Three essays on behavioural finance

Hemmens, Christopher

Three essays on behavioural finance

by

Christopher Hemmens

A thesis submitted to the
Geneva School of Economics and Management,
University of Geneva, Switzerland,
in fulfillment of the requirements for the degree of
PhD in Finance

Members of the thesis committee:
Prof. Rajna Gibson Brandon, Supervisor, University of Geneva
Prof. Olivier Scaillet, Chair, University of Geneva
Prof. Kerstin Preuschoff, University of Geneva
Prof. Peter Bossaerts, University of Melbourne

Thesis No. 39
January 2017

Acknowledgements

None of this would have been possible without my hard-working advisor Rajna Gibson Brandon who always made sure that I was making progress and who was both motivating and critical when she needed to be. Thank you very much and please accept my sincerest gratitude. Thank you to Alexis who has supported me so thoroughly in this crucial final stretch and to my friends and family who have stood by me and shown such interest in my success.

To the professors and co-authors who gave me their advice and encouragement starting with my thesis committee: Peter Bossaerts, Kertsin Preuschoff, and Olivier Scaillet; then Tony Berrada, Ines Chaieb, Harald Hau, Philipp Krueger, Erwan Morellec, and Mathieu Trépanier, a big thank you. I would also like to thank the Geneva Finance Research Institute for its unerring support and Christina Gremaud and Michael Steinhauser for their administrative help and guidance. Thank you to Matthias Efing who was an extremely good influence on me and one of the best office-mates a PhD student could ask for. I would also like to mention just a couple of people I've had the pleasure of working with: Elisabeth, James, Paola, Alper, Stefano, Sébastien, Adrien, Gabriela, Philip, Anita, Cyril, Cliona, Marilyn, Ridima, Ariana, Rémy, Emmanuel, Claudio, Dario, Max, Denis, Julien, Jan, Cornelius, and many more.

It remains for me to acknowledge the comments made for my individual research papers.

Chapter 1: My co-authors and I would like to thank Peter Bossaerts, Michal Czerwonko, Kerstin Preuschoff, Olivier Scaillet, and Fabio Trojani; Ilaria Piatti from the Università della Svizzera italiana for her discussion of their paper at the SFI Research Days conference in Gerzensee in June 2014; Petko Kalev, Talis J. Putnins, Kym Thorne, Vikash Ramiah, Robert B. Durand, as well as other participants who attended the 5th Behavioral Finance and Capital Markets Conference in Adelaide, Australia on the 8th and 9th September 2015.

Chapter 2: I would like to thank Rajna Gibson Brandon and Olivier Scaillet of the University of Geneva for guidance on data analysis; Manuel Grieder, Christian Zehnder, John Antonakis, and the students of the University of Lausanne for helping with the main experiment; the Swiss Finance Institute and the Geneva Finance Research Institute for financing and support; Matthias Efing, Elisabeth Proehl, Cyril Pasche, Christina Gremaud, Sebastien Coupy, Zhicheng Zhang, and Paola Pederzoli for help with translations and quality control; and Didier Grandjean, Andy Christen, the psychology department at the University of Geneva, the participants of the 4th International Conference on Music and Emotion held in Geneva in October 2015, and Peter Bossaerts and Kerstin Preuschoff.

Chapter 3: I would like to thank the following people for their support and advice: Tony Berrada, Peter Bossaerts, Ines Chaieb, Jerome Detemple, Rajna Gibson Brandon, Kerstin Preuschoff, and Olivier Scaillet as well as the Swiss Finance Institute and Geneva Finance Research Institute.

Abstract

The fact that human economic behaviour has a significant irrational element - one that is simultaneously hard-to-explain and highly predictable - has fascinated economists for decades from Fechner, 1860 to Shiller, 2005 and beyond. In this dissertation, I investigate the field from various perspectives: chapter 1 examines the impact that language describing irrational behaviour in the media has on stock markets; chapter 2 looks at whether musical harmonics can predict what choices participants in money-sharing games will make; and chapter 3 takes an existing theroetical model of stochastic decision-making and changes it to help explain phenomena such as the overweighting of small probabilities, the willingness-to-accept—willingess-to-pay (WTA-WTP) disparity, and preference reversals.

Résumé

Le fait que le comportement économique de l'être humain comporte un élément irrationnel significatif - qui est à la fois difficile à expliquer et hautement prévisible - fascine les économistes depuis des décennies, de Fechner, 1860 à Shiller, 2005 et au-delà. Dans cette thèse, j'étudie cette problématique sous différentes perspectives: Le chapitre 1 examine l'influence du langage décrivant le comportement irrationnel dans les médias sur les marchés boursiers; Le chapitre 2 examine si les harmoniques musicaux peuvent prédire les choix des participants aux jeux de partage d'argent qu'ils feront; Et le chapitre 3 prend un modèle theroétique existant de prise de décision stochastique et le modifie pour aider à expliquer phénomènes comme surpondération de petites probabilités, la disparité entre la volonté d'accepter et la volonté de payer (VDA-VDP) et le renversement de préférences.

Contents

To Alexis and my Dad.

Introduction

The refinement of game theory in the 1940s and '50s was seen as a great advance in the field of economics and won John F. Nash Jr. the Nobel prize in economics in 1994 for his work on non-cooperative games (Nash, 1951). With its ability to provide concrete predictions about what rational agents would do in a variety of disparate scenarios, it stood as a touchstone for predicting economic behaviour (Von Neumann and Morgenstern, 1944).

However, its central conceit of agent rationality has been shown in numerous experimental situations to be false and unable to accurately predict behaviour in all but a few cases. The thing that surprised experimenters and practitioners of game theory, however, was the consistency with which individuals demonstrated supposedly irrational behaviour (Hirshleifer, 2001, 2015).

Take the dictator game, for example, wherein one individual is given an amount of money and, in one scenario, is told to share it with someone they don't know, don't meet, and are unlikely to ever meet. The other person has no recourse if they are given nothing so one would suppose that, given a strict focus on self-interest, a rational individual tasked with splitting the amount would choose to keep all the money for themselves. However, less than two-thirds of people do so (Eckel and Grossman, 1996).

Behavioural economics, and by extension behavioural finance, attempts to provide a way of predicting human behaviour on the assumption that there are some innate human characteristics such as social preferences, emotions, and moral preferences that defy traditional ideas of rationality but that the majority of people exhibit. In this thesis, I approach the field from a range of perspectives.

In Chapter 1, my co-authors Rajna Gibson Brandon, Mathieu Trépanier and I investigate whether news suggestive of irrationality within financial markets affect stock returns. Our work follows from that of Tetlock, 2007 who looks at whether the types of words used in the financial media affects stock returns. He finds that a higher-than-average use of negative words in particular precedes a downturn in stock returns followed by a reversal several days later.

There is abundant literature showing that irrationality plays a significant role in investor behaviour, for example, Shiller, 2005 details the phenomenon of 'Irrational Exuberance'. Using the methodology in Tetlock, 2007, we construct a lexicon describing 'market irrationality' and score daily news articles based on the proportion of words they contain from the lexicon. We find that market irrationality has a significant negative impact on subsequent stock market returns and exacerbates future stock market volatility. Furthermore, stocks with large positive irrationality risk betas outperform those with large negative irrationality risk betas by 0.353% monthly. We hypothesise that our market irrationality risk measure is a proxy for noise trader risk, which generates a positive risk premium, and conduct a number of robustness tests that support this conjecture.

In Chapter 2, I look closer at investor behaviour by investigating the value of heuristics as decision-making tools and create and test my own aesthetic heuristic based on music

theory. The concept of a heuristic is simple: you eliminate the number of options you have to choose from using superficial means thereby freeing up cognitive faculties allowing you to make an optimal choice among the remaining options. This is just one cognitive bias that people are subject to based on the summary in Hirshleifer, 2001.

The idea that money-sharing choices and musical harmonics are connected came when I reviewed Fehr and Gächter, 1999 whose results showed that people were choosing options that would be considered pleasant in Western music were they transformed into musical harmonics. To find out if there is a connection, I conduct an independent experiment to find out which musical harmonics are preferred and then test those results against existing data on money-sharing games.

I compare my heuristic with another, more established heuristic in the literature and find that my heuristic improves the fit of the model for all games and is a better fit for the model than the other, more established heuristic in roughly a half of the games.[1] Since my heuristic works as well as existing heuristics and given its enhanced flexibility and range, this is an important contribution to the analysis of money-sharing games.

Finally, in chapter 3, I turn to theoretical models of decision-making and, in particular, the strong utility model of P. R. Blavatskyy, 2014. This probabilistic decision-making model improves on existing stochastic models by eliminating the choice of stochastically-dominated options while maintaining the defining characteristics of stochastic decision-making models.

The author demonstrates that his model can resolve the preference reversal phenomenon (Grether and Plott, 1979), however, U. Schmidt and Hey, 2004 demonstrate that the preference reversal phenomenon is different for willingness-to-accept (WTA) and willingness-to-pay (WTP) cases. With that in mind, I add a new parameter to the P. R. Blavatskyy, 2014 model that accounts for the endowment effect. Not only does this change help account for the preference reversal phenomenon in both WTA and WTP cases, it also predicts the WTA-WTP disparity phenomenon and the overweighting of small probabilities.

[1] The existing heuristic I test against is one that reimagines each money-sharing game as one that has 10 choices. For example, if a game has 100 choices, this heuristic only considers every tenth choice, if it has 20 choices, this heuristic only considers every other choice. This heuristic requires a given game's number of choices to be divisible by 10. This is an example of rigidity in some existing heuristics that my heuristic does not have.

Chapter 1

Does Market Irrationality in the Media Affect Stock Returns?

Abstract. This paper investigates whether news suggestive of irrationality within financial markets affect stock returns. We construct a lexicon describing 'market irrationality' and score daily news articles based on the proportion of words they contain from the lexicon. We find that market irrationality has a significant negative impact on subsequent stock market returns and exacerbates future stock market volatility. Furthermore, stocks with large positive irrationality risk betas outperform those with large negative irrationality risk betas by 0.353% monthly. We hypothesise that our market irrationality risk measure is a proxy for noise trader risk, which generates a positive risk premium, and conduct a number of robustness tests that support this conjecture.

JEL Classification: G12, G14

Keywords: Linguistics, Market Irrationality, Media, Stock Market Returns, Volatility, Noise Trader Risk.

The work in this chapter was co-authored by Rajna Gibson Brandon[1] and Mathieu Trépanier[2]

1.1 Introduction

An abundant literature has studied irrational behaviour in finance. For instance, Daniel and Titman, 2000 examine how investor overconfidence can generate stock market momentum, Laibson, 1997 documents empirical evidence for time-inconsistent discounting and over-valuation of the present, and De Long et al., 1990 and Barber et al., 2006 show that noise traders can cause mispricing in financial markets. Shiller, 2005 covers the topic in extensive detail with explorations of the effects of irrationality on housing markets, the media, and financial institutions. Yet, there is no empirical evidence linking stock returns and the frequency of media characterisations of financial markets as irrational.

[1]SFI chaired professor in finance at the Geneva Finance Research Institute, University of Geneva, Geneva, Switzerland; Rajna.Gibson@unige.ch

[2]Associate researcher at the Swiss Institute for International Economics and Applied Economic Research, University of St Gallen, St. Gallen, Switzerland; mathieu.trepanier@unisg.ch

This study aims to bridge this gap and investigates how the use of irrational language to describe the market in news media affects stock returns. We pursue two main objectives: first, we assess the forecasting ability of a market irrationality sentiment measure extracted from the media on stock returns and volatility and, second, we examine whether innovations in the market irrationality sentiment measure represent a priced risk factor in stock returns.

To do so, we construct a lexicon of words suggestive of irrationality. By 'irrationality', we mean any action or decision that appears to go against what any reasonable person would do and is, most likely, an emotional reaction to something or someone. We do this irrespective of whether the word is positive, negative, or neutral. Due to the connotations of irrational behaviour, it's unsurprising to see that roughly 60% of the words in our 'irrationality' lexicon also appears in the Harvard IV-4 *Negativ* lexicon used by Tetlock, 2007. It should be noted, however, that of the 2291 words found in this lexicon, only about 4% appear in our 'irrationality' lexicon. With this in mind, its clear that the 'irrationality' lexicon is not equivalent to 'pessimism' or 'negative' words. Indeed, although most of our words have a negative connotation, they are explicitly chosen to denote irrationality, not pessimism.

We then look at the financial press and identify articles in the US over the 15-year period from 1998 to 2012 that describe irrational behaviour in the stock markets.[3] We begin by compiling a list of words that suggest irrationality and then validate it independently with three experts working in the fields of psychology and neuroscience. To construct the "market irrationality" sentiment measure, we use articles posted on the *Dow Jones Newswire*, which includes articles from a wide range of sources including the *Wall Street Journal*. To create a numerical measure, we follow Tetlock et al., 2008 and Loughran and McDonald, 2011 who compute the proportion of words from a specific lexicon in a given text. The irrationality sentiment measure, IRM, is determined by calculating the percentage of words from the irrationality lexicon that appear daily in a series of financial news articles.

We find that the market irrationality sentiment measure has a significant negative effect on subsequent stock market returns - proxied by the S&P 500 and the DJIA - and exacerbates stock market volatility, the full impact of which culminates over the first four days followed by a weak reversal in the following week. The impact for large stocks occurs after 1 lag, while for small stocks the effect is relatively protracted suggesting complex information that takes longer to be integrated in less liquid stocks.

We next define and estimate the irrationality risk factor by calculating the residuals of an autoregressive process. We then construct portfolios using stocks drawn from the S&P 500 index and which are sorted on the basis of their returns' sensitivity to this risk factor. We find the distribution of irrationality risk betas to be almost symmetric about zero with some stocks having large, positive irrationality risk betas and some having large, negative irrationality risk betas. Stocks with the most positive irrationality risk betas achieve an average 25-day return of 1.48% compared to 1.06% for the stocks with the most negative irrationality risk betas. This difference is highly statistically significant, amounts to about 4.31% annually, and persists through most of our robustness checks: controlling for standard risk factors, volatility risk, different holding periods, windsorisation, and running cross-sectional regressions. We lose significance in the subsample analysis and in using non-overlapping data due to, respectively, the learning phenomenon and insufficient observations. This finding suggests that market irrationality is a source of risk that we

[3]Further details on the construction of the lexicon are provided in Section 1.3.1.3.1.

hypothesise is a proxy for noise trader risk as described in De Long et al., 1990. Stocks, in which noise traders invest, experience an increase in risk and higher expected returns. This is also consistent with the large variation of market irrationality risk factor betas for stocks that we observe over time as noise traders' attention moves around the market.[4] When sorting on size, book-to-market ratios, and volatility risk, we find a concentration of the irrationality risk premium on the high-minus-low IRM beta portfolio for mid-to-large-sized firms, for firms with a mid-level book-to-market ratio, and for firms with volatility risk at the extremes.

This study contributes both to the literature on the impact of language used in the media on financial markets and to the literature on investor irrationality. It also contributes to the literature on noise trader risk, for example, De Long et al., 1990 and Biais et al., 2010, which we associate with language in the media. We show that irrational stock market sentiment as expressed in the media has a negative subsequent impact on the investment opportunity set. We further document that market irrationality is a priced risk factor and present an investment strategy - long on the High IRM beta portfolio and short on the Low IRM beta portfolio - that produces a significant positive return, even when accounting for standard risk factors, volatility risk, and outliers.

Earlier literature on the role of the media has focused primarily on news conveying optimistic and pessimistic sentiment and on the impact of investor attention to such news. To our knowledge, our study is the first to examine language focusing on market irrationality in the media and the first one to document how the sentiment and the risk it conveys affects stock market prices.

The outline of this paper is as follows: Section II provides a brief literature review. Section III describes the data used in the study and details the construction of the market irrationality lexicon and sentiment measure. Section IV describes the impact of market irrationality on subsequent stock market returns and volatility. Section V examines whether irrationality risk is priced in stock returns and conducts robustness tests. Section VI concludes the study.

1.2 Literature Review

This study relates to two different fields: media analysis and irrational behaviour in economic and financial decision-making.

Using text to provide insights into stock market movements above what can be taken from numerical data has attracted many researchers. There are several ways to approach this task: some look at news and user-generated content such as newswires or online message boards while others look at company documents such as 10-K filings or earnings reports. Some look at the volume of internet searches while others examine the content of documents. The latter splits into those that take a lexical approach and those that take a classification approach.

Antweiler and Frank, 2004 and Das and Chen, 2007 classify user-generated content (UGC) on internet message boards about finance into pessimistic, optimistic, and neutral signals and find that a high volume of posts about individual firms is linked to lower returns and higher volatility the following day. Tirunillai and Tellis, 2012 and Da et al., 2011 use online search volume for certain firms and find that an increase in search volume for a particular company or its most popular product leads to significant positive abnor-

[4]Data on market irrationality risk beta movement is available on request.

mal returns. Tirunillai and Tellis, 2012 also find that negative UGC leads to significant negative abnormal returns while positive UGC has no significant effect on these metrics.

The following papers all look at positive or negative sentiment using news sources. Tetlock, 2007 uses a single column - *Abreast of the Market* - in the *Wall Street Journal* (*WSJ*) and a lexical approach to predict returns in the Dow Jones Industrial Average. He uses the Harvard-IV-Psychosocial dictionary, constructs sentiment factors for all its word categories, and finds that negative sentiment predicts lower stock returns on the following day and a reversal some days later. Tetlock et al., 2008 expand the news source to include all articles in the *WSJ* and *Dow Jones Newswire* and obtain similar results with the additional finding that market prices underreact to firm-specific news. García, 2013 continues by looking at columns in the *New York Times* over a much longer period (1905-1958) and concludes that the result found by Tetlock, 2007 is concentrated during recessions.

Loughran and McDonald, 2011 also use a lexical approach but adapt the word lists used in Tetlock, 2007 to be representative for financial texts.[5] Loughran and McDonald, 2011 use 10-K reports and find that firms whose reports contain a high proportion of negative words experience lower subsequent returns and find little effect from positive words. Li, 2006 also looks at 10-K filings but specifically at words that reflect 'riskiness'. He finds that risk sentiment is associated with lower future earnings.

Da et al., 2015 and Manela and Moreira, 2013 both look at investor concern and its impact on the stock market. The former use Google search volume and look for queries about household concerns such as "recession", "unemployment", and "bankruptcy". They construct a sentiment index and find that it predicts short-term return reversals, temporary increases in volatility, and mutual fund flows out of equity funds and into bond funds. Manela and Moreira, 2013 construct a sentiment index about the concerns of the average investor using the front-page of the *WSJ* and a longer time period (1890-2009). They find that periods where people are more concerned about a rare disaster are either followed by above-average stock returns or by periods of large economic disasters. In general, they find evidence consistent with the view that rare disaster risk is an important driver of asset prices.

Yuan, 2015 looks both at Dow record events and front-page news events, the latter of which he defines as occasions on which both the *New York Times* and the *Los Angeles Times* cover the change in price level of the domestic stock market within front-page articles. He finds that these events are linked to an increase in selling and a drop in market returns, which he links to an increase in investor attention, their information processing activity, and subsequently the management of their portfolios. While Dow record events have this effect in all cases, front-page news events only seem to have this effect when the market is high.

We next provide a brief review of the literature on irrational behaviour in economic and financial decision-making. Hirshleifer, 2001 gives a detailed summary of the cognitive biases that impair investors' ability to make rational decisions. He also summarises many of the ways researchers have tried to adapt their models in order to replicate this kind of behaviour. Laibson, 1997 provides evidence that hyperbolic discounting may be the reason for the ongoing decline in savings rates in the U.S and Diamond and Köszegi, 2003 expand on this model with a specific interest in savings and retirement.

Of particular interest is the work of De Long et al., 1990 who find that noise traders can earn higher expected returns than their more-informed colleagues by profiting from

[5] García, 2013 also uses the Loughran and McDonald, 2011 lexicon in his study.

the increased risk that their presence generates. They demonstrate that, although arbitrageurs should be able to profit from mispricing due to noise traders in theory, noise traders' beliefs may not revert to their mean for a long time and may, indeed, become more extreme. Thus noise traders create a space for themselves that other traders must interact with. This model accounts for several financial anomalies including the Mehra-Prescott equity premium puzzle (Mehra and Prescott, 1985); consistent with which they find that asset prices diverge from fundamental values and later mean-revert. Biais et al., 2010 build on this by developing an investment strategy that successfully out-performs the index as well as a comparable momentum strategy.

Brennan and Lo, 2012 look at the evolutionary reasons for bounded rationality by using mathematical models that demonstrate that the paradigm in which humans use their own intelligence in order to maximise their own self-interest is just one of many possible outcomes of natural selection. Brennan and Lo, 2011 present a single evolutionary concept that explains behaviour including risk-sensitive foraging, risk aversion, loss aversion, probability matching, randomisation, and diversification. Modern thinking in humans and other great apes occurs in several decision-making locations in the brain such as the prefrontal cortex - believed to be the source of individually rational behaviour - and other components such as the amygdala, which is responsible for the "fight-or-flight" response. Financial decision-making and how it shapes financial markets is the product of these neural networks and, from an evolutionary perspective, is neither efficient nor irrational - it is adaptive.

Dow and Gorton, 2006 give an overview of the impact of noise traders on financial markets. Since they allow informed traders to capitalise on their private information, they play an essential role in modern finance theory, however, their identities, motivations, and persistence remain topics of research. Brown, 1999 finds that the sentiment of individual investors is related to increased volatility in closed-end investment funds. This only happens during trading hours showing that irrational investors only affect prices through trading. Barber et al., 2006 report evidence consistent with noise trader models in which the trading of stocks by uninformed investors causes mispricing.

Tetlock et al., 2008 document an underreaction of stock markets to news, as do Huynh and D. R. Smith, 2013 who find that this is the main driver of momentum effects globally.

Finally, we derive inspiration from Robert J. Shiller's book, *Irrational Exuberance* (2005). The title is a reference to Federal Reserve Board chairman Alan Greenspan who, in December 1996, is quoted as saying:

> "Clearly, sustained low inflation implies less uncertainty about the future, and lower risk premiums imply higher prices of stocks and other earning assets. We can see that in the inverse relationship exhibited by price/earnings ratios and the rate of inflation in the past. But how do we know when *irrational exuberance* has unduly escalated asset values, which then become subject to unexpected and prolonged contractions as they have in Japan over the past decade?"

Shiller, 2005 further refers to information cascades whereby the news media make connections between current events and sequences of events in the past, which can then cause similar sequences to occur in the future.[6]

[6]Shiller first published his book in 2000 and in Chapter Five (pp. 85-105) he assesses the role of the news media on stock market speculation.

1.3 Data

In this study, we use text data from the Dow Jones Newswire in order to construct the market irrationality sentiment measure. We do not identify articles whose subject is a specific company, but focus on articles that make reference to the US stock market as a whole, which allows us to construct a market-wide irrationality sentiment measure against which we can measure individual stocks' sensitivity.

1.3.1 Constructing the market irrationality sentiment measure and risk factor (IRM)

The first step is to identify words that describe irrational behaviour in the market. Other studies have used pre-existing lexicons, however, to our knowledge, a lexicon describing irrational behaviour did not already exist. We thus compiled an initial list of words using the Harvard-IV-4 Psychosocial Dictionary and General Inquirer categories as a template and asked three experts from the fields of neuroscience and psychology to validate which words would comprise the final lexicon.[7] The complete lexicon of words used in this study appears in Appendix I.

We performed a search of all articles on the Dow Jones Newswire, written in English, and reporting on North America, that included either of the words 'market', 'markets', 'Dow', 'NASDAQ', or 'NYSE', but not 'Moody's', "Dow Jones reported", nor "Dow Jones said", within a five-word proximity of any of the words from our irrationality lexicon. We apply a series of filters to eliminate tables, summaries of news stories, articles of 50 words or fewer, as well as any weblinks, subheadings, and any attribution text that were not relevant to the news story.[8]

We then used the LIWC 2007 program presented in Tausczik and Pennebaker, 2010 to assign a score to every article equal to the percentage of words in the article that appear in the irrationality lexicon. We turned this into a daily irrationality sentiment measure (from here on referred to as IRM^{raw}) by taking every article published after 1700 EST on day $t-1$ and before 1529 EST on day t and taking the simple average of their scores. In our sample, days $t-1$ and t refer to trading days, not calendar days. This means that most observations cover a single day's worth of news while others cover several days, usually over weekends and public holidays. This gave us daily data for the whole period from 2nd January 1998 to 31st December 2012, for which we provide the summary statistics in the first row of Table 1.1.

IRM^{raw} includes 494 (13.1%) null observations out of a total of 3773 observations. The large number of zeros indicates that there were many days where no articles fitting our conditions were published. In total, 11727 articles fitting our conditions were available on the Dow Jones Newswire over the given period.

We continue by normalising IRM^{raw} over a rolling 6-month period - a compromise to minimise look-ahead bias while maximising the number of observations. We then estimate the residuals using the following autoregression with $p = 2$

$$IRM_t^{norm} = \alpha + \sum_{s=1}^{2} \phi_s \cdot IRM_{t-s}^{norm} + \epsilon_t. \tag{1.1}$$

[7] Instructions provided to the experts and their short CVs can be found in Appendix II.

[8] The complete algorithm for which can be found in Appendix III.

$p = 2$ was chosen above other values because it represented the largest increase in the adjusted R squared (from $p = 1$ to $p = 2$) than any other value between 1 and 10. Table 1.2 shows IRM^{norm} is weakly autocorrelated and thus is largely unpredictable.

We proceed by denoting the innovations estimated in (1.1) as simply IRM. We refer to the latter variable as the **market irrationality risk factor**. Table 1.1 shows the summary statistics of the IRM^{raw} sentiment measure along with its normalisation and the market irrationality risk factor and, due largely to the relative lack of autocorrelation in the risk factor, the statistics for the innovations do not differ greatly from those of the normalised IRM^{norm} sentiment measure. The mean and standard deviation remain roughly the same with large positive skewness and kurtosis denoting the appearance of several large, positive values.

[Insert Tables 1.1, 1.2, and 1.3]

Table 1.3 shows the correlation between the three Fama-French factors, the Carhart momentum factor, the normalised market irrationality sentiment measure, and the market irrationality risk factor. It can be seen that the correlations between the market irrationality sentiment measure (and therefore also the market irrationality risk factor) and other standard risk factors are close to zero.

1.3.2 Other data

Stock data such as prices, returns, trading volumes, and the number of outstanding shares are collected from the Center of Research in Securities Prices (CRSP) Daily Stocks Combined File which includes all stocks actively traded on the NYSE, AMEX, and Nasdaq. Only ordinary common shares (with CRSP share code 10 or 11) are considered in this study. In addition, only companies that form part of the S&P500 index and have at least 250 days of trading data between 1st January 1998 and 31st December 2012 are included. For each company, we identify the most recent occasion on which it was included in the S&P500 index and use all its data starting from up to one calendar year prior to its inclusion. The sample includes 637 individual firms. Appendix IV presents the process in more detail. Table 1.4 shows the summary statistics of the daily firm sample size. Data on the Dow Jones Industrial Average was obtained from CRSP and the St. Louis Fed (FRED).

[Insert Table 1.4]

Data on the Fama-French factors - market excess return, size factor, book-to-market factor - as well as the implied risk-free rate and Carhart momentum factor are taken from Kenneth French's website. We also construct a $FINNEG^{raw}$ measure using the same method for constructing the IRM^{raw} measure but instead using the lexicon of 'negative financial' words provided by Loughran and McDonald, 2011. This lexicon was constructed to correct for the fact that many lexicons of 'negative' words developed in the field of psychology include words such as *tax, cost, foreign,* and *liability*, which, in a financial context, have a neutral meaning. The Loughran and McDonald, 2011 word list specifically focuses on negative words in a financial context making it an appropriate lexicon to use for the alternative negative sentiment measure. A risk factor, $FINNEG$, is derived from $FINNEG^{raw}$ in the same manner that IRM is derived from IRM^{raw} except using $p = 3$ instead of $p = 2$.

1.4 Market Activity and Irrationality

In this section, we follow Tetlock, 2007's methodology by investigating the ability of the market irrationality sentiment measure to forecast stock market returns and volatility. Tetlock, who considers investor pessimism, focuses on the Dow Jones Industrial Average because he derives his measure from the WSJ column 'Abreast of the Market' which covers the Dow Jones Index. Since the market irrationality sentiment measure incorporates all articles recovered from the Dow Jones Newswire, we look at the Dow Jones Industrial Average as well as the S&P 500 Index and a series of portfolios using S&P 500 firms sorted on size and book-to-market ratios.

We hypothesise that higher market irrationality sentiment embedded in the language of the media is associated with higher uncertainty. As a result, we expect to see a deterioration in the investment opportunity set, that is to say, lower future stock market returns and higher future stock market volatility.

1.4.1 VARs

We conduct a series of Vector Autoregressions (VARs) using portfolio or stock market returns (R), the normalised irrationality sentiment measure (IRM^{norm}), and a proxy for volatility using the CBOE VIX Index (VIX). We include a series of exogenous variables ($Exog$) that comprise 5 lags of share volume specific to the portfolio being analysed, dummies for the days of the week, a dummy for January, and dummies for extreme negative stock market events on the following dates: 31st August 1998 (Russian financial crisis), 14th April 2000 (dot-com bubble), 17th September 2001 (September 11th attacks), 29th September 2008, and 15th October 2008 (the subprime financial crisis).[9][10]

As in Tetlock, 2007, we define the lag operator, L5, to be the transform of variable x_t to the vector consisting of the five lags of x_t, that is,

$$L5(x_t) = [x_{t-1}\, x_{t-2}\, x_{t-3}\, x_{t-4}\, x_{t-5}].$$

In this way, the first two sets of VARs can be expressed as

$$R_t = \alpha_1 + \beta_1 \cdot L5(R_t) + \gamma_1 \cdot L5(IRM_t^{norm}) + \delta_1 \cdot L5(VIX_t) + \lambda_1 \cdot Exog_t + \epsilon_{1t} \quad (1.2)$$

and

$$VIX_t = \alpha_2 + \beta_2 \cdot L5(R_t) + \gamma_2 \cdot L5(IRM_t^{norm}) + \delta_2 \cdot L5(VIX_t) + \lambda_2 \cdot Exog_t + \epsilon_{2t} \quad (1.3)$$

We focus on the γs as they describe the dependence of the portfolio returns and market volatility on the market irrationality sentiment measure. Table 1.5 summarises the estimates of γ_1 when R describes stock index returns and shows that the market

[9] 29.09.2008: The U.S. House of Representatives' failure to pass the Bush Administration's $700 billion bailout plan triggered the biggest one-day point drop in the history of the Dow Jones industrial average. This happened two weeks after Lehman Brothers filed for bankruptcy. Source: TIME: http://content.time.com/time/specials/packages/article/0,28804,1845523_1845619_1845541,00.html. The Guardian: http://www.theguardian.com/business/2008/sep/15/lehmanbrothers.creditcrunch

[10] 15.10.2008: "The Dow Jones dropped in response to a report that retail sales have reached a 3-year low and a speech by Federal Reserve Chairman Ben Bernanke in which he says the economic recovery will be slow." Source: http://www.infoplease.com/business/economy/declines-dow-jones-industrial-average.html

irrationality sentiment measure has, over the following four days, a negative impact on the Dow Jones Industrial Average amounting to a 11.4 basis point drop significant at the 10% level as well as a negative impact on the S&P 500 index amounting to a 16.5 basis point drop significant at the 1% level. Table 1.6 shows the values for γ_2 and, again over the following four days, we find that the market irrationality sentiment measure predicts an increase of 0.15 in stock market volatility significant at the 5% level regardless of which index we use for R_t. For all these results, the impact peaks on the third day. This shows that an increase in the market irrationality sentiment measure is associated with a deterioration in the opportunity set. We conjecture that IRM is a proxy for uncertainty and specifically for overreactions to financial news which is driven by noise traders.

Table 1.7 summarises the estimates of γ_1 when R describes portfolio returns sorted on size and book-to-market ratio and shows that the market irrationality sentiment measure has a significant negative impact on all firm portfolios regardless of size or book-to-market ratio over the following four days at at least the 5% level. This amounts to a 20.4 basis point drop for small stocks, a 15.2 basis point drop for large stocks, a 13.6 basis point drop for value stocks, and a 27.6 basis point drop for growth stocks. We find a quicker impact for large firms than for small firms and conjecture that the timing of changes is dependent on liquidity.

Whereas in Tetlock, 2007 the first lag of the 'bad news' measure is most pertinent for stock returns, the market irrationality sentiment measure has a protracted effect on all stocks which peaks on the third lag. We interpret this delayed response by recognising that irrationality is not a straight-forward concept to interpret so its impact may take longer to be integrated into the market - this is compounded for small stocks who are more illiquid in general. While Tetlock finds a significant impact of the "bad news" measure on the first lag and a weak reversal on the 4th lag, we show that the market irrationality sentiment measure has a longer-lasting negative impact on returns over the first week. By considering up to 10 lags instead of 5, we find a reversal in the following week for all portfolios and indices.[11]

The third set of VARs examine whether the market irrationality sentiment measure depends on stock returns or market volatility. These are expressed as

$$IRM_t^{norm} = \alpha_3 + \beta_3 \cdot L5(R_t) + \gamma_3 \cdot L5(IRM_t^{norm}) + \delta_3 \cdot L5(VIX_t) + \lambda_3 \cdot Exog_t + \epsilon_{3t} \quad (1.4)$$

There does not appear to be a clear causal relationship going from stock returns or volatility to the market irrationality sentiment measure as there are no significant coefficients when using either the S&P 500 index or the DJIA.[12]

Using a Granger causality test, we see that market irrationality significantly predicts both stock market returns and volatility over the full 5 lags (at the 10% level for the DJIA, at the 1% level for the S&P 500 index, and at the 5% level for the VIX index); we do not observe causality in the opposite direction.

1.4.2 Robustness Analysis

As a robustness check, we include a news measure to account for "bad news". We adopt the lexicon developed by Loughran and McDonald, 2011 as it is designed to take into

[11] All results on the reversals are available on request.

[12] Coefficient estimates are available upon request.

account the financial nature of the news articles we are interested in.[13] We apply this lexicon to score the articles we collected for the market irrationality sentiment measure and call it $FINNEG^{raw}$ ($FINance\ NEGative$).

The fourth set of VARs look at how stock returns depend on the market irrationality and negative sentiment measures. These are expressed as

$$R_t = \alpha_4 + \beta_4 \cdot L5(R_t) + \phi_4 \cdot L5(FINNEG_t^{norm}) + \gamma_4 \cdot L5(IRM_t^{norm}) + \delta_4 \cdot L5(VIX_t) + \lambda_4 \cdot Exog_t + \epsilon_{4t} \tag{1.5}$$

We find no significant results for $FINNEG^{norm}$ when looking at either stock market index for any of the 5 lags.[14] Tetlock, 2007 finds a significant negative impact on the DJIA at the first lag for his "bad news" measure with a significant reversal on the 4th lag. One of the reasons we might not see this is that the two datasets are structurally different with respect to both the lexicon used to retrieve the articles and the source of the text data. Even so, after controlling for negative news, we still obtain significantly negative coefficients for IRM^{norm} on S&P 500 and DJIA returns over the four days that follow.

We conclude that market irrationality affects both stock returns and stock market volatility but that the timing of the impact varies according to stock characteristics. For instance, for large stocks or stocks with low book-to-market ratios, there is an immediate impact while for small stocks or stocks with a high book-to-market ratio, the impact takes time to materialise. Similar results are obtained when we consider stock indices: the DJIA which, on average, contains larger stocks than the S&P500 shows a quicker impact while the impact for the S&P500 is significant only on the third lag. In both cases, the impact of our market irrationality sentiment measure on market volatility occurs on the first lag and continues on the third lag. These results are consistent with the irrationality conveyed in the media being relatively complex information to integrate into prices, particularly for less liquid stocks.

1.5 Is Market Irrationality Priced in Stock Returns?

We have seen that market irrationality depresses the subsequent investment opportunity set. We next examine whether innovations in the irrationality sentiment measure are a common and priced source of risk in stock returns. We hypothesise that the market irrationality risk factor is a proxy for noise trader risk (see De Long et al., 1990 and Biais et al., 2010), as a result, we expect IRM to have a positive risk premium that persists when controlling for other risk factors. We conjecture that stocks that have a high exposure to market irrationality sentiment are also those with a large investment by noise traders that, according to the noise trader risk hypothesis, is itself a source of risk that leads to higher returns.

We start by sorting the sample of S&P 500 firms into ten portfolios based on their dependence on the market irrationality risk factor, IRM. For each day in our sample and for every firm trading on that day, we run the regression

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_{i,MKT} MKT_t + \beta_{i,IRM} IRM_t + \epsilon_{i,t} \tag{1.6}$$

[13] 13 words appear in both the IRM and $FINNEG$ lexicons. This compares to a total of 144 words in the IRM lexicon and 2337 in the $FINNEG$ lexicon.

[14] The values for ϕ_4 and γ_4 are available on request.

where MKT_t is the daily excess market return, IRM_t is the market irrationality risk factor, and $r_{i,t} - r_{f,t}$ is the daily excess return for firm i. The market irrationality factor beta, β_{IRM}, is estimated using the 63 preceding observations (3 months) while controlling for the market beta, β_{MKT}. Firms whose betas are in the first decile are assigned to the first portfolio, firms whose betas are in the second decile are assigned to the second portfolio, and so on. These portfolios are constructed the day after the beta estimation is conducted and held for twenty-five days; the 25-day portfolio return is calculated as the equally-weighted average return of the firms in that portfolio.

Since we hold the portfolios for twenty-five trading days, we compute the 25-day Fama-French and Carhart factors by calculating the return on each factor running from trading day $t - 25$ to trading day t.[15]

We find that the correlation between the market irrationality risk factor and the Fama-French and Carhart factors is negligible for both the daily and 25-day estimates.[16] In addition, the $\hat{\beta}_{IRM}$s estimated by taking systematic market risk into account, using equation (1.6), are almost symmetric about zero. Thus, creating any portfolio that is long in both portfolio n and portfolio $11 - n$: 5 and 6, 4 and 7, etc., has a market irrationality risk beta close to zero.[17]

1.5.1 The Impact of the Market Irrationality Risk Factor on Expected Stock Returns

Table 1.8 reports summary statistics of the daily returns of ten portfolios based on the estimated $\hat{\beta}_{IRM}$: minimum, median, maximum, mean standard deviation, skewness, and kurtosis. For the majority of these portfolios, skewness is negative suggesting that the portfolios are subject to occasional, large negative returns. The one exception to this is the portfolio with the largest, negative market irrationality beta which has very high positive skewness caused by occasional, large positive returns. This portfolio also exhibits the largest standard deviation and largest kurtosis with a large maximum and large minimum in absolute terms. All portfolios have large tails and, interestingly, many of the summary statistics are U-shaped or inverse U-shaped.

The portfolio mean returns appear to decrease almost monotonically from the High (large, positive IRM beta) portfolio with an average 25-day return of 1.48% to the Low (large, negative IRM beta) portfolio with an average 25-day return of 1.06%. This amounts to an average 25-day difference between the High and Low portfolios of 0.42% or about 4.3% annually.[18] This difference is significant at the 1% level.

We next calculate the risk-adjusted performances (alphas) for all ten portfolios sorted on the market irrationality risk factor using the performance evaluation models developed by Fama and French, 1993 - the three-factor model (henceforth FF), and by Carhart, 1997

[15]When comparing these longer-term risk factors calculated in this way with those published on Kenneth French's website, we find that the values match in all cases where data is available. The risk factor data we use are available upon request.

[16]The largest correlation by absolute value is -0.0420 between IRR^{norm} and HML. All correlation data are available on request.

[17]Nonetheless, all portfolios combining portfolio n and $(11 - n)$, except $High + Low$, have negative IRM betas significant at the 1% level. The beta of the $High + Low$ portfolio is not significantly different from 0.

[18]Calculated on the assumption that there are 252 trading days per year.

- the four-factor model (henceforth Carhart), which are, respectively,

$$r_{i,t} - r_{f,t} = \alpha_{i1} + \beta_{i,MKT} MKT_t + \beta_{i,SMB} SMB_t + \beta_{i,HML} HML_t + \epsilon_{i,t} \tag{1.7}$$

and

$$r_{i,t} - r_{f,t} = \alpha_{i2} + \beta_{i,MKT} MKT_t + \beta_{i,SMB} SMB_t + \beta_{i,HML} HML_t + \beta_{i,UMD} UMD_t + \epsilon_{i,t} \tag{1.8}$$

where $r_{i,t}$ is the return of portfolio i, $r_{f,t}$ is the one-day risk-free interest rate, MKT_t is the excess market return, SMB_t is the excess return of all small-cap stocks over large-cap stocks, HML_t is the excess return of value stocks over growth stocks, and UMD_t is the excess return of the prior month's winning stocks over losing stocks.

Table 1.9 shows the results wherein we see a strong monotonic trend running from the Low IRM beta portfolio with an alpha of 0.324% in the FF case and 0.699% in the Carhart case, to the High IRM beta portfolio with an alpha of 0.744% in the FF case and 0.884% in the Carhart case. The portfolio constructed by going long in the High IRM beta portfolio and short in the Low IRM beta portfolio produces a daily alpha of 0.420% in the FF case and 0.184% in the Carhart case. These alphas are significant at the 1% and 5% levels respectively and suggest that a strategy that goes long in the High IRM beta portfolio and short in the Low IRM beta portfolio could deliver a significant annual alpha of about 4.3% (1.9% in the Carhart model), providing further evidence that the traditional risk factors do not fully cover the risk characteristics that drive stock returns.[19]

Our findings are consistent with a positive risk premium because our market irrationality sentiment measure is acting as a proxy for noise trader risk. To further test the noise trader risk hypothesis, we examine whether we find evidence of mean-reversion in stock returns. Indeed, we find by looking at the 25-day returns prior to the formation of the portfolios i.e. the portfolio returns from trading day $t-25$ to trading day t, that the return on the High-minus-Low IRM beta portfolio is negative (-0.196%) with a t-stat of -1.5712. This is not significant, however, it goes in the direction we would expect.

Sorting by Size, Book-to-Market Ratio, and the Market Irrationality Risk Factor Beta

We next split the firms into 25 portfolios sorted by their size and their estimated market irrationality risk factor betas. First, we construct a ranking of all firms according to their size; this ranking is updated every twenty-five trading days.[20] Every day, we sort all the firms into five equally-sized portfolios according to their market irrationality risk factor beta estimated on the prior 25 trading days and then use the size ranking to sort these five portfolios into a total of twenty-five equally-sized portfolios. We hold these portfolios for twenty-five days and calculate the 25-day return.

Panel A of Table 1.10 shows the mean 25-day returns for all Size-IRM beta portfolios and it is clear that the portfolios with positive risk betas outperform those with negative risk betas for all firms above a certain size, ranging from a difference of -0.043% for small firms (not significantly different from zero) to 0.852% (significant at the 1% level) for mid-sized firms. In general, we see a significant impact on all mid-cap stocks and larger.

We perform the same operation but use the stocks' book-to-market ratios to double-sort the portfolios rather than their sizes. Panel B of Table 1.10 shows the mean 25-day

[19] This strategy may, however, be less profitable once transaction costs are accounted for.

[20] Size is calculated by multiplying the number of shares outstanding by the share price.

returns for all B/M-IRM beta portfolios and the difference between the High and Low IRM beta portfolios persists albeit without any clear monotonicity running from growth firms to value firms. For value firms, the difference between the High and Low IRM beta portfolios' returns is 0.227% and for growth firms, the difference is 0.231% (both significant at the 5% level). In this sorting, mid-level firms are disproportionately affected with returns of 0.333% significant at the 1% level.

Cross-Sectional Regression Test

In order to evaluate the importance of IRM relative to standard risk factors, we continue by using the cross-sectional regression method of Fama and MacBeth, 1973 and, as in Gibson and Wang, 2016, we estimate the market irrationality sentiment measure's beta for each individual stock, rather than combining them into portfolios, using the Fama-French model:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_{iMKT} MKT_t + \beta_{iSMB} SMB_t + \beta_{iHML} HML_t + \beta_{iIRM} IRM_t + \epsilon_{i,t}$$

where $r_{i,t}$ is the return of stock i, $r_{f,t}$ is the one-day riskless interest rate, MKT_t is the excess market return, SMB_t is the excess return of small-cap stocks over big-cap stocks, HML_t is the excess return of value stocks over growth stocks, and IRM_t is the market irrationality risk factor.

We test to see if the individual stocks' expected returns are cross-sectionally related to the risk factor betas as such:

$$r_{i,t} - r_{f,t} = \gamma_0 + \gamma_1 \beta_{iMKT,t-1} + \gamma_2 \beta_{iSMB,t-2} + \gamma_3 \beta_{iHML,t-1} + \gamma_4 \beta_{iIRM,t-1} + u_{i,t}$$

Theoretically, if IRM is a priced risk factor consistent with what we've already observed, γ_4 should be significantly positive.

We estimate betas over a rolling 63-day period (equivalent to 3 months) for each stock which we then normalise and use in the cross-sectional regression using the following 25-day period. We combine the coefficients of all stocks and, due to the high number of observations, we use non-overlapping 25-day periods.

Regarding our hypothesis, we find that the market irrationality risk beta coefficient is positive, equal to 0.574%, and highly statistically significant with a t-stat 25.2706. This compares to the market risk beta of -0.430% and a t-stat of -10.5091. Both based on a total of 64,545 observations. This premium appears to correspond to the values we get in the 3-factor risk-adjusted time-series analysis of 0.420% and -0.131% for the IRM and market risk factors respectively and suggests that this is an economically meaningful effect.

1.5.2 Robustness Checks

The results so far suggest that stocks with positive exposure to the market irrationality risk factor earn higher subsequent returns than stocks with negative exposure. However, there is a possibility that this effect could have been induced by model mis-specifications. We perform a series of robustness checks to address these concerns.

Sorting by Volatility Risk Betas

Ang et al., 2006 examine whether stock market volatility is a priced risk factor and estimate the price of aggregate volatility risk. This is accomplished by constructing a risk factor from the innovations in the VIX index. They estimate the stocks' betas with respect to this factor and then construct 5 portfolios of individual stocks sorted by their volatility betas. Using the returns on these 5 portfolios as the vector X_t, they estimate the regression

$$\Delta VIX_t = a + b'X_t + \epsilon_t$$

and use $b'X_t$ as their primary volatility risk factor, which they call $FVIX_t$. We use their method to construct the $FVIX$ factor from our dataset.

Ang et al., 2006 find that $FVIX$ carries a statistically significant negative price of risk and that this is robust to controlling for size, value, momentum, and liquidity effects. Our own research confirms this result.[21]

One concern is that stock market irrationality risk may merely capture stock market volatility risk. Two sets of results mitigate this concern, however. First, the correlation between volatility risk as proxied by the $FVIX$ factor and stock market irrationality risk is very low and equals 0.0135. Second, we construct 25 portfolios sorting stocks on their $FVIX$ and IRM betas. Panel C of Table 1.10 shows the mean 25-day returns for all FVIX beta-IRM beta portfolios.

For Low FVIX beta firms, the difference between the High and Low IRM beta portfolios' returns is 0.274% (significant at the 1% level). For High FVIX beta firms, the difference between the High and Low IRM beta portfolios' returns is 0.205% (significant at the 5% level). We conclude that IRM risk is not merely capturing volatility risk as its pricing prevails in both low and high $FVIX$ beta-sorted portfolios.

Holding Portfolios for Different Period Lengths

In the main results, we hold the portfolios for twenty-five days; in Table 1.11 we use the same ten portfolios but instead hold them for 5, 10, 15, and 20 (trading) days to see if different holding periods change our results. Looking at Table 1.11, we see that holding periods shorter than 15 days do not carry market irrationality risk premia that are significantly different from zero. A holding period of 5 days gives us an average daily difference of -0.000%; the difference only becomes significant (at the 5% level) as we consider a holding period of 15 days for which the average daily difference is 0.011% with a t-stat of 2.4058. The average daily difference for 20 days is 0.015% with a t-stat of 3.6225 (significant at the 1% level) and it's more significant still for 25 days. The increasing trend in significance as the holding period lengthens suggests that the pricing of irrationality risk is a rather persistent phenomenon.

Subsample Analysis

In Table 1.12, we examine the portfolios' performance over two separate subsamples: the first one from the 9th November 1998 to the 5th December 2005 and the second one from the 6th December 2005 to the 31st December 2012. The start of the period is dictated by the time required to estimate the IRM risk betas and the mid-point is chosen such that

[21] Relevant portfolio returns are available on request.

each subsample has an equal number of observations.We see that the previous results still hold albeit mostly for the pre-2006 period where the High IRM portfolio outperforms the Low IRM portfolio by 0.664% with a t-stat of 4.5794. The post-2005 period also has the High IRM portfolio outperforming the Low IRM portfolio by 0.175%, however, this result is not significant with a t-stat of 1.6341.

In order to explain the difference in the two periods, we conjecture that we are observing a learning process. The original 'irrational exuberance' comment made by then-Fed chairman Alan Greenspan, occurred on the 5th December 1996. The full comment was

> "Clearly, sustained low inflation implies less uncertainty about the future, and lower risk premiums imply higher prices of stocks and other earning assets. We can see that in the inverse relationship exhibited by price/earnings ratios and the rate of inflation in the past. But how do we know when *irrational exuberance* has unduly escalated asset values, which then become subject to unexpected and prolonged contractions as they have in Japan over the past decade?"

These comments were immediately followed by a sharp fall in the Tokyo stock market and later in other world markets. Since this happened less than 2 years before our first subsample begins, we argue that irrationality sentiment still had a strong effect. Over ten years later, we may be observing some acclimatisation to the media language focussing on irrationality, which may inhibit the effect we see later in the sample.

Winsorization

To examine if our results are driven by extreme returns delivering false positives, for each firm we windsorise the data at the 1% and 2% levels (that is to say that a dataset windsorised at the n% level replaces all the observations below the nth percentile with the value at the nth percentile, and all the observations above the $(100-n)$th percentile with the value at the $(100-n)$th percentile) using the data covering the current year and those immediately preceding and following the current year. Then, we perform all the sorting and portfolio return calculations using the methods outlined previously. We can see from Table 1.13, that the results remain significant at the 1% level, even when stock returns are windsorised at the 2% level.

Non-overlapping data

The main dataset uses overlapping data, however, this runs the risk of over-emphasising extreme observations. We look at a non-overlapping dataset that reduces the number of observations from 3558 to 143 and the results are presented in Panel A of Table 1.14. The difference between the High and Low IRM beta portfolios' returns remains positive, however, the difference is no longer significant due to the large reduction in observations. The t-stat is 1.0615. This may be due to non-overlapping data leading to too few observations (only 143 compared to the full sample's 3558) and hence to low power for our test statistics.

We investigate this last issue further in Panel B of Table 1.14 by looking at partially overlapping data taken from every fifth trading day. This sample contains 712 observations and shows a return on the High-minus-Low IRM Beta portfolio of 0.437%, which is significant at the 5% level with a t-stat of 2.2779.

1.6 Conclusion

We construct a measure of market irrationality sentiment by downloading text from the Dow Jones Newswire and calculating the proportion of words each day that describe irrational stock market behaviour in the media. We use data from the S&P500 and DJIA indices and investigate how market irrationality sentiment influences subsequent stock market returns and volatility. We further examine whether the resulting market irrationality risk factor is priced.

Performing vector autoregressions using the Dow Jones Industrial Average, the S&P 500 index, and several portfolios constructed by sorting firms on size and book-to-market ratios, we first find evidence that an increase in market irrationality sentiment is associated with a subsequent decrease in stock market returns (as proxied by the S&P 500 and DJIA stock indices as well as portfolios sorted on size and book-to-market ratio). We find that the market irrationality sentiment measure takes longer on average to fully affect stock market returns and volatility than other news-based sentiment measures and relate that to the greater complexity associated with the information it conveys which needs more time to be incorporated into stock prices. After controlling for share volume, volatility, and dummies for days-of-the-week, January, and five market crashes, we find that a one standard deviation increase of the market irrationality risk factor is associated with a 16.5 basis point drop in the S&P 500 index (significant at the 1% level), a 11.4 basis point drop on the DJIA (significant at the 10% level), a 20.4 basis point drop in the portfolio of small stocks (significant at the 5% level), a 15.2 basis point drop in the portfolio of large stocks (significant at the 1% level), a 13.6 basis point drop in the portfolio of value stocks (significant at the 5% level), and a 27.6 basis point drop in the portfolio of growth stocks (significant at the 1% level) over the following four days. We also observe that a one standard deviation increase of the market irrationality risk factor is associated with a 0.15 increase in the VIX volatility index (significant at the 5% level) over the following four days. We find no evidence that stock market returns, share volume, or volatility have any impact on the market irrationality sentiment measure. Our first main conclusion from these results is that market irrationality as expressed in the media deteriorates the subsequent investment opportunity set. We also find that the impact is immediate for more liquid firms with more illiquid stocks experiencing the impact primarily on the third lag.

The next objective is to examine whether innovations in the stock market irrationality sentiment measure represent a priced risk factor. We observe that the high-minus-low IRM beta portfolio generates a positive and significant alpha after accounting for standard risk factors (0.420% monthly in the FF 3-factor model and 0.184% monthly in the Carhart 4-factor model); this result holds in the cross-section where we find that the market irrationality risk beta coefficient is 0.574% and significant at the 1% level. We also observe that IRM continues to have a positive risk premium when controlling for volatility risk (from 0.038% monthly on mid-volatility-risk firms up to 0.315% monthly on mid-to-large-volatility-risk firms significant at the 1% level). We hypothesise that stocks that react strongly to market irrationality sentiment are also subject to high noise trader risk on the basis that noise traders may be more sensitive to irrational language in the financial news. The noise trader risk hypothesis is accompanied by a mean-reversion in stock returns, which we also observe in the data.

A primary extension to this paper would be to investigate if market irrationality reported in the media also affects the returns of other asset classes. Moreover, it would

be worthwhile to repeat the exercise by focusing on irrational words characterising stock markets that appear on the Internet, on financial blogs, or on social media. Finally, the pricing of market irrationality risk as conveyed by the media represents a theoretical challenge for standard asset pricing models that deserves to be further explored.

Table 1.1: Summary statistics of IRM

This table shows the minimum, median, maximum, mean, standard deviation, skewness, and kurtosis of IRM^{raw}, its normalisation, and its innovations starting on the 2nd January 1998 for the raw factor, the 2nd July 1998 for the normalisation, and the 7th July 1998 for the innovations and running until the 31st December 2012 inclusive for all three.

	Min	Median	Max	Mean	Std	Skew	Kurt
IRM^{raw}	0	0.308	5.755	0.445	0.509	3.556	23.704
IRM^{norm}	-1.429	-0.282	9.161	-0.004	1.009	2.816	15.382
IRM	-1.488	-0.278	9.092	0	1.008	2.813	15.406

Table 1.2: Autocorrelation of IRM$^{\text{norm}}$ and coefficients in the AR(2) process

IRM^{raw} is normalised using the prior 6 months of data. Autocorrelation and AR(2) coefficients are computed from 2nd July 1998 until the 31st December 2012 inclusive.

	ρ_1	ρ_2
IRM^{norm}	0.036	0.044
	ϕ_1	ϕ_2
Coefficients in Equation (1.1)	0.035**	0.043***

Table 1.3: Correlation between key factors

This table reports a correlation matrix of the following variables: market factor (MKT) defined as the excess market return; size factor (SMB) defined as the excess returns of small-cap stocks over large-cap stocks; value factor (HML) defined as the excess returns of the value stocks over growth stocks; momentum factor (UMD) defined as the excess returns of prior month winning stocks over losing stocks; IRM^{norm}; and IRM. The sample period is all trading days from 7th July 1998 to 31st December 2012 inclusive.

	MKT	SMB	HML	UMD	IRM$^{\text{norm}}$	IRM
MKT	1.0000					
SMB	0.0629	1.0000				
HML	-0.0953	-0.1453	1.0000			
UMD	-0.3131	0.1036	-0.2652	1.0000		
IRM$^{\text{norm}}$	-0.0112	0.0177	-0.0420	0.0092	1.0000	
IRM	-0.0091	0.0172	-0.0418	0.0093	0.9984	1.0000

Table 1.4: Summary Statistics of daily firm sample size

The table reports summary statistics of daily firm sample size: minimum, median, maximum, mean, standard deviation, skewness, and kurtosis. The data period is 2nd January 1998 - 31st December 2012.

	Min	Median	Max	Mean	Std	Skew	Kurt
# Firms	329	465	498	452.02	40.96	-1.47	4.18

Table 1.5: Coefficients of IRM^{norm} in VAR Equation (1.2). Values represent basis points.

The table reports the coefficients for the market irrationality sentiment measure in equation (1.2) when using the returns on the DJIA or S&P 500 index. It also reports the test-statistics that all coefficients are jointly 0 along with the accompanying p-values. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

	Dependent variable: R	
Irrationality	DJIA	S&P 500
IRM^{norm}_{t-1}	-3.46*	-3.43
IRM^{norm}_{t-2}	-1.86	-3.14
IRM^{norm}_{t-3}	-4.16**	-6.61***
IRM^{norm}_{t-4}	-1.87	-3.31
IRM^{norm}_{t-5}	-0.38	1.12
$\chi^2(5)$ $[Joint]$	9.92*	16.09***
p-value	0.077	0.007

Table 1.6: Coefficients of IRM^{norm} in VAR Equation (1.4)

The table reports the coefficients for the market irrationality sentiment measure in equation (1.4) when using the returns on the DJIA or S&P 500 index. It also reports the test-statistics that all coefficients are jointly 0 along with the accompanying p-values. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

R	DJIA	S&P 500
Irrationality	Dependent variable: VIX	
IRM^{norm}_{t-1}	0.060**	0.057**
IRM^{norm}_{t-2}	0.011	0.010
IRM^{norm}_{t-3}	0.069**	0.068**
IRM^{norm}_{t-4}	0.014	0.013
IRM^{norm}_{t-5}	-0.004	-0.005
$\chi^2(5)$ $[Joint]$	11.77**	11.17**
p-value	0.038	0.048

Table 1.7: Coefficients of IRM^{norm} in VAR Equation (1.2). Values represent basis points.

The table reports the coefficients for the market irrationality sentiment measure in equation (1.2) when using the returns on portfolios whose components are ranked by their size (Panel A) and book-to-market ratios (Panel B). It also reports the test-statistics that all coefficients are jointly 0 along with the accompanying p-values. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

Panel A	Dependent variable: R					
Irrationality	Small	Size Dec. 2	Size Dec. 3	Size Dec. 8	Size Dec. 9	Large
IRM_{t-1}^{norm}	-3.96	-4.52*	-4.45*	-4.53**	-4.72**	-4.97**
IRM_{t-2}^{norm}	-4.69	-3.58	-4.46*	-3.31	-2.92	-2.43
IRM_{t-3}^{norm}	-8.84***	-5.59**	-5.84**	-5.07**	-4.48**	-4.72**
IRM_{t-4}^{norm}	-2.94	-2.96	-2.46	-4.12*	-4.38*	-3.04
IRM_{t-5}^{norm}	1.58	0.93	0.34	-0.14	-1.04	0.10
$\chi^2(5)$ $[Joint]$	14.36**	12.37**	14.34**	16.37***	15.39***	15.09***
p-value	0.014	0.030	0.014	0.006	0.009	0.010
Panel B	Dependent variable: R					
Irrationality	Low	B/M Dec. 2	B/M Dec. 3	B/M Dec. 8	B/M Dec. 9	High
IRM_{t-1}^{norm}	-4.74**	-5.66**	-4.87**	-4.12*	-3.84	-5.79*
IRM_{t-2}^{norm}	-1.07	-3.46	-2.91	-3.29	-3.20	-6.41*
IRM_{t-3}^{norm}	-3.24	-3.89*	-4.97**	-6.02***	-6.14**	-10.26***
IRM_{t-4}^{norm}	-4.58**	-2.37	-2.49	-2.06	-3.59	-5.13
IRM_{t-5}^{norm}	-0.12	-1.49	-0.93	1.96	3.06	3.19
$\chi^2(5)$ $[Joint]$	11.07**	13.81**	13.81**	14.03**	13.45**	19.63***
p-value	0.050	0.017	0.017	0.015	0.020	0.002

Table 1.8: Summary Statistics of 25-Day Portfolio Excess Returns sorted on IRM

Every day from the 9th November 1998 to the 31st December 2012, using prior 63 trading days (3 months) of observations, we regress excess stock returns on the excess stock market returns and innovations in the normalised *IRM*, and stocks are assigned into ten portfolios based on the sensitivities of their excess returns to these innovations. Stocks are chosen as companies that are part of the S&P500 index or will be within one year, and that are trading Ordinary Common Shares for which CRSP has data. Stocks that appear in the S&P500 multiple times over the sample period are only included for their most recent appearance. Portfolios are held for twenty-five days and the 25-day portfolio return is calculated as the equal-weighted average of the returns of all stocks in the portfolio. The table reports summary statistics of 25-day portfolio returns: minimum, median, maximum, mean, standard deviation, skewness, kurtosis, and average exposure to the market irrationality risk factor for the 10 decile portfolios, the portfolio going long in High and short in Low, the portfolio going long in both High and Low, the portfolio going long in High and short in 6, and the portfolio going long in 5 and short in Low. The numbers in parentheses are *t*-statistics. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

IRM Beta	Min	Median	Max	Mean	Std	Skew	Kurt	IRM Beta
Low	-0.404	0.01426	0.931	0.01056	0.093	0.944	11.760	-0.00521
2	-0.375	0.01310	0.311	0.00862	0.066	-0.573	5.838	-0.00258
3	-0.342	0.01236	0.289	0.00874	0.061	-0.531	5.863	-0.00162
4	-0.322	0.01235	0.277	0.00861	0.059	-0.627	6.067	-0.00094
5	-0.321	0.01331	0.295	0.00943	0.059	-0.488	6.339	-0.00035
6	-0.322	0.01439	0.369	0.00992	0.059	-0.390	6.756	0.00021
7	-0.331	0.01521	0.390	0.01062	0.060	-0.340	7.163	0.00082
8	-0.317	0.01623	0.390	0.01145	0.063	-0.359	6.794	0.00153
9	-0.323	0.01692	0.329	0.01153	0.067	-0.502	5.730	0.00254
High	-0.457	0.01797	0.422	0.01475	0.084	-0.206	5.646	0.00522
High - Low				0.00420***				0.01043***
t-statistic				(4.6485)				(92.9809)
High + Low				0.02531***				0.00002
t-statistic				(8.9292)				(0.8616)
High - 6				0.00484***				0.00501***
t-statistic				(6.5663)				(85.2323)
5 - Low				-0.00113				0.00485***
t-statistic				(-1.2889)				(96.7723)

Table 1.9: Time-Series Tests of Three- and Four-Factor Models of Equal-Weighted Portfolios sorted on IRM

Every day from the 9th November 1998 to the 31st December 2012, using prior 63 trading days (3 months) of observations, we regress excess stock returns on the excess market returns and innovations in the normalised *IRM*, and stocks are assigned into ten portfolios based on the sensitivities of their excess returns to these innovations. Stocks are chosen as companies that are part of the S&P500 index or will be within one year, and that are trading Ordinary Common Shares for which CRSP has data. Stocks that appear in the S&P500 multiple times over the sample period are only included for their most recent appearance. Portfolios are held for twenty-five days and the 25-day portfolio return is calculated as the equal-weighted average of the returns of all stocks in the portfolio. The table reports the evaluation results of the three- and four-factor models. The numbers in parentheses are t-statistics. Significance at the 10% and 5% levels are indicated with * and ** respectively. Values whose significance is above the 10% threshold are indicated with † and all other values are significant at the 1% level.

IRM Beta	α (%)	MKT	SMB	HML	UMD	R^2_{adj}
Low	0.324	1.440	-0.067	0.511		0.7796
	(4.34)	(107.22)	(-3.04)	(26.86)		
	0.699	1.184	0.066	0.320	-0.456	0.8591
	(11.62)	(97.36)	(3.74)	(20.25)	(-44.78)	
2	0.325	1.070	-0.098	0.428		0.8500
	(7.43)	(135.81)	(-7.62)	(38.40)		
	0.473	0.969	-0.046	0.353	-0.179	0.8743
	(11.69)	(118.67)	(-3.82)	(33.30)	(-26.23)	
3	0.386	0.997	-0.134	0.420		0.8616
	(9.94)	(142.64)	(-11.73)	(42.45)		
	0.505	0.916	-0.092	0.360	-0.145	0.8802
	(13.85)	(124.30)	(-8.51)	(37.58)	(-23.45)	
4	0.392	0.963	-0.136	0.407		0.8643
	(10.59)	(144.35)	(-12.48)	(43.07)		
	0.497	0.892	-0.099	0.353	-0.127	0.8798
	(14.12)	(125.35)	(-9.49)	(38.27)	(-21.41)	
5	0.470	0.962	-0.133	0.416		0.8787
	(13.53)	(153.68)	(-13.06)	(46.93)		
	0.564	0.898	-0.100	0.368	-0.114	0.8913
	(16.97)	(133.78)	(-10.21)	(42.24)	(-20.25)	
6	0.507	0.973	-0.107	0.399		0.8809
	(14.59)	(155.28)	(-10.45)	(45.00)		
	0.605	0.907	-0.072	0.350	-0.118	0.8942
	(18.28)	(135.53)	(-7.39)	(40.25)	(-21.17)	
7	0.557	0.980	-0.053	0.377		0.8799
	(15.78)	(153.98)	(-5.07)	(41.84)		
	0.656	0.912	-0.018*	0.327	-0.120	0.8933
	(19.52)	(134.29)	(-1.77)	(37.03)	(-21.10)	
8	0.603	1.029	$-0.011^{\dagger}$	0.369		0.8807
	(16.27)	(154.02)	(-1.03)	(39.05)		
	0.704	0.960	0.025**	0.318	-0.123	0.8933
	(19.89)	(134.15)	(2.35)	(34.22)	(-20.47)	
9	0.564	1.098	0.050	0.342		0.8848
	(14.45)	(156.24)	(4.38)	(34.33)		
	0.645	1.043	0.079	0.300	-0.099	0.8919
	(16.90)	(135.21)	(7.02)	(29.98)	(-15.30)	
High	0.744	1.309	0.263	0.211		0.8565
	(13.69)	(133.71)	(16.46)	(15.25)		
	0.884	1.214	0.313	0.140	-0.170	0.8700
	(16.92)	(115.00)	(20.27)	(10.23)	(-19.23)	
High-Low	0.420	-0.131	0.330	-0.299		0.1137
	(4.86)	(-8.40)	(12.96)	(-13.57)		
	0.184**	0.030*	0.246	-0.180	0.286	0.2069
	(2.23)	(1.77)	(10.09)	(-8.28)	(20.47)	

Table 1.10: Mean 25-Day Portfolio Excess Returns in % sorted by Size and IRM

Every day from the 9th November 1998 to the 31st December 2012, using prior 63 trading days (3 months) of observations, we regress excess stock returns on the excess market returns and innovations in the normalised *IRM*, and stocks are assigned into five portfolios based on the sensitivities of their excess returns to innovations in the normalised *IRM*. Stocks are chosen as companies that are part of the S&P500 index are will be within one year, and that are trading Ordinary Common Shares for which CRSP has data. Stocks that appear in the S&P500 multiple times over the sample period are only included for their most recent appearance. In each sensitivity quintile, stocks are assigned into five further portfolios based on their market capitalisations (Panel A), book-to-market ratios (Panel B), or risk exposure to the FVIX risk factor (Panel C, Ang et al., 2006), updated every twenty-five trading days over the sample period. Portfolios are held for twenty-five days and the 25-day portfolio return is calculated as the equal-weighted average of the returns of all stocks in the portfolio. This table reports mean 25-day portfolio returns. The numbers in parentheses are *t*-statistics. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

Panel A	Size				
IRM Beta	Small	2	3	4	Large
Low	2.376	1.347	0.478	0.194	0.198
2	1.697	1.020	0.900	0.601	0.475
3	1.619	1.270	0.946	0.735	0.441
4	1.789	1.281	0.977	0.767	0.594
High	2.333	1.437	1.330	0.987	0.468
High-Low	-0.043	0.090	0.852***	0.793***	0.270***
t-statistic	(-0.4182)	(1.1307)	(9.7977)	(9.8989)	(3.1637)
Panel B	Book-to-Market Ratio				
IRM Beta	Low	2	3	4	High
Low	0.797	0.672	0.678	0.929	1.648
2	0.913	0.723	0.820	0.980	1.211
3	0.881	0.944	1.032	1.104	1.125
4	0.925	0.964	0.970	1.154	1.392
High	1.028	1.095	1.010	1.424	1.875
High-Low	0.231**	0.422***	0.333***	0.495***	0.227**
t-statistic	(2.431)	(4.6883)	(4.1318)	(6.9136)	(2.1770)
Panel C	FVIX Risk				
IRM Beta	Low	2	3	4	High
Low	0.538	0.898	1.149	1.024	1.591
2	0.552	0.740	0.946	1.002	1.255
3	0.601	0.814	0.905	1.033	1.452
4	0.814	0.907	1.021	1.046	1.464
High	0.812	1.140	1.187	1.339	1.796
High-Low	0.274***	0.242***	0.038	0.315***	0.205**
t-statistic	(2.7905)	(3.3180)	(0.5860)	(4.5539)	(2.0457)

Table 1.11: Mean Holding Period Portfolio Returns in % for different holding periods sorted on IRM

Every day from the 9th November 1998 to the 31st December 2012, using prior 63 trading days (3 months) of observations, we regress excess stock returns on the excess market returns and innovations in the normalised *IRM*, and stocks are assigned into ten portfolios based on the sensitivities of their excess returns to these innovations. Stocks are chosen as companies that are part of the S&P500 index or will be within one year, and that are trading Ordinary Common Shares for which CRSP has data. Stocks that appear in the S&P500 multiple times over the sample period are only included for their most recent appearance. Portfolios are held for a set number of days and the multi-day portfolio return is calculated as the equal-weighted average of the returns of all stocks in the portfolio. The table reports mean portfolio returns. The numbers in parentheses are *t*-statistics. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

	Holding Period (days)			
IRM Beta	5	10	15	20
Low	0.264	0.473	0.670	0.865
2	0.200	0.371	0.549	0.730
3	0.213	0.411	0.604	0.738
4	0.198	0.388	0.525	0.703
5	0.212	0.398	0.578	0.760
6	0.220	0.431	0.612	0.807
7	0.219	0.417	0.622	0.839
8	0.232	0.465	0.697	0.921
9	0.235	0.480	0.732	0.959
High	0.263	0.538	0.837	1.156
High-Low	-0.002	0.065	0.166**	0.291***
t-statistic	(-0.0330)	(1.1186)	(2.4058)	(3.6225)

Table 1.12: Subsample Analysis - Portfolios sorted on IRM

Every day from the 9th November 1998 to the 31st December 2012, using prior 63 trading days (3 months) of observations, we regress excess stock returns on the excess market returns and innovations in the normalised *IRM*, and stocks are assigned into ten portfolios based on the sensitivities of their excess returns to these innovations. Stocks are chosen as companies that are part of the S&P500 index or will be within one year, and that are trading Ordinary Common Shares for which CRSP has data. Stocks that appear in the S&P500 multiple times over the sample period are only included for their most recent appearance. Portfolios are held for twenty-five days and the 25-day portfolio return is calculated as the equal-weighted average of the returns of all stocks in the portfolio. This table reports mean 25-day portfolio returns. The numbers in parentheses are *t*-statistics. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

	9th Nov 1998 - 5th Dec 2005	6th Dec 2005 - 31st Dec 2012
IRM Beta	Mean (%) / N=1779	Mean (%) / N=1779
Low	1.201	0.910
2	0.954	0.771
3	1.072	0.676
4	0.946	0.776
5	1.060	0.825
6	1.113	0.870
7	1.323	0.801
8	1.470	0.821
9	1.510	0.797
High	1.866	1.085
High-Low	0.664***	0.175
t-statistic	(4.5794)	(1.6341)

Table 1.13: Windsorised - Portfolios sorted on IRM

Every day from the 9th November 1998 to the 31st December 2012, using prior 63 trading days (3 months) of observations, we regress excess stock returns on the excess market returns and innovations in the normalised *IRM*, and stocks are assigned into ten portfolios based on the sensitivities of their excess returns to these innovations. Stocks are chosen as companies that are part of the S&P500 index or will be within one year, and that are trading Ordinary Common Shares for which CRSP has data. Stocks that appear in the S&P500 multiple times over the sample period are only included for their most recent appearance. Portfolios are held for twenty-five days and the 25-day portfolio return is calculated as the equal-weighted average of the returns of all stocks in the portfolio. The portfolio returns are windsorised on a year-by-year basis. This table reports mean 25-day portfolio returns. The numbers in parentheses are *t*-statistics. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

	1%	2%
IRM Beta	Mean (%)	Mean (%)
Low	0.935	0.796
2	0.883	0.840
3	0.910	0.856
4	0.912	0.862
5	0.969	0.946
6	0.953	0.919
7	1.096	1.041
8	1.166	1.087
9	1.144	1.099
High	1.396	1.290
High-Low	0.461***	0.494***
t-statistic	(5.4878)	(6.1340)

Table 1.14: Summary Statistics of 25-Day Portfolio Excess Returns sorted on IRM with non-overlapping data

Every day from the 9th November 1998 to the 31st December 2012, using prior 63 trading days (3 months) of observations, we regress excess stock returns on the excess market returns and innovations in the normalised *IRM*, and stocks are assigned into ten portfolios based on the sensitivities of their excess returns to these innovations. Stocks are chosen as companies that are part of the S&P500 index or will be within one year, and that are trading Ordinary Common Shares for which CRSP has data. Stocks that appear in the S&P500 multiple times over the sample period are only included for their most recent appearance. Portfolios are held for twenty-five days and the 25-day portfolio return is calculated as the equal-weighted average of the returns of all stocks in the portfolio. The table reports summary statistics of 25-day portfolio returns: minimum, median, maximum, mean, standard deviation, skewness, and kurtosis. The number in parentheses is the t-statistic. Non-overlapping (Panel A) and partially overlapping data (Panel B) reduce the number of observations from 3558 to 143 and 712 respectively. Significance at the 10%, 5%, and 1% levels are indicated with *, **, and *** respectively.

Panel A	Non-overlapping - 143 observations						
IRM Beta	Min	Median	Max	Mean	Std	Skew	Kurt
Low	-0.307	0.01575	0.320	0.01115	0.092	0.185	4.881
2	-0.253	0.01447	0.214	0.00992	0.064	-0.468	4.884
3	-0.251	0.01449	0.204	0.00988	0.059	-0.614	5.460
4	-0.252	0.01198	0.170	0.00694	0.059	-0.596	5.322
5	-0.239	0.01493	0.174	0.01121	0.057	-0.435	5.159
6	-0.199	0.01529	0.161	0.01010	0.056	-0.384	4.560
7	-0.271	0.01656	0.150	0.01014	0.057	-0.916	6.501
8	-0.226	0.01264	0.197	0.01098	0.061	-0.428	4.746
9	-0.260	0.01403	0.188	0.01181	0.068	-0.515	4.652
High	-0.351	0.01676	0.236	0.01567	0.085	-0.446	5.436
High - Low				0.00451			
t-statistic				(1.0615)			
Panel B	Partial overlapping - 712 observations						
IRM Beta	Min	Median	Max	Mean	Std	Skew	Kurt
Low	-0.307	0.01524	0.610	0.01053	0.090	0.669	8.284
2	-0.258	0.01357	0.258	0.00843	0.065	-0.472	4.811
3	-0.251	0.01284	0.208	0.00852	0.060	-0.493	4.830
4	-0.252	0.01242	0.198	0.00872	0.058	-0.510	4.983
5	-0.239	0.01339	0.278	0.00981	0.058	-0.348	5.522
6	-0.214	0.01482	0.259	0.00990	0.058	-0.280	5.123
7	-0.271	0.01470	0.270	0.01041	0.058	-0.364	5.693
8	-0.258	0.01611	0.305	0.01116	0.062	-0.331	5.175
9	-0.260	0.01580	0.262	0.01132	0.066	-0.448	4.972
High	-0.351	0.01791	0.327	0.01490	0.083	-0.151	4.740
High - Low				0.00437**			
t-statistic				(2.2779)			

Chapter 2

Introduction of the Aesthetic Heuristic in Analysing Money-sharing Experiments

Abstract. Simple money-sharing games have long been used in economics to study participants' utility preferences in areas like altruism and inequality aversion. However, we regularly see behaviour that appears to be irrational or contradictory to standard game theoretic assumptions. Based on data from existing money-sharing games, I hypothesise that, on average, people show a preference for ratios that, in music, would be considered pleasant or appealing. I conduct an experiment to find which musical ratios are preferred and run logit models on existing money-sharing data to see if these ratios are selected for over alternatives. I find this new factor significantly increases the fit of the model, however, in roughly 50% of cases studied, swapping my factor for a simpler factor provides a better fit. I also find that, when the two factors are combined, they convey more predictive power than either of the factors alone, however, in most cases, this is not a significant difference.

JEL Classification: G02, H30

Keywords: Ultimatum game, Heuristic, Music, Aesthetic, Behavioural economics.

2.1 Introduction

Cognitive load refers to the available capacity for the brain to make calculations and a healthy brain will usually attempt to make approximately optimal decisions while minimising cognitive load. This would not only save cognitive resources, but energy as well. To achieve this, it is supposed that the brain relies on simple heuristics that it can employ in cases where calculations may be particularly complex and heuristics are able to be applied. These may include making similar decisions to those made by others or looking at superficial characteristics that serve only to remove certain options from the decision-making process.

The ultimatum game is a frequently-used, economic game designed to test certain behavioural characteristics such as altruism and envy. However, in some instances of this game, the number of options available to the proposer would be cognitively overwhelming were it not possible to simplify the decision. There are some existing heuristics that help

explain behaviour in these games but I look at an alternative, as-yet-unexplored heuristic to see if it performs better.

The ultimatum and dictator games are special in experimental economics because they automatically generate fund-splitting ratios between the proposer and responder. I examine if it's possible to construct a heuristic based on another field that features ratios: harmonics.

The possibility that the two might be connected came from the results of Fehr and Gächter, 1999. Participants in their experiment allocated a portion of 20 tokens to donate to a public good - the rest of the 20 tokens they would keep - and I noticed that, in one set of results, there were spikes in the number of people allocating the following amounts: 8, 10, 12, and 15. In another set of results, there appeared to be spikes at 5, 8, 10, and 12. If the number of participants were low enough, you would see spikes occurring in random places, however, since the spikes here were symmetric about 10 and appeared to correlate with ratios that are traditionally considered pleasant in musical harmonics (Barker, 2010), I decided to investigate the possibility that there was a verifiable correlation between the two.

In the event of successfully establishing a connection, this heuristic has two distinct advantages that, as far as I'm aware, are not generally seen in other heuristics previously studied; that is that it both identifies individual options across the entire range of possibilities (rather than narrowing down to an enclosed subset of the entire range) and is perfectly adaptable to any collection of possible options. For example, the factor I use as a robustness check, Ten, holds the first property but not the second. Anchoring would be an example of a heuristic that holds the second property but not the first.

The ratio hypothesis also relates to the neurofinance literature that says that human brains make comparisons on a logarithmic scale (Ward, 1973, Luce and D. M. Green, 1974), that is to say that we naturally compare amounts in terms of ratios. One of the possible reasons for this is that this method for comparing amounts is computationally less expensive than doing so in a linear fashion (Dehaene, 2003).

In addition, this paper raises the question of whether it's possible to predict how people will make choices based just on their vocation. This paper specifically looks at musicians, however, it's reasonable to suppose that people with backgrounds in fine art or engineering can have unique and similarly informative decision patterns. For example, the golden ratio is a prominent ratio in the visual arts and deeply ingrained in our interactions with the world (Bejan, 2009). This paper provides a framework to generate predicative factors for this type of decision-making and is, to my knowledge, the first to do so.

In harmonics, some ratios sound pleasant (consonant) and some do not (dissonant). I conduct an experiment in which I present participants with a series of harmonics and ask them to rank them from most to least pleasant. This provides me with an ordering that is statistically significantly different from random as determined by a Shapiro-Wilk test. Using this ordering, I construct a factor that equals 1 for the most pleasant-sounding ratios, 0.5 for those that are slightly less pleasant, and 0 for all other ratios.

Using the existing results from a series of ultimatum games in other papers that satisfy a set of criteria, I perform alternative-specific conditional logit models to determine if ratios belonging to the pleasant-sounding set, hereon referred to as Aes (aesthetic), are more likely to be chosen. I find that, indeed, they are but only when the number of available options is at least 20.

I perform a robustness check wherein I include a simple heuristic/dummy variable called Ten. If N is the number of tokens available to share in an ultimatum or dictator

game and N is divisible by 10, then Ten equals 1 for every offer, x, for which $x \times 10/N$ is an integer and 0 for all other offers.

For example, if an ultimatum game has 20 tokens, Ten would equal 1 for every other option i.e. 0, 2, 4, 6, 8, 10, ..., 20. If the ultimatum game had 100 tokens, Ten would equal 1 for every tenth option: 0, 10, 20, 30, 40, 50, ..., 100.

I find that Aes and Ten both significantly predict sharing behaviour about as well as each other. Since it's easier to construct, this suggests that Ten is better heuristic to employ, however, there are examples where Ten would not be an appropriate heuristic to use, in which case one could use Aes without loss of predictive power.

When Aes and Ten are combined, I find that the combined factor is a better predictor than either of the factors alone, however, on a case-by-case basis, the combined factor is not a significantly better predictor of behaviour than Ten alone.

Section 2 provides an overview of the existing literature on harmonics, investor biases, and the ultimatum game. Section 3 provides the details of the music experiment. Section 4 provides an overview of the literature used as the source for the ultimatum game data analysed in this paper. Section 5 outlines the data used and the methods for analysis. Section 6 shows the results and section 7 concludes the paper.

2.2 Background and Literature Review

This paper takes inspiration from at least three different fields: aesthetic philosophy, cognitive science, and behavioural economics. The most important form of aesthetic philosophy in this paper is that of harmonics, which was studied as far back as the ancient Greeks. Barker, 2010 provides an overview of the Greeks' building upon the work of Ptolemy's *Harmonics* including input from Pythagoras and Aristotle.

Hirshleifer, 2001 gives a detailed summary of the cognitive biases that impair investors' ability to make rational decisions. He also summarises many of the ways researchers have tried to adapt their models in order to replicate this kind of behaviour. Early on, Hirshleifer, 2001 refers to *heuristic simplification* in which limited attention and processing capacities force individuals to focus on subsets of available information. Unconscious associations can also drive selective focus and habits help to economise on thinking. Hagen and Hammerstein, 2006 say that game theory, a mainstay of traditional economics, is ill-suited to humans because it assumes they possess computational ability and preference consistency. Basic evolutionary biology also doesn't fully apply because it fails to take account of humans' ability to adapt.

Behavioural economics has a significant effect on the analysis of ultimatum games - the data for which I use to test my hypothesis. Falk and Fischbacher, 2006 look at a theory of reciprocity in which people reward kind actions and punish selfish ones, which puts a great deal of importance on context. Gintis et al., 2003 show that this even applies when the fruits of that behaviour are unlikely to be seen. Fehr and K. M. Schmidt, 2000 present a model that incorporates elements of fairness, competition, and cooperation into the utility model. This model with inferiority aversion and inequity aversion is the one I use as the base model of utility in this paper.

Bethwaite and Tompkinson, 1996 look at what drives investors to act the way they do in ultimatum games and find that over half of participants have a concern for fairness, which exceeds those that make their decisions based on envy or altruism, while only a quarter of participants can be said to be motivated by selfishness. However, Güth et al., 2001 find that, when it's not possible to propose an equal split, people give more unfair

proposals. Ubeda, 2014 find that there are a multitude of fairness rules where selfishness and strict egalitarianism win out. She also finds that faster decisions tend to be more selfish. A modified ultimatum game in Nelissen et al., 2009 demonstrates a preference for egalitarianism.

Roch et al., 2000 propose evidence for a 2-step decision-making process wherein people first apply the equality heuristic and then, depending on how much cognitive load is available, will engage in self-seeking behaviour. Handgraaf et al., 2004 go on to say that decisions depend on self-interest and fairness and the final decision depends on the evaluability of both. Brocas and Carrillo, 2014 provides an overview of decision-making systems.

Oosterbeek et al., 2004 provide a meta-analysis of ultimatum game results and show that proposers' offers do not vary greatly across regions or countries. Things that affect the generosity of the amounts proposed include the level of respect for authority, the amount available to split, and the number of times the proposer has played the game. I source many of the papers used in my analysis from this paper.

2.3 Music Experiment

Determining whether aesthetically-pleasing ratios as experienced through music are also chosen more often when the same ratios are presented in money-sharing experiments, first one must establish which ratios are most pleasant in music. To do this, I conduct an experiment that allows me to rank the ratios.

2.3.1 The method

I start by choosing ratios that can be formed by splitting 120 because it has a large number of factors.[1] I take digital recordings of all the notes in the C major scale on the piano and taper the sound files so that they all fade out and last exactly 2 seconds each. Since most of the ratios that can be derived from a split of 120 cannot be played on a piano, I use a C as my bass note (130.8 Hz defines the frequency of the bass note in the C2 list and 261.6 Hz does the same for the C3 list), and calculate the frequency above it that my second note would have to be in order to produce the ratio I want. I select the closest note from the C major scale and modulate it until it's the correct frequency. I then combine the two sound files to create a single sound file featuring two notes in the correct ratio.

In the experiment, each participant is presented with 6 trials, 3 from the C2 list and 3 from the C3 list in the order C2, C3, C3, C2, C2, C3. In each trial, the participant is presented with five note-pairs drawn at random from the given list by the computer and asked to rank them from most appealing to least appealing. After an initial volume test, all 6 trials are presented one after the other, followed by a set of demographic questions including musical background and ability. Appendix I shows what the participants saw during the experiment. The participants are not told that the note-pairs are sorted into 2 lists, nor that there are 33 note-pairs in each list. The experiment lasted 10-15 minutes on average and all participants were paid a flat fee of 10 Swiss Francs. Participants were told this information before they agreed to participate and, since the selection system targeted

[1] Namely 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 30, 40, and 60. This will help with doing experiments in the future.

students, it was deemed that an average wage of 40-60 Francs per hour was sufficient for the participants to complete the experiment earnestly. A total of 187 participants were recruited at the University of Lausanne, Switzerland using the ORSEE system (Greiner et al., 2003) and all completed the experiment.

The method for generating note-pairs was carefully considered to avoid note-pairs being preferred for reasons other than their ratio. First, all note-pairs are played in the same musical key; else all note-pairs sound equally bad and the selections become effectively random. Sine waves were rejected as the source of the notes because, although the timbre of the sine waves is neutral, sine waves with a high frequency are extremely unpleasant and risked note-pairs including high frequencies being disregarded for that reason only. Therefore, the decision to use a piano to generate the notes was made.

I made sure all note-pairs had identical bass notes because early testers preferred note-pairs with a lower bass note and based their decision more on that characteristic than the ratio of the note-pair. On the contrary, if the same bass note was used throughout the experiment, this became aggravating to testers and significantly affected the results from later trials. Having two lists also allows larger ratios to be analysed, however, it creates a problem of comparison between lists. For this reason, I kept the number of lists limited to two since, at least, it solved the problem of repeated bass notes while minimising any issues with cross-list comparisons.

The C3 list comprises 33 note-pairs that use C3 (261.6 Hz) as the bass note and includes all ratios from 60:60 (261.6 Hz : 261.6 Hz) to 92:28 (859.5 Hz : 261.6 Hz). The C2 list comprises 33 note-pairs that use C2 (130.8 Hz) as the bass note and includes all ratios from 80:40 (261.6 Hz : 130.8 Hz) to 112:8 (1831.2 Hz : 130.8 Hz). As the ratios diverge further from 1 ($\equiv$ 60:60), the increase in frequency of the uppermost note from one ratio to the next becomes greater than exponential so I must find a reasonable place to stop. 112:8 was chosen because frequencies higher than those found in this region start to get difficult to hear and 112:8 is more likely to appear in a real ultimatum game experiment than either 111:9 or 113:7.

There is an overlap of 13 note-pairs that appear in both so some cross-comparisons can be made but some ambiguity still exists by necessity.

2.3.2 Participant demographics

For the purpose of robustness, I start by ignoring all the data generated by participants that took less than 16 seconds to complete the final trial in the experiment. This is a very fast time and suggests that the participant may not have been taking the experiment seriously. I also ignore the data of any participant that failed to follow the instructions stated at the beginning of the experiment before attempting to continue. This removes 32 participants and leaves 155.

The demographic split of the 155 participants and the correlation between the different demographics can be found in Tables 2.1 and 2.2 respectively.

[Insert Tables 2.1 and 2.2]

The experiment took place in a university in a French-speaking region, hence the predominance of French-speakers aged 18-25. The split for the other demographics is close to half if we compare ‘Science’ with ‘Non-science’ and ‘No musical experience’ with ‘Some musical experience’. I also find that the median time to complete one trial (averaged over all 6 trials) is about 40 seconds.

The strongest correlation in the demographics is between people with musical experience and those with musical family members (0.377). I also find that people with musical experience are slightly more likely to take their time completing the experiment.

2.3.3 Experiment Results

I perform a statistical test to see if the results I find in the music experiment could have been created at random. In the experiment, when the participant ranks the 5 note-pairs, the note-pairs are given the scores +2, +1, 0, -1, -2 in order from 'most appealing' to 'least appealing'. If the answers are selected at random, then when all the scores are added together and the 33 note-pairs are taken together as the distribution of a random variable, then it should not be significantly different from normal with mean 0 and variance $2 \times (N \times 3) \times (5/33) = N \times 10/11$ where N is the number of participants.

I use a Shapiro-Wilk test (Shapiro and Wilk, 1965) to check for normality and find that, in general, the rankings come out as significantly different random for only two demographics: people with musical training and, at a weaker significance threshold, men. These results are independent from each other since the correlation between the two demographics is only 0.002.

Tables 2.3 to 2.6 show what the most appealing note-pairs in the C2 and C3 lists were and shows the p-value of the Shapiro-Wilk test. Unsurprisingly, the group with the most consistent ranking of the note-pairs is the group with at least 1 year of musical training with the C3 list significant at the 5% level and the C2 list significant at the 1% level. This implies that, should it be that ratios that are musically appealing are selected with greater probability in money-sharing games, musicians are more likely to be susceptible to this heuristic than non-musicians.

[Insert Tables 2.3, 2.4, 2.5, and 2.6]

Regardless of statistical power, Table 2.7 shows the six ratios that appear in the top six for every demographic in the C2 list in ascending order.

[Insert Table 2.7]

Similarly, for the C3 list, Table 2.8 shows (above the line) the same four ratios that appear in the top six for every demographic and (below the line) the three ratios from which the remaining two in the top six come.

[Insert Table 2.8]

The connection between all the note-pairs listed above is that they correspond to every octave, major third, and perfect fifth that appear in these lists. This indicates that these are the kinds of ratios I should be focussing on in my analysis. To formally construct my aesthetic utility factor, I use the rankings produced by the combination of all 155 participants, which you can see in Table 2.9. Both lists are significantly different from random at at least the 10% level, so I'm confident that this ranking is representative of the ranking of these note-pairs in general.

[Insert Table 2.9]

2.3.4 Within-subject analysis

The music experiment tests to see if subjects agree with each other in regard to which ratios they prefer the most. If the Shapiro-Wilk test returns a p-value larger than 0.1, one possibility is that the ratios were sorted at random; an alternative possibility is that even though the preferences of participants were individually consistent, their preferences were idiosyncratic, which, when all the responses were combined, would make the overall results appear less consistent.

To test for this possibility, I look for participants for whom their answers can be checked for internal consistency. This could be as straight-forward as one participant being given the same ratios twice and checking to see if they rank the two ratios in the same way both times. Alternatively, a chain of responses could be checked: for example, if they prefer A to B in one trial, B to C in another, C to D in another, and D to A in a fourth trial, then this would be considered an inconsistent preference ordering.

I check all 187 participants to find whose answers can be checked for internal consistency up to a chain of 7 comparisons. Since there are 13 ratios that appear in both lists C2 and C3, it is possible to check consistency across both lists. Here is a summary of the full algorithm:

- If a pair of ratios is ranked one way round in one trial and the other way round in another trial, the within-subject choices are deemed to be inconsistent.

- If ratio A is preferred to ratio B in one trial, ratio B is preferred to ratio C in a different trial, and ratio C is preferred to ratio A in a third trial different from the other two, then the within-subject choices are deemed to be inconsistent.

- This logic is extended to chains of 4, 5, 6, and 7 comparisons of ratios. I will demonstrate the extended logic with the chain of 7 comparisons:

  - If ratio A is preferred to ratio B,... ratio F is preferred to ratio G, and ratio G is preferred to ratio A, then the within-subject choices are deemed to be inconsistent.
  - Adjacent comparisons must appear in different trials, however, non-adjacent comparisons can appear in the same trials (the comparison of G to A is considered adjacent to the comparison of A to B). For example, the comparison of B to C can appear in the same trial as that of D to E, however, the comparison of C to D cannot appear with either.

- If any of the above conditions can be checked and the result is that the within-subject choices are not inconsistent, then they are deemed to be consistent. If none of the above checks can be made, then it is deemed impossible to check the within-subject consistency.

I find that it is possible to check the internal consistency of 48 participants of which 36 were internally consistent and 12 were not. Table 2.10 details the demographics of these 48 participants.

[Insert Table 2.10]

In general, I find very little difference in the demographics between the participants whose answers were consistent and those whose answers were inconsistent. In any case, the numbers are too low to do a proper analysis. The main reason for this is that each participant only heard at most 15 ratios out of 33 from each list and those for whose answers it was possible to check for consistency must have, by definition, heard fewer than that. This means overlaps were infrequent enough that within-subject consistency becomes a moot point.

This is a result of the design of the experiment. In an effort to avoid a learning effect, participants' exposure to the experiment was limited. Its design was such that it was possible to establish whether answers were made at random for all the participants as a whole, which it does adequately. Future experiments should be designed to also check for consistency among individual participants if possible.

2.4 Review of ultimatum and dictator game data source literature

Here, I provide a short summary of each of the 14 papers I used to test to see if *Aes* has predictive power in the ultimatum or dictator game. All of these were found in the meta-analysis made by Oosterbeek et al., 2004 and each one was selected because it was freely available online, its data could be extracted in its entirety from the paper itself, the ultimatum or dictator game performed in the analysis was of the standard form with no complicating factors, and there were enough data points per treatment that it was reasonable to fit a logit model.

Each summary is preceded by the title 'Paper X' where X is a number. This is how I'll refer to these papers later.

Paper 1

Anderson et al., 2000 conduct ultimatum games on groups in the United States and in Honduras to see how cultural differences affect attitudes toward bargaining. They conduct 2 treatments with students from each country - playing an ultimatum game and then a dictator game, or vice versa. Each game consisted of 10 rounds.

They find that, in the US, playing the ultimatum game first generated lower offers in the dictator game whereas no game order effect was found in Honduras. Offers were, in general, higher in Honduras than in the United States.

Anderson et al., 2000 provide data for the 1st and 10th rounds for both countries, both treatments, and both games in that treatment. That makes a total of 16 treatments of 11 options analysed in this paper.

Paper 2

Andreoni et al., 2003 look at what happens when, instead of just being able to accept or reject the offer, the responder can choose the amount that gets split depending on the proposer's offer. The authors call this the convex ultimatum game.

They conclude that the responses in the convex ultimatum game can be justified using a simple utility function that is continuous, convex, regular, but not monotonic. They

also find a large range of preferences between different participants, however, altruism and selfishness are both highly robust characteristics.

Andreoni et al., 2003 provide data for the standard and convex ultimatum games. That makes a total of 2 treatments of 11 options analysed in this paper.

Paper 3

Bornstein and Yaniv, 1998 look at how offers in the ultimatum game change if proposers make decisions in groups of three rather than just on their own. They find that groups are more likely to demand more money but are also more likely to accept lower offers.

Bornstein and Yaniv, 1998 provide data for the individual and group games. That makes a total of 2 treatments of 101 options analysed in this paper.

Paper 4

Cameron, 1999 attempts to address the issue that participants in other studies are not incentivised to perform as they would in real life by being offered small or hypothetical payoffs. She conducts an ultimatum game experiment in Indonesia where it's possible for the experimenters to offer significantly large amounts of money to the participants.

The experimenters conduct 4 games, each with two rounds, three using real money and one using hypothetical payoffs. She finds that systematic deviations from game-theoretic optima persist when the stakes are high, however, responders are more likely to accept low offers when the total amount on offer is higher.

Cameron, 1999 provides data for all four games and for both rounds of each game. That makes a total of 8 treatments of 21 options analysed in this paper.

Paper 5

Eckel and Grossman, 2001 conduct ultimatum games to see if men and women respond differently if receiving an offer from either a man or a woman. They find that women are, on average, more generous than men and almost never fail to reach an agreement when paired with another woman. Men are more likely to accept an offer if it's made by a woman.

Eckel and Grossman, 2001 provide data for male and female proposers. They also provide data for black and non-black proposers, however, since this is the same dataset with different groupings, I only use the data sorted by gender. That makes a total of 2 treatments of 11 options analysed in this paper.

Paper 6

Ellingsen and Johannesson, 2001 produce a model whereby a proposer incurs a cost when he or she decides to make an offer. This produces the possibility that, in the ultimatum game, the proposer rationally decides not to make an offer in the first place. They conduct treatments in which there is either symmetric or asymmetric information.

They find that when costs are known to be high or low, in general responders act fair-mindedly, however, when information is asymmetric, low-cost proposers act more aggressively and high-cost proposers less aggressively.

Ellingsen and Johannesson, 2001 provide data for high-cost and low-cost proposers with either symmetric or asymmetric information. That makes a total of 4 treatments of 21 options analysed in this paper.

Paper 7

Fershtman and Gneezy, 2001 conduct an ultimatum game in which participants can or must delegate responsibility to someone else. They find that using a delegate increases the participant's share unless the delegate is not being observed, in which case, it decreases it.

Fershtman and Gneezy, 2001 provide data for proposers and responders in cases where they used a delegate or not and in the cases where the delegate could be observed or not observed. That makes a total of 8 treatments of 11 options analysed in this paper.

Paper 8

Forsythe et al., 1994 run ultimatum and dictator games with either real (pay) or hypothetical money (no pay). All participants receive a $3 turn-up fee but participants in the pay treatments can receive an additional $5 or $10. The authors test to see if participants aim for fairness under which case results for the ultimatum and dictator games should be the same. They reject this hypothesis for the pay treatment at least.

Forsythe et al., 1994 provide data for the ultimatum and dictator games under the pay and no pay treatments. Like the authors, I pool the results for the games performed in April and in September as the time of year is the only thing that distinguishes the two. I also only use the data for the $5 experiments due to a lack of observations in the $10 experiments. That makes a total of 4 treatments of 101 options analysed in this paper.

Paper 9

Güth et al., 1997 examine if experiencing responder competition (RC) or a random responder (RR) affects the proposer's approach to fairness in a subsequent ultimatum game. In each treatment, 5 responders give the lower limit for offers that they would accept. For RC, the responder with the highest limit is selected to respond to the proposer; for RR, one of the 5 responders is chosen at random.

The authors find that participants get used to unfair offers and adjust their behaviour accordingly. They suggest that the asymmetry in behaviour - reliable fairness of proposers and willingness of responders to accept unfair offers, at least in the RC case - can be partly explained by false expectations.

Güth et al., 1997 provide data for the RR and RC treatments and a control experiment. In order to ensure an adequate number of observations, I ignore the control experiment and pool data across rounds in the RR and RC treatments. That makes a total of 2 treatments of 51 options analysed in this paper.

Paper 10

Hoffman et al., 1994 ask whether a sense of 'fairness' leads opening offers in ultimatum and dictator games to be 'too high'. They set up an experiment whereby participants earn the right to be the first mover by being the highest scorer in a general knowledge

quiz. They found that first movers in this game tended to be more self-regarding than in the control.

Hoffman et al., 1994 provide data on ultimatum and dictator game control experiments using the methodology from Forsythe et al., 1994, and treatments using either the ultimatum or dictator game, a random or earned entitlement to be first mover, as well as framing the economic game in the standard way or as an exchange. That makes a total of 10 treatments of 11 options analysed in this paper.

Paper 11

Ruffle, 1998 investigates the idea of tipping by letting the responder determine the amount to be divided through a test of skill. The proposer can then choose to reward or punish the responder accordingly.

They find that, in the dictator game, proposers reward skilled responders but only mildly punish unskilled responders. Unskilled responders nevertheless think they deserve more than they receive. In the ultimatum game, offers tend toward fairness over strategy and higher offers eliminate the significance of punishment for unskilled responders and mitigate rewards for skilled responders.

Ruffle, 1998 provides data on ultimatum and dictator games for the large pot size ($10) and small pot size ($4). The treatments depend on whether the amount to be divided is determined by responder skill or by a coin toss, and whether the amount is real (proposers only) or hypothetical (proposers and responders). I do not consider the data for the ultimatum games because they are not precise and I ignore the data for the $4 games because I do not deem 5 options to be enough for a proper analysis of the impact of Aes. That makes a total of 6 treatments of 11 options analysed in this paper.

Paper 12

Schotter et al., 1996 look at whether perceived unfairness may be fair in the context of economic survival. They do this by conducting ultimatum and dictator games - sometimes with one stage and sometimes with two stages where admission to the second stage requires success in the first stage.

They find that the hypothesis of more brutal decisions made as a result of the need to survive is founded for the dictator game but not clear for the ultimatum game. They find, however, that offers in the first stage of the ultimatum game are more likely to tend toward equity if the game is played with simultaneous moves rather than with the more common sequential moves.

Schotter et al., 1996 provide data on the one-stage and two-stage ultimatum and dictator games including simultaneously and sequentially played versions of the ultimatum game. Data from the second stage of the two-stage games were ignored due to a lack of observations. That makes a total of 6 treatments of 201 options analysed in this paper.

Paper 13

Slembeck, 1999 investigates the role of reputation in ultimatum games by making participants play against the same opponent repeatedly. He conducts ultimatum games comprising 20 rounds with either fixed pairs or rotating pairs. He finds that about half of pairs conform to the learned norm hypothesis by converging to an utility-optimising

equilibrium. The other half engage in costly battles that lowers income. Participants in fixed pairs earn about 12% less than those in rotating pairs because 'fair' offers are rejected more frequently.

Slembeck, 1999 provides precise data for the fixed pairs only for all 20 rounds. To avoid an excess of data, I analyse the results for the first and last rounds only. That makes a total of 2 treatments of 101 options analysed in this paper.

Paper 14

Weg and V. Smith, 1993 see if making a participant weaker or stronger than they would be in a standard ultimatum game can elicit more extreme responses than in the standard game. They do this by setting up ultimatum games in which, if the responder rejects the first offer, they may propose a counter-offer to which the original proposer must respond, or the proposer gets to make a second proposal. They also conduct a control experiment.

They find that proposers offer larger amounts the weaker they are, however, offers are not extreme in the manner the authors predicted. This supports the tendency toward fairness in the ultimatum game seen across the literature.

Weg and V. Smith, 1993 provide precise data on the first offers from all types of games, each of which are played twice by each participant. That makes a total of 6 treatments of 11 options analysed in this paper.

2.5 Economic Data and Analysis Methods

In this section, I use the rankings derived from the music experiment to construct *Aes*. I then use existing data from previous ultimatum and dictator game experiments and run alternative-specific conditional logit models using a base utility model from Fehr and K. M. Schmidt, 2000 in conjunction with *Aes*. The goal is to see if *Aes* has predictive power in money-sharing experiments.

I use the results of the ultimatum and dictator games conducted by the independent studies detailed in Section 2.4. For each treatment in each paper, I conduct an alternative-specific conditional logit model based on the options available. I assume the existence of a base utility function modelled on that used by Fehr and K. M. Schmidt, 2000. This has the form

$$U(x;y) = x - \alpha \max(x-y,0) - \beta \times \mathbf{1}\{y > x\} \tag{2.1}$$

where α measures the level of inequity aversion, β measures the level of inferiority aversion, x is the amount the proposer proposes for themselves, and y is the amount the proposer proposes for the receiver.[2] The original form of this utility function in Fehr and K. M. Schmidt, 2000 is

$$U(x;y) = x - \alpha \max(x-y,0) - \beta \max(y-x,0),$$

however, when this factor is used instead of the factor for inferiority aversion seen in equation (2.1), the logit models almost never converge. Given the structure of most of the datasets analysed in this paper, I deem this change to be reasonable and appropriate.

[2]Inequity aversion refers to an individual's dislike of having more than another person. Inferiority aversion refers to an individual's dislike of having less than another person. In Tables 2.15-2.21, the coefficients listed for Ineq. Av. and Inf. Av. are equal to $-\alpha$ and $-\beta$ in equation (2.1) respectively.

In some cases, when the model doesn't converge, I force β in equation (2.1) to equal 0. If the model still does not converge, I do not include the results.

The choice of this base utility function is such that it is simple enough to be easily applicable to a wide range of different, independently-conducted experimental datasets, while not being so simple that it doesn't provide adequately robust results. Based on the large body of literature suggesting that fairness is one of the single most important concerns for the majority of ultimatum and dictator game participants (see Section 2.2), I believe that this base utility function is the simplest way of generating adequately robust results.

In the first analysis, I add the aesthetic utility factor (Aes) to the utility function. Aes is created by looking at the ranking in Table 2.9 and assigning the different options the values of +1, +0.5, or 0 depending on where in the ranking they appear. Tables 2.11 and 2.12 demonstrate what this looks like for economic games using 10 and 20 tokens respectively. They also show the rankings of the different options in the C2 and C3 lists so you're able to see how the rankings translate into the Aes factor.

[Insert Tables 2.11 and 2.12]

As a robustness check, in models where it's appropriate, I include an additional, simpler heuristic, Ten, in which only options at periodic intervals are considered. More specifically, if the amount to be shared can be divided by 10, then there exists a subset of options, x, for which $x \times 10/N$ is an integer. So, if the amount to be shared is 100, I include a dummy variable that is 1 if the splits are 100:0, 90:10, 80:20, 70:30, etc and 0 for any other split.

This heuristic is simpler than the aesthetic heuristic, however, the aesthetic heuristic has a few advantages: first, it's application is independent of the amount being split and secondly, it's roots are more subconscious and so would therefore rely less on the visual parts of the brain as the options are being considered.

I perform the logit models using these utility factors on a treatment-by-treatment basis. This ensures that the models are applied to experiments like-for-like.

To disentangle the two factors, I create 3 variables: $Aes \times Ten$, 'Aes only', and 'Ten only'. The first is self-explanatory and the second two are equal to Aes and Ten respectively unless $Aes \times Ten$ is non-zero, in which case they are equal to zero.

In some cases, since the most popular ratios in the music experiment consistently related to octaves, major thirds, and perfect fifths, I also look at the model with a factor for ratios corresponding to octaves, a factor for ratios corresponding to major thirds, and a factor corresponding to perfect fifths. In the experiments analysed, this only applies to those where the amount to share is at least 100. Table 2.13 shows for which ratios the different dummy variables equal 1; they equal 0 otherwise.

[Insert Table 2.13]

I do the same treatment for these as I do for Aes and Ten. Namely, I construct the factors $Oct \times Ten$, $PF \times Ten$, 'Oct only', 'MT only', 'PF only', and 'Ten only'.[3] Several of these factors are 1 in only one place, for example, 'Oct only' is only 1 for the ratio 67:33, while $PF \times Int$ is only 1 for the ratio 60:40, however, it's an interesting test to see if the coefficients change for different kinds of ratios.

[3] $MT \times Ten$ is not included because there is no overlap between the two factors.

2.6 Results

Having run the model for all 78 treatments over 14 papers, I now compare the coefficients for Aes and Ten; in each run, the base utility model as expressed in equation (2.1) is always present. I look to see if either heuristic - Aes or Ten - is significantly different from 0 and if one is a better predictor of behaviour than the other. I also look to see if the combined factor, $Aes \times Ten$, is a strong predictor of behaviour and whether it's significantly better or worse than 'Ten only'.

Table 2.14 shows the distribution of significant results. When the number of options is low (e.g. 11 options), the Aes coefficient is neither consistently positive nor negative and is rarely significant. There's one paper (paper 2) in which the Aes coefficient is negative and highly significant in both treatments and one paper (paper 5) in which the Aes coefficient is positive and highly significant in both treatments.

When the amount to be shared is at least 20, however, the Aes coefficient becomes positive in all treatments and is significant in the majority of cases. When the amount to be shared is increased further (at least 50), the Aes coefficient becomes positive and significant in all treatments.

This shows that, when accounting for inequity aversion and inferiority aversion, Aes predicts which options are likely to be chosen, if only when the number of available options is large.

[Insert Tables 2.14, 2.15, 2.16, and 2.17]

I do the same analysis but this time using Ten in place of Aes, which can only be done for the papers where the participants are given at least 20 to share. Tables 2.15 - 2.17 show the coefficients of the logit models for all of these treatments; first when the additional factor is Aes and then when the additional factor is Ten.[4]

In the majority of treatments, the coefficient of the additional factor, either Aes or Ten, is positive and significant at the 1% level. When comparing the fit of the model, out of 28 treatments, 14 are a better fit when Aes is the additional factor and 14 are a better fit when Ten is the additional factor. Both are more reliable predictive factors of the proposed split than either inequity aversion, inferiority aversion, or the amount the proposer gets to keep, with more significant coefficients over the 28 treatments than any of the other factors.

Tables 2.18 and 2.19 show the model coefficients when both Aes and Ten are included in the form 'Aes only', 'Ten only', and $Aes \times Ten$. It also tests to see if $Aes \times Ten$ and 'Ten only' are significantly different. I find that $Aes \times Ten$ is positive and highly significant in 21 out of the 25 treatments in which the model converged. In the remaining 4 treatments, the coefficient was not significantly different from 0.

The 'Aes only' coefficient is significant in 5 out of 25 treatments and is positive in all of them; the 'Ten only' coefficient is significant in 16 out of 25 treatments and is positive in 15. When calculating the difference between $Aes \times Ten$ and 'Ten only', I find that the former is larger than the latter in 23 out of 25 treatments. The difference is significant in 7 out of 25 treatments and in all 7 treatments the $Aes \times Ten$ coefficient is larger than the 'Ten only' coefficient.

[4] If the model does not converge when all the factors are included, I remove the inferiority aversion factor. If it still doesn't converge, I do not publish the results for that model.

Tables 2.20 and 2.21 show the model coefficients when 'Oct only', 'MT only', 'PF only', 'Ten only', $Oct \times Ten$, and $PF \times Ten$ are added to the base model. Of all the 'Oct only', 'MT only', and 'PF only' coefficients, there is a total of one significant result out of the 11 treatments that converged and that is for the 'PF only' factor.

Of the 'Ten only', $Oct \times Ten$, and $PF \times Ten$ coefficients, they are collectively positive and highly significant in 6 out of 11 treatments and collectively not significantly different from zero in 3 out of 11 treatments. In the remaining 2 treatments, one of the factors is not significantly different from zero while the other 2 are positive and highly significant.

When comparing $Oct \times Ten$ to 'Ten only', I find that $Oct \times Ten$ is significantly larger than 'Ten only' in 3 out of 11 treatments; it is larger but not significantly so in 6 out of 11 treatments; it is smaller but not significantly so in the remaining 2 treatments.

When comparing $PF \times Ten$ to 'Ten only', I find that $PF \times Ten$ is significantly larger than 'Ten only' in 3 out of 11 treatments; it is larger but not significantly so in 3 out of 11 treatments; it is smaller but not significantly so in the remaining 5 treatments.

Overall, I find that neither Aes nor Ten is a significantly better predictor of behaviour than the other although both heuristics are significantly better predictors than the amount the proposer receives, inequity aversion, and inferiority aversion. I find that the combined heuristic, $Aes \times Ten$ is a better predictor of behaviour than either heuristic alone, however, the difference in predictive power between $Aes \times Ten$ and 'Ten only' is not significant.

I also split Aes into three types of musical harmonic for robustness and find that, again, the combined heuristic of Aes and Ten is a better predictor than either heuristic alone, however, the difference between that and 'Ten only' is not significant.

[Insert Tables 2.18, 2.19, 2.20, and 2.21]

2.7 Conclusion

In this paper, I construct a factor that identifies ratios that are appealing when played as musical harmonics and transpose this factor into something that can be applied to economic money-sharing games such as the ultimatum game or dictator game. I do this by conducting an experiment in which students from the University of Lausanne in Switzerland are given musical harmonics and asked to rank them from most appealing to least appealing. The factor joins a set of other factors known as heuristics: a superficial method for simplifying the choices available to us in such a way that we cut down on the amount of processing our brains have to do while simultaneously making decisions that are approximately optimal.

Identifying heuristics is a useful way to aid the predictive power of economic models and it means we can understand what elements of decisions come from superficial aspects of the choice and which ones come from a much more fundamental and economically meaningful aspect of the problem.

In this paper I look at two heuristics: an original factor based on musical harmonics (Aes) and a simpler factor that identifies options occurring at regular intervals which I call Ten. I add these to a base model featuring the amount of money the proposer gets to keep in the ultimatum or dictator game, inferiority aversion, and inequity aversion. In general, both Aes and Ten have positive and highly significant coefficients in any experiment that has at least 20 options available to the proposer. When it comes to the best fit of the model, out of 28 treatments, both heuristics prove to be the best fit in 14 of them.

One of the advantages of *Aes* over *Ten* is that it is highly adaptable and can be easily applied to cover any selection of options the proposer may be faced with. *Ten* on the other hand, can only be applied to experiments where the number of options is divisible by ten and are equally spaced in terms of return to the proposer. Thus, these results suggest that, in most standard cases, *Ten* is a better heuristic to employ because it's simpler to implement, however, when the options fit a non-standard template e.g. when the number of options is not divisible by ten or not all the possible options fit a regular, periodic structure, *Aes* would be a suitable replacement of equal predictive power.

When both heuristics are included, I find that options where *Aes* and *Ten* overlap are most likely to be chosen followed by options covered only by *Ten*, although the difference between the two is rarely significant. The coefficient for '*Aes* only' is rarely significantly different from 0.

Next, I break *Aes* into a more granular form by splitting the harmonics into octaves, major thirds, and perfect fifths - the most preferred harmonics in the music experiment. The results for this are not significantly better than those for *Aes* with the exception that the overlap between the octaves factor (*Oct*) and *Ten* is the best predictor of all the heuristics. This, however, is most likely due to the fact that this factor includes the equal 50:50 split that is disproportionately chosen in ultimatum games.

I conclude that, if you wish to fit a heuristic to a predictive model of money-sharing behaviour, the regular heuristic *Ten* is a good predictor, however, if the experimental money-sharing game is of non-standard form e.g. total amount to be shared is not divisible by ten or options are not available in a regular way, *Aes* is an appropriate substitute.

Extensions to this research could include conducting an experiment only on musicians and see if *Aes* is a much better heuristic for them than *Ten* in every regard, not just in terms of flexibility. The music experiment showed that non-musicians either did not easily identify harmonics that they found pleasant or were substantially more idiosyncratic in their preferences and this may have a significant effect on how well *Aes* predicts their behaviour in the money-sharing experiment. It's also important to see how well *Aes* performs in an experiment where the amount to be shared is not divisible by ten or where the options are not equally spaced out.

One could also investigate if other aesthetic sensibilities such as those found in painting and sculpture could have an influence on decision-making for individuals who perform those activities. An alternative method for discovering if there is some innate attraction toward certain ratios might be, for example, giving participants a large number of small objects, beads say, and asking them to split them into two piles with the minimum of restrictions. This may help reveal natural ratios that could improve the performance of the aesthetic heuristic in future experiments.

Table 2.1: Participant Demographics

Language	English	French	German	
	8	146	1	
Musical Family Members?	Yes	No		
	65	90		
Years of musical experience	0	1-2	3-5	6+
	82	13	28	32
Best subject	Arts	Science	Humanities	Other
	14	85	46	10
Age	18-25	26-35		
	149	6		
Gender	Male	Female		
	89	66		
Average trial time	< 40s	≥ 40 s		
	76	79		

This table reports the number of participants in each demographic sought by the experimenter.

Table 2.2: Demographic Correlations

	Family Musicians	Some Experience	Science	Male
Some Music Experience	0.377	1		
Science	-0.017	0.025	1	
Male	0.018	0.002	0.136	1
< 40s	-0.075	-0.176	0.060	0.036

This table reports the correlation of different demographics examined in this experiment.

Table 2.3: Family musicians

	List C3			
	Yes	$N = 90$	No	$N = 65$
Rank	Tone	Score	Tone	Score
1	90:30	1.22	86:34	1.50
2	67:53	1.09	80:40	1.34
3	72:48	1.07	67:53	1.25
4	60:60	1.07	90:30	1.24
5	80:40	0.97	72:48	1.18
6	87:33	0.67	87:33	1.00
p-value		0.054		0.275
	List C2			
Rank	Tone	Score	Tone	Score
1	86:34	1.19	96:24	1.34
2	90:30	1.02	80:40	1.32
3	96:24	0.97	103:17	1.20
4	100:20	0.97	86:34	1.10
5	80:40	0.74	100:20	1.00
6	103:17	0.69	90:30	0.93
p-value		0.388		0.165

This table reports the top 6 note-pairs in each list based on the average score given to them by participants in the music experiment sorted according to whether they reported having musicians in the family or not. In each trial, participants were given five note-pairs to sort; the note-pair they put top was given 2 points, the second was given 1 point, the third was given 0 points, the fourth was given -1 point, and the bottom was given -2 points. The p-value is the result of the Shapiro-Wilk test that tests to see if the distribution of scores is significantly different from that of a normal distribution. If the distribution is not significantly different from a normal distribution, I cannot reject the hypothesis that the note-pairs were ranked at random.

Table 2.4: Musical Experience

	List C3			
	Some	$N = 73$	None	$N = 82$
Rank	Tone	Score	Tone	Score
1	72:48	1.52	90:30	1.08
2	90:30	1.39	80:40	0.97
3	86:34	1.38	67:53	0.94
4	67:53	1.35	60:60	0.88
5	80:40	1.33	72:48	0.80
6	60:60	1.03	87:33	0.71
p-value		0.048		0.292
	List C2			
Rank	Tone	Score	Tone	Score
1	96:24	1.39	86:34	1.03
2	80:40	1.32	96:24	0.91
3	86:34	1.27	100:20	0.72
4	100:20	1.24	90:30	0.72
5	103:17	1.23	103:17	0.64
6	90:30	1.21	80:40	0.61
p-value		0.006		0.412

This table reports the top 6 note-pairs in each list based on the average score given to them by participants in the music experiment sorted according to whether they reported having some experience of learning a musical instrument or not. In each trial, participants were given five note-pairs to sort; the note-pair they put top was given 2 points, the second was given 1 point, the third was given 0 points, the fourth was given -1 point, and the bottom was given -2 points. The p-value is the result of the Shapiro-Wilk test that tests to see if the distribution of scores is significantly different from that of a normal distribution. If the distribution is not significantly different from a normal distribution, I cannot reject the hypothesis that the note-pairs were ranked at random.

Table 2.5: Gender

	List C3			
	Male	$N = 89$	Female	$N = 66$
Rank	Tone	Score	Tone	Score
1	72:48	1.23	90:30	1.27
2	90:30	1.20	67:53	1.19
3	80:40	1.20	86:34	1.19
4	67:53	1.12	80:40	1.03
5	60:60	0.91	72:48	1.00
6	86:34	0.89	60:60	1.00
p-value		0.048		0.219
	List C2			
Rank	Tone	Score	Tone	Score
1	96:24	1.41	86:34	1.18
2	90:30	1.14	103:17	1.08
3	86:34	1.13	100:20	1.03
4	80:40	1.00	80:40	0.79
5	100:20	0.94	90:30	0.75
6	103:17	0.85	96:24	0.74
p-value		0.092		0.317

This table reports the top 6 note-pairs in each list based on the average score given to them by participants in the music experiment sorted according to gender. In each trial, participants were given five note-pairs to sort; the note-pair they put top was given 2 points, the second was given 1 point, the third was given 0 points, the fourth was given -1 point, and the bottom was given -2 points. The p-value is the result of the Shapiro-Wilk test that tests to see if the distribution of scores is significantly different from that of a normal distribution. If the distribution is not significantly different from a normal distribution, I cannot reject the hypothesis that the note-pairs were ranked at random.

Table 2.6: Speed

	List C3			
	$\geq$ 40s	$N = 79$	$<$ 40s	$N = 76$
Rank	Tone	Score	Tone	Score
1	67:53	1.28	80:40	1.33
2	90:30	1.13	90:30	1.30
3	60:60	1.10	72:48	1.18
4	72:48	1.05	67:53	1.03
5	86:34	1.03	86:34	0.97
6	80:40	0.94	60:60	0.78
p-value		0.437		0.072
	List C2			
Rank	Tone	Score	Tone	Score
1	96:24	1.20	86:34	1.17
2	86:34	1.13	96:24	1.09
3	100:20	1.12	90:30	1.09
4	90:30	0.89	103:17	1.07
5	103:17	0.84	80:40	1.02
6	80:40	0.82	100:20	0.84
p-value		0.088		0.158

This table reports the top 6 note-pairs in each list based on the average score given to them by participants in the music experiment sorted according to how quickly they completed each trial on average. In each trial, participants were given five note-pairs to sort; the note-pair they put top was given 2 points, the second was given 1 point, the third was given 0 points, the fourth was given -1 point, and the bottom was given -2 points. The p-value is the result of the Shapiro-Wilk test that tests to see if the distribution of scores is significantly different from that of a normal distribution. If the distribution is not significantly different from a normal distribution, I cannot reject the hypothesis that the note-pairs were ranked at random.

Table 2.7: Top ratios in C2 list

80	40
86	34
90	30
96	24
100	20
103	17

This table shows the ratios that were ranked in the top 6 for every demographic when sorting the C2 list.

Table 2.8: Top ratios in C3 list

67	53
72	48
80	40
90	30
60	60
86	34
87	33

This table shows the ratios that were ranked in the top 6 for every demographic when sorting the C2 list. Ratios above the line were always in the top 6, ratios below the line were not.

Table 2.9: Ranking of ratios

	List C3		List C2	
Rank	Tone	Score	Tone	Score
1	90:30	1.23	86:34	1.15
2	67:53	1.15	96:24	1.15
3	80:40	1.13	90:30	0.99
4	72:48	1.11	100:20	0.98
5	86:34	1.00	103:17	0.95
6	60:60	0.95	80:40	0.92
7	87:33	0.79	108:12	0.61
8	75:45	0.69	101:19	0.49
9	68:52	0.20	105:15	0.45
10	83:37	0.18	98:22	0.41
11	69:51	0.15	106:14	0.12
12	65:55	0.14	87:33	0.07
13	74:46	0.09	99:21	0.03
14	63:57	-0.07	84:36	-0.07
15	76:44	-0.13	93:27	-0.11
16	91:29	-0.17	109:11	-0.11
17	89:31	-0.18	95:25	-0.13
18	84:36	-0.18	112:08	-0.14
19	88:32	-0.22	110:10	-0.24
20	77:43	-0.22	83:37	-0.25
21	73:47	-0.23	107:13	-0.26
22	70:50	-0.23	111:09	-0.29
23	85:35	-0.26	104:16	-0.34
24	64:56	-0.41	85:35	-0.37
25	71:49	-0.45	97:23	-0.43
26	92:28	-0.46	94:26	-0.48
27	78:42	-0.53	102:18	-0.48
28	66:54	-0.59	89:31	-0.54
29	62:58	-0.74	88:32	-0.57
30	79:41	-0.81	91:29	-0.63
31	81:39	-0.83	92:28	-0.68
32	82:38	-0.84	82:38	-0.86
33	61:59	-1.36	81:39	-0.97
p-val		0.046		0.078

This table reports the full ranking of note-pairs for all experiment participants who demonstrated that they'd properly read the instructions at the beginning and took at least 16 seconds to complete the final trial ($N = 155$). In each trial, participants were given five note-pairs to sort; the note-pair they put top was given 2 points, the second was given 1 point, the third was given 0 points, the fourth was given -1 point, and the bottom was given -2 points. The p-value is the result of the Shapiro-Wilk test that tests to see if the distribution of scores is significantly different from that of a normal distribution. If the distribution is not significantly different from a normal distribution, I cannot reject the hypothesis that the note-pairs were ranked at random. If a ratio is underlined, then it appears in both lists.

Table 2.10: Demographics of participants included in the consistency check

Total	Consistent		Inconsistent	
	36		12	
	Total	Percentage	Total	Percentage
Included in final dataset	31	86%	11	92%
French	32	89%	11	92%
Family Musicians	13	36%	5	42%
Some Musical Experience	13	36%	6	50%
Studying Science	19	53%	5	42%
Male	20	56%	6	50%
18-25	32	89%	12	100%
Median average time	40.5		38.5	
Median final time	28.5		30	

This table reports the demographic breakdown of participants whose responses in the music experiment where capable of being checked for within-subject consistency.

Table 2.11: Value of *Aes* for a game using 10 tokens

x	y	Rank in C2 list	Rank in C3 List	*Aes*
10	0	-	-	0
9	1	7	-	0.5
8	2	2	-	1
7	3	14	18	0
6	4	-	4	1
5	5	-	6	0.5
4	6	-	4	1
3	7	14	18	0
2	8	2	-	1
1	9	7	-	0.5
0	10	-	-	0

This table shows how *Aes* is constructed for an economic game with 10 tokens. Ratios ranked 1-5 are given the value 1, those ranked 6-7 are given the value 0.5; all other ratios are given 0.

Table 2.12: Value of *Aes* for a game using 20 tokens

x	y	Rank in C2 list	Rank in C3 List	*Aes*
20	0	-	-	0
19	1	-	-	0
18	2	7	-	0.5
17	3	27	-	0
16	4	2	-	1
15	5	3	1	1
14	6	14	18	0
13	7	-	27	0
12	8	-	4	1
11	9	-	28	0
10	10	-	6	0.5
9	11	-	28	0
8	12	-	4	1
7	13	-	27	0
6	14	14	18	0
5	15	3	1	1
4	16	2	-	1
3	17	27	-	0
2	18	7	-	0.5
1	19	-	-	0
0	20	-	-	0

This table shows how *Aes* is constructed for an economic game with 20 tokens. Ratios ranked 1-5 are given the value 1, those ranked 6-7 are given the value 0.5; all other ratios are given 0.

Table 2.13: Summary of dummy variables when splitting *Aes* up

Factor	Equal to 1 for ratios		
Octave (Oct)	50:50	67:33	80:20
Major Third (MT)	56:44	72:28	83:17
Perfect Fifth (PF)	60:40	75:25	86:14

This table shows the ratios for which the dummy variables Oct, MT, and PF equal one.

Table 2.14: Distribution of significant results

Number of Options		11						
	Paper	1	2	5	7	10	11	14
positive	(0.1%)	-	-	2	-	-	-	-
	(1%)	1	-	-	-	-	-	-
	(10%)	2	-	-	-	-	-	-
Insignificant		11	-	-	8	7	3	6
negative	(10%)	2	-	-	-	2	2	-
	(1%)	-	-	-	-	1	1	-
	(0.1%)	-	2	-	-	-	-	-
Min		-1.829	-1.492	1.034	-0.498	-1.696	-1.665	0.095
Max		1.962	-1.183	1.350	0.937	0.881	-0.169	0.878

Number of Options		21		51	101			201
	Paper	4	6	9	3	8	13	12
positive	(0.1%)	2	2	2	2	4	-	5
	(1%)	3	2	-	-	-	-	-
	(10%)	1	-	-	-	-	2	1
Insignificant		2	-	-	-	-	-	-
negative	(10%)	-	-	-	-	-	-	-
	(1%)	-	-	-	-	-	-	-
	(0.1%)	-	-	-	-	-	-	-
Min		0.230	1.475	1.091	2.651	2.175	1.258	1.390
Max		1.778	3.323	1.183	3.548	5.260	1.376	5.607

This table reports the level of significance and sign of the *Aes* coefficient in the logit model for every treatment studied. It also reports the maximum and minimum values of the *Aes* coefficient for every paper studied. The individual paper references can be found in Section 2.4.

Table 2.15: Comparison of *Aes* and *Ten*

Paper 4 : 21 Options

Treatment	N	Heur.	x	Inf. Av.	Ineq. Av.	Heuristic	Wald stat
1	29	*Aes*	0.769	-1.658*	-3.083*	1.342***	23.33***
		Ten	2.439	-1.022	-3.502**	1.683***	26.91***
2	29	*Aes*	6.464	-1.721	-6.055**	1.157***	29.16***
		Ten	7.975	-1.082	-6.340**	2.419***	34.92***
3	35	*Aes*	5.851	-2.147**	-7.527***	1.778***	46.70***
		Ten	7.735	-1.174	-7.227***	17.665	35.88***
4	35	*Aes*	1.128	-4.675**	-6.124*	0.230	47.93***
		Ten	1.131	-4.717**	-6.232*	-0.145	46.78***
5	37	*Aes*	-3.123	-4.042***	-0.661	1.515***	39.17***
		Ten	-0.245	-3.081**	-1.753	2.206***	42.36***
6	37	*Aes*	15.187	-2.363	-11.946	0.672	39.27***
		Ten	15.455	-2.154	-11.715	0.622	43.40***
7	40	*Aes*	-0.597	-2.820**	-1.038	0.790**	27.72***
		Ten	1.178	-2.207**	-1.807	3.485***	33.83***
8	40	*Aes*	-3.881	-4.041***	-0.138	1.029***	35.80**
		Ten	-2.040	-3.472***	-0.929	0.737**	35.48***
All	282	*Aes*	1.079	-2.890***	-3.514***	1.010***	309.20***
		Ten	2.377*	-2.370***	-3.769***	1.417***	363.25***

Paper 6 : 21 Options

Treatment	N	Heur.	x	Inf. Av.	Ineq. Av.	Heuristic	Wald stat
1	22	*Aes*	0.002	-17.892	-0.024	1.475***	13.43***
		Ten	-0.000	-17.936	-0.019	2.798***	15.47***
2	45	*Aes*	0.003	-14.689	0.056	3.323***	26.46***
		Ten	-0.000	-16.255	-0.017	1.019***	21.66***
3	59	*Aes*	0.408	1.408	-0.206	1.959***	42.84***
		Ten	0.548*	3.006	-0.280*	2.708***	31.65***
4	21	*Aes*	0.003	-16.905	0.012	1.525***	8.25*
		Ten	-0.000	-18.014	0.005	0.976*	4.33
All	147	*Aes*	0.382	0.695	-0.184	1.797***	91.68***
		Ten	0.380	0.665	-0.189	1.715***	66.03***

This table reports coefficients of the alternative-specific conditional logit model for the given paper and treatment with the base utility model combined with one heuristic: either *Aes* or *Ten*. All treatments in this table feature the ultimatum game. * means that the coefficient or stat is significant at the 10% level; ** and *** respectively the 5% and 1% levels.

Table 2.16: Comparison of *Aes* and *Ten*

Paper 9 : 51 Options

Treatment	N	Heur.	x	Inf. Av.	Ineq. Av.	Heuristic	Wald stat
1	60	*Aes*	0.296***	-0.023	-0.198***	1.091***	49.19***
		Ten	0.372***	0.806	-0.230***	2.357***	92.87***
2	72	*Aes*	1.410*	0.952	-0.725*	1.183***	31.87***
		Ten	1.273*	1.119	-0.658*	2.096***	81.19***
All	132	*Aes*	0.339***	-0.289	-0.201***	1.110***	87.54***
		Ten	0.412***	0.403	-0.236***	2.203***	182.29***

Paper 3 : 101 Options

Treatment	N	Heur.	x	Inf. Av.	Ineq. Av.	Heuristic	Wald stat
1	20	*Aes*	0.000	-18.446	-0.025	2.651***	26.97***
		Ten	-0.001	-19.194	-0.022	3.671***	45.27***
2	20	*Aes*	0.243	-0.816	-0.196**	3.548***	50.17***
		Ten	0.290*	0.386	-0.202**	3.785***	62.57***
All	40	*Aes*	0.227	-1.292	-0.155*	3.059***	81.59***
		Ten	0.287*	-0.072	-0.178**	3.756***	114.29***

Paper 8 : 101 Options

Treatment	N	Heur.	x	Inf. Av.	Ineq. Av.	Heuristic	Wald stat
1	45	*Aes*	2.410	0.030	-0.768	2.175***	46.70***
Dc		*Ten*	2.865**	-	-1.286**	19.639	11.67***
2	46	*Aes*	-1.284	-6.330**	0.087	3.594***	84.35***
Dc		*Ten*	0.000	-4.509*	-0.480	20.035	35.51***
3	43	*Aes*	3.524**	-0.881	-2.965***	5.260***	91.31***
		Ten	8.999*	2.802	-5.429**	6.779***	64.61***
4	48	*Aes*	1.998*	-1.724**	-1.964***	3.340***	113.24***
		Ten	2.840**	-0.533	-2.185***	4.401***	128.29***
All	182	*Aes*	1.986***	-1.571***	-1.429***	3.092***	314.61***
		Ten	3.300***	-0.163	-2.049***	5.658***	261.08***

This table reports coefficients of the alternative-specific conditional logit model for the given paper and treatment with the base utility model combined with one heuristic: either *Aes* or *Ten*. "Dc" denotes that the treatment featured a dictator game, otherwise, it featured an ultimatum game. * means that the coefficient or stat is significant at the 10% level; ** and *** respectively the 5% and 1% levels.

Table 2.17: Comparison of *Aes* and *Ten*

Paper 13 : 101 Options

Treatment	N	Heur.	x	Inf. Av.	Ineq. Av.	Heuristic	Wald stat
1	19	*Aes*	0.002	-19.862	0.022	1.376**	6.48
		Ten	0.000	-18.300	0.035	2.338***	25.30***
2	19	*Aes*	0.941*	-1.149	-0.956***	1.258**	27.93***
		Ten	1.091*	-0.515	-0.940***	2.351***	49.41***
All	38	*Aes*	0.943*	-1.134	-0.630**	1.288***	32.59***
		Ten	1.093*	-0.637	-0.670**	2.363***	73.66***

Paper 12 : 201 Options

Treatment	N	Heur.	x	Inf. Av.	Ineq. Av.	Heuristic	Wald stat
1	15	*Aes*	0.003	-17.525	-0.212	2.329***	22.73***
Dc		*Ten*	0.394	-15.133	-0.375	32.505	4.46
2	14	*Aes*	0.001	-15.431	0.199	1.390**	5.97
Dc		*Ten*	0.569	-15.653	-0.171	31.525	1.92
3	17	*Aes*	1.300	-1.376	-1.198**	3.809***	53.97***
		Ten	1.829***	-	-1.336***	23.904	16.93***
4	16	*Aes*	1.495*	0.703	-0.755*	2.721***	25.78***
		Ten	2.312*	1.879	-1.210*	4.622***	45.24***
5	18	*Aes*	-0.012	-2.993***	-0.845***	3.806***	67.90***
		Ten	0.335	-1.668	-0.782***	5.426***	57.65***
6	23	*Aes*	1.460	-2.528	-1.632**	5.607***	60.75***
		Ten	2.897***	-	-2.108***	20.354	25.17***
All	103	*Aes*	0.638**	-1.918***	-0.598***	3.075***	269.25***
		Ten	1.040***	-0.746	-0.755***	6.001***	220.90***

This table reports coefficients of the alternative-specific conditional logit model for the given paper and treatment with the base utility model combined with one heuristic: either *Aes* or *Ten*. "Dc" denotes that the treatment featured a dictator game, otherwise, it featured an ultimatum game. * means that the coefficient or stat is significant at the 10% level; ** and *** respectively the 5% and 1% levels.

Table 2.18: Comparison of Ten only and $Aes \times Ten$ factors

Paper 4 : 21 Options

Treatment	N	x	Inf. Av.	Ineq. Av.	Aes only	Ten only	$Aes \times Ten$	$Aes \times Ten$ $-Ten$ only
1	29	0.694	-1.594**	-2.838*	0.106	0.683	1.820***	1.136*
2	29	6.934	-1.818	-6.721**	0.707	1.858**	2.062***	0.205
3	35	5.543	-2.358**	-7.777***	-13.670	2.193***	2.719***	0.526
4	35	1.530	-5.776**	-9.230	2.636***	0.206	-0.649	-0.855
5	37	-2.257	-3.775***	-0.981	-13.101	1.378**	2.365***	0.987*
6	37	15.284	-2.353	-12.004	0.624	0.112	0.706	0.594
7	40	0.282	-2.853***	-1.948	-13.266	2.324***	2.329***	0.005
8	40	-3.740	-4.143***	-0.491	1.455***	0.317	0.949**	0.632
All	282	1.424	-2.883***	-3.839***	0.697**	1.034***	1.478***	0.445**

Paper 6 : 21 Options

Treatment	N	x	Inf. Av.	Ineq. Av.	Aes only	Ten only	$Aes \times Ten$	$Aes \times Ten$ $-Ten$ only
1	22	0.003	-19.852	-0.044	3.515**	4.808***	4.586***	-0.222
2	45	0.004	-16.551	0.051	-16.807	-2.147**	3.064***	5.211***
3	59	0.676**	3.720	-0.340**	0.469	2.022***	3.374***	1.352***
4	21	0.004	-16.073	0.008	1.653*	1.622**	2.612***	0.990
All	147	0.542*	1.973	-0.263*	0.268	1.056***	2.690***	1.634***

Paper 9 : 51 Options

Treatment	N	x	Inf. Av.	Ineq. Av.	Aes only	Ten only	$Aes \times Ten$	$Aes \times Ten$ $-Ten$ only
1	60	0.370***	0.879	-0.221***	0.026	2.243***	2.468***	0.226
2	72	1.478*	2.020	-0.749**	0.915*	2.324***	2.801***	0.476
All	132	0.441***	0.679	-0.241***	0.479	2.236***	2.588***	0.352

Paper 3 : 101 Options

Treatment	N	x	Inf. Av.	Ineq. Av.	Aes only	Ten only	$Aes \times Ten$	$Aes \times Ten$ $-Ten$ only
1	20	0.001	-17.851	-0.016	1.628	4.065***	4.562***	0.497
2	20	0.280	0.454	-0.184*	-0.524	-12.388	3.880***	16.268
All	40	0.318*	0.341	-0.185**	0.756	3.360***	4.237***	0.877

This table reports coefficients of the alternative-specific conditional logit model for the given paper and treatment with the base utility model combined with Aes only, Ten only, and $Aes \times Ten$. It also reports the difference between the $Aes \times Ten$ and Ten only coefficients. All treatments in this table feature the ultimatum game. * means that the coefficient or stat is significant at the 10% level; ** and *** respectively the 5% and 1% levels.

Table 2.19: Comparison of Ten only and $Aes \times Ten$ factors

Paper 8 : 101 Options

Treatment	N	x	Inf. Av.	Ineq. Av.	Aes only	Ten only	$Aes \times Ten$	$Aes \times Ten$ $-Ten$ only
(Dc) 1	45			Convergence not achieved				
(Dc) 2	46	3.670***	-	-2.033***	3.916	38.217	39.250	1.034**
3	43	8.145*	2.446	-4.720*	-24.557	4.462**	6.403***	1.941*
4	48	2.275*	-0.673	-1.695***	-28.817	3.039***	4.174***	1.136*
All	182	2.703***	-0.333	-1.532***	-25.602	4.835***	5.681***	0.846***

Paper 13 : 101 Options

Treatment	N	x	Inf. Av.	Ineq. Av.	Aes only	Ten only	$Aes \times Ten$	$Aes \times Ten$ $-Ten$ only
1	19	0.024	-16.337	0.060	-0.780	-14.537	2.508***	17.045
2	19	0.989*	-0.571	-0.854***	-1.409	1.617	2.097***	0.480
All	38	0.959*	-0.689	-0.562*	-1.142	0.665	2.307***	1.642

Paper 12 : 201 Options

Treatment	N	x	Inf. Av.	Ineq. Av.	Aes only	Ten only	$Aes \times Ten$	$Aes \times Ten$ $-Ten$ only
(Dc) 1	15	-0.21	-18.167	-0.215	-27.134	6.629***	5.665***	-0.964
(Dc) 2	14			Convergence not achieved				
3	17			Convergence not achieved				
4	16	1.965*	1.664	-0.933	-27.976	3.125***	4.312***	1.187
5	18	0.070	-1.869*	-0.568**	-27.974	3.548**	4.745***	1.197
6	23	19.795	17.029	-10.507	2.433	38.971	39.324	0.353
All	103	0.916**	-0.889	-0.656***	-24.893	5.348***	5.312***	-0.036

This table reports coefficients of the alternative-specific conditional logit model for the given paper and treatment with the base utility model combined with Aes only, Ten only, and $Aes \times Ten$. It also reports the difference between the $Aes \times Ten$ and Ten only coefficients. (Dc) denotes that the treatment featured a dictator game, otherwise, it featured an ultimatum game. * means that the coefficient or stat is significant at the 10% level; ** and *** respectively the 5% and 1% levels.

Table 2.20: Comparison of Ten only and $Oct \times Ten$ and $PF \times Ten$ factors

101 Options	Treatment - Paper 3		
	1	2	All
N	20	20	40
x	0.002	0.251	0.287*
Inf. Av.	-18.823	0.579	0.227
Ineq. Av.	-0.020	-0.163*	-0.170**
Oct only	-16.523	-13.474	-13.061
MT only	-16.420	-13.328	-13.017
PF only	2.596**	-13.710	2.107*
Ten only	3.428***	-13.608	2.941***
$Oct \times Ten$	4.097***	4.137***	4.142***
$PF \times Ten$	4.164***	3.538***	3.848***
$Oct \times Ten - Ten$ only	0.669	17.745	1.201*
$PF \times Ten - Ten$ only	0.736	17.146	0.907

101 Options	Treatment - Paper 8				
	1 (Dc)	2 (Dc)	3	4	All
N	45	46	43	48	182
x	2.225**	-0.543	8.371	2.112	2.828***
Inf. Av.	-	-4.600*	2.757	-0.931	-0.125
Ineq. Av.	-0.729	0.026	-4.796*	-1.652**	-1.610***
Oct only	-0.146	0.013	-12.065	-13.658	-13.382
MT only	-0.080	0.018	-11.974	-13.610	-13.387
PF only	-0.014	-0.096	-12.302	-13.775	-13.443
Ten only	21.109	19.479	4.552**	3.137***	4.670***
$Oct \times Ten$	22.449	20.925	6.861***	4.317***	5.930***
$PF \times Ten$	22.045	20.713	6.598***	4.614***	5.660***
$Oct \times Ten - Ten$ only	1.339***	1.446***	2.309**	1.180	1.259***
$PF \times Ten - Ten$ only	0.936	1.234**	2.046*	1.477**	0.990***

This table reports coefficients of the alternative-specific conditional logit model for the given paper and treatment with the base utility model combined with Oct only, MT only, PF only, Ten only, $Oct \times Ten$, and $PF \times Ten$. It also reports the differences between the $Oct \times Ten$ and Ten only coefficients and between the $PF \times Ten$ and Ten only coefficients. (Dc) denotes that the treatment featured a dictator game, otherwise, it featured an ultimatum game. * means that the coefficient or stat is significant at the 10% level; ** and *** respectively the 5% and 1% levels.

Table 2.21: Comparison of Ten only and $Oct \times Ten$ and $PF \times Ten$ factors

101 Options	Treatment - Paper 13		
	1	2	All
N	19	19	38
x	0.108	1.115*	1.141**
Inf. Av.	-17.178	-0.283	-0.270
Ineq. Av.	-0.021	-0.919***	-0.677**
Oct only	-15.493	-14.664	-14.578
MT only	-15.501	-14.632	-14.576
PF only	0.861	-14.723	0.498
Ten only	1.961***	2.349***	2.058***
$Oct \times Ten$	2.905***	2.503***	2.802***
$PF \times Ten$	-15.477	1.694**	1.247*
$Oct \times Ten - Ten$ only	0.943	0.154	0.745
$PF \times Ten - Ten$ only	-17.438	-0.655	-0.811

201 Options	Treatment - Paper 12			
	4	5	6	All
N	16	18	23	103
x	2.207**	0.357	12.021	1.303***
Inf. Av.	2.299*	-2.416*	-	-0.171
Ineq. Av.	-1.149**	-1.051**	-11.096	-0.902***
Oct only	-13.152	-11.377	3.787	-12.956
MT only	-13.144	-11.216	1.148	-12.918
PF only	-13.159	–11.462	6.035	-13.066
Ten only	4.713***	5.930***	39.767	6.200***
$Oct \times Ten$	5.148***	4.801***	22.157	6.120***
$PF \times Ten$	3.924***	5.688***	31.420	5.528***
$Oct \times Ten - Ten$ only	0.435	-1.129	-17.610	-0.079
$PF \times Ten - Ten$ only	-0.788	-0.242	-8.347	-0.672*

NOTE: Treatments 1, 2, and 3 did not converge

This table reports coefficients of the alternative-specific conditional logit model for the given paper and treatment with the base utility model combined with Oct only, MT only, PF only, Ten only, $Oct \times Ten$, and $PF \times Ten$. It also reports the differences between the $Oct \times Ten$ and Ten only coefficients and between the $PF \times Ten$ and Ten only coefficients. All treatments in this table feature the ultimatum game. * means that the coefficient or stat is significant at the 10% level; ** and *** respectively the 5% and 1% levels.

Chapter 3

Stronger Utility and the Endowment Effect

Abstract. P. R. Blavatskyy, 2014 introduced a model based on Fechner, 1860 that allowed for stochastic choice but was also able to prevent any stochastically dominated choices from being chosen. P. R. Blavatskyy, 2014 showed (among other things) that it predicted preference reversals, however, U. Schmidt and Hey, 2004 show that the preference reversal phenomenon is different for the willingness-to-accept (WTA) and willingness-to-pay (WTP) cases. I define the probability equivalent valuation, or PEV, as the expected value of the probability equivalent defined in P. R. Blavatskyy, 2014; this measure predicts the overweighting of small probabilities seen in the literature. I go on to modify the P. R. Blavatskyy, 2014 model with a parameter representing the endowment effect that generates 2 valuations - one for WTP and one for WTA. I show that my model predicts both the WTA-WTP disparity seen in the data and the full preference reversal phenomenon without compromising any of the features of the P. R. Blavatskyy, 2014 model.

JEL Classification: D81, G02, G12

Keywords: Probabilistic choice, Strong utility, Preference reversal phenomenon, WTA-WTP disparity, endowment effect.

3.1 Introduction

Stochastic utility has been considered as far back as Fechner, 1860 and has appeared consistently throughout the literature since: Thurstone, 1927, Luce, 1959, Machina, 1985, etc. However, well-fitting models to the data have been elusive. Any time a solution to one phenomenon is found, it goes on to predict unreasonable behaviour elsewhere.[1]

In this paper, I concentrate on the model in P. R. Blavatskyy, 2014, which allows for stochastic choice but gives us the option to eliminate the possibility that the decision-maker will make a choice that is wholly dominated by another.[2] Adapting the model in a plausible way to account for both is a big step forward, however, there are some market

[1] See section 3.2 for more information.

[2] It is possible to adapt the model to allow for such choices - whose presence can be seen empirically - which is an additional strength of the model.

phenomena that the model cannot account for, for example, the full preference reversal phenomenon.

In P. R. Blavatskyy, 2014, the author states that his model can explain the prevalence of standard preference reversals over non-standard preference reversals, however, this is not the case. U. Schmidt and Hey, 2004 demonstrate that this prevalence only exists in the willingness-to-accept (WTA) case, not the willingness-to-pay (WTP) case.[3] By introducing an additional parameter accounting for the endowment effect, I find that it is possible to recreate the results in U. Schmidt and Hey, 2004 without compromising the central features of the P. R. Blavatskyy, 2014 model.

This additional parameter has a number of other benefits: by using the expectation of the probability equivalent as a proxy for value, I find that it's also possible to predict the WTA