learning and Individual Differences*, *18*(2), 187–196. http://doi.org/10.1016/j.lindif.2007.08.006

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. http://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. http://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293–315. http://doi.org/10.1037/0033-295X.90.4.293

Tversky, A., & Koehler, D. J. (1994). Support theory: A non-extensional representation of subjective probability. *Psychological Review*, *101*(4), 547–567. http://doi.org/10.1037/0033-295X.101.4.547

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*(1), 124–129. http://doi.org/10.1016/j.cognition.2008.10.008

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, *23*(4), 515–557. http://doi.org/10.1016/0010-0285(91)90005-9

Vitaliano, P. P., Russo, J., Carr, J. E., Maiuro, R. D., & Becker, J. (1985). The ways of coping checklist: revision and psychometric properties. *Multivariate Behavioral Research*, *20*(1), 3–26. http://doi.org/10.1207/s15327906mbr2001_1

Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review*, *92*(4), 574–584. http://doi.org/10.1037/0033-295X.92.4.574

Wearden, G. (2015, January 5). Millions fear missing January's rent or mortgage payments, says Shelter. *The Guardian*. Retrieved from http://www.theguardian.com/society/2015/jan/05/missing-rent-mortgage-payments-shelter-interest-rate-housing-costs

Weber, B. J., & Chapman, G. B. (2005). Playing for peanuts: Why is risk seeking more common for low-stakes gambles? *Organizational Behavior and Human Decision Processes*, *97*(1), 31–46. http://doi.org/10.1016/j.obhdp.2005.03.001

Wegwarth, O., Schwartz, L. M., Woloshin, S., Gaissmaier, W., & Gigerenzer, G. (2012). Do physicians understand cancer screening statistics? A national survey of primary care physicians in the united states. *Annals of Internal Medicine*, *156*(5), 340–349. http://doi.org/10.7326/0003-4819-156-5-201203060-00005

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, *39*(5), 806–820. http://doi.org/10.1037/0022-3514.39.5.806

Weinstein, N. D., & Klein, W. M. (1996). Unrealistic optimism: Present and future. *Journal of Social and Clinical Psychology*, *15*(1), 1–8. http://doi.org/10.1521/jscp.1996.15.1.1

Wilkie-Thomson, M. E., Önkal-Atay, D., & Pollock, A. C. (1997). Currency forecasting: An investigation of extrapolative judgement. *International Journal of Forecasting*, *13*(4), 509–526. http://doi.org/10.1016/S0169-2070(97)00036-8

Witteman, H. O., Fuhrel-Forbis, A., Wijeysundera, H. C., Exe, N., Dickson, M., Holtzman, L., Kahn, V. C., Zikmund-Fisher, B. J. (2014). Animated randomness, avatars, movement, and personalization in risk graphics. *Journal of Medical Internet Research*, *16*(3), e80. http://doi.org/10.2196/jmir.2895

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, *13*(4), 167–174. http://doi.org/10.1016/j.tics.2009.01.008

Yamagishi, K. (1997). When a 12.86% mortality is more dangerous than 24.14%: Implications for risk communication. *Appl. Cognit. Psychol*, *11*(6), 495–506. http://doi.org/10.1002/(SICI)1099-0720(199712)11:63.0.CO;2-J

Yang, S., Markoczy, L., & Qi, M. (2007). Unrealistic optimism in consumer credit card adoption. *Journal of Economic Psychology*, *28*(2), 170–185. http://doi.org/10.1016/j.joep.2006.05.006

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*(3), 287–308. http://doi.org/10.1016/j.cognition.2004.12.003

Zikmund-Fisher, B. J., Ubel, P. A., Smith, D. M., Derry, H. A., McClure, J. B., Stark, A., Pitsch, R. K., Fagerlin, A. (2008). Communicating side effect risks in a tamoxifen prophylaxis decision aid: The debiasing influence of pictographs. *Patient Education and Counseling*, *73*(2), 209–214. http://doi.org/10.1016/j.pec.2008.05.010

Appendices

Appendix 1

Chapter 3, Experiment 1. Method of conversion of the onscreen stimuli and the participants' responses in the percentage condition into absolute values.

Percentage condition				
	Percentage numbers (observed)	Base rate changes in absolute values (hidden)	Conversions of percentages seen onscreen into absolute values of base rates	Base rate changes calculated in absolute vales
Linear increase	101.56 % (101.5625)	200	101.5625*200/100 = 203.125	203.125 + 200 = 403.125
	50.39 % (50.387597)	403.125	50.387597*403.125/100 = 203.125	203.125+403.125 = 606.25
	33.51 % (33.505155)	606.25	33.505155*606.25/100 = 203.125	203.125+606.25 = 809.375
4^{th} point prediction	25.01 % (un-rounded = 25.096525)	809.375	25.096525*809.375/100 = 203.125	203.125+809.375 = 1012.5
Exponential increase	50 % (1.5)	200	50*200/100 = 100	100+200 = 300
	50 % (1.5)	300	50*300/100 = 150	150+300 = 450
	50 % (1.5)	450	50*450/100 = 225	225+450 = 675
4^{th} point prediction	50 % (1.5)	675	50*675/100 = 337.5	337.5+675 = 1012.5
Linear decrease	-20.06 % (-20.061728)	1012.5	-20.061728*1012.5/100 = -203.124996	-203.125+1012.5 = 809.375
	-25.01 % (-25.096525)	809.375	-25.096525*809.375/100 = -203.1249992	-203.125+809.375 = 606.25
	-33.51 % (-33.505155)	606.25	-33.505155*606.25/100 = -203.125	-203.125+606.25 = 403.125

4th point prediction	-50.39 % (un-rounded = -50.387597)	403.125	-50.387597*403.125/100 = -203.125	-203.125+403.125 = 200
Exponential decrease	-33.3 % (0.666667)	1012.5	-33.3*1012.5/100 = -337.5	-337.5+1012.5 = 675
	-33.3 % (0.666667)	675	-33.3*675/100 = -225	-225+675 = 450
	-33.3 % (0.666667)	450	-33.3*450/100 = -150	-150+450 = 300
4th point prediction	-33.3 % (0.666667)	300	-33.3*300/100 = -100	-100+300 = 200

Appendix 2

Chapter 3, Experiment 1. Generation of the formal model predictions for the human-model fitting procedure.

The 2-Regression Model Construction

The 2-regression model was constructed by applying linear regression to the second and third data points observed in the absolute condition exponential increase and exponential decrease sequences and the percentage condition linear increase and linear decrease sequences. The fourth point predictions were then generated from the regression of these second and third points. Shown below are the numbers that were entered into the 2-regression model and the regression equation and fourth point predictions for the absolute and percentage condition. (x is the presentation order of the numbers in each trial, i.e., the last two observations are numbered x = 1 and 2 which represents the first and second numbers presented in each trial).

Absolute condition:

Exponential increase (1, 150) (2, 225); $y = 75x + 75$; model prediction = 300.00.

Exponential decrease (1, -225) (2, -150); $y = 75x – 300$; model prediction = -75.00.

Percentage condition:

Linear increase (1, 50.387597) (2, 33.505155); $y = -16.882x + 67.27$; model prediction = 16.622713000; conversion into absolute value = 134.54008379752700.

Linear decrease (1, -25.096525) (2, -33.505155); y = -8.4086x - 16.688; model prediction = -41.91378500; conversion into absolute value = -168.96494591716300.

The 3-Regression Model Construction

The 3-regression model was constructed in the same way as the 2-regression model, except all three observed values were used to generate the predictions. The numbers that were entered into the 3-regression model and the regression equations and fourth point predictions are given below for the absolute and percentage condition (x = 1, 2 and 3 which represents the first, second and third numbers presented in each trial).

Absolute condition:

Exponential increase (1, 100) (2, 150) (3, 225); y = 62.5x + 33.333;

model prediction = 283.33333333333300.

Exponential decrease (1, -337.5) (2, -225) (3, -150); y = 93.75x - 425; model prediction = -50.00.

The percentage values given in the following linear regression model parameters are the un-rounded percentage values. In the trials, the values shown onscreen were rounded to two decimal places and labelled with a % sign and a plus/minus to signify increases and decreases. (E.g., +101.56%, +50.39%, +33.51%).

Percentage condition:

Linear increase (1, 101.5625) (2, 50.387597) (3, 33.505155); y = -34.029x + 129.88;

model prediction = -6.23892766666668000; conversion into absolute value =

-50.4963209723977000.

Linear decrease (1, -20.061728) (2, -25.096525) (3, -33.505155); y = -6.7217x - 12.778;

model prediction = -39.6645630000; conversion into absolute value = -159.8977697223700000.

The 3-Mean Model Construction

For both the absolute and percentage conditions, the fourth point predictions for the 3-mean model were calculated as the arithmetic mean of all three observed values. For the arithmetic mean calculations in the percentage condition, the outputs were converted in absolute values.

3-mean model predictions				
	Linear increase	Linear decrease	Exponential increase	Exponential decrease
Absolute condition	+203.125	-203.125	+158.33	-237.5
Percentage condition	+61.82	-26.2	+50	-33.3

Appendix 3

Chapter 3, Experiment 1. Conversions of 3-mean model predictions from percentages into absolute values.

Conversion of 3-mean model predictions into absolute values			
Percentage condition	Arithmetic mean calculation	Design function fourth point base rate	Conversion into absolute value
Linear increase	101.56 + 50.39 + 33.51/3 = 61.82	809.375	809.375 * 61.82/100 = 500.355625
Exponential increase	50 + 50 + 50/3 = 50	675	675 * 50/100 = 337.5
Linear decrease	-20.06 + -25.01 + -33.51/3 = -26.19333333	403.125	403.125 * -26.19333333/100 = -105.591875
Exponential decrease	-33.3 + -33.3 + -33.3/3 = -33.3	300	300 * -33.3/100 = -99.9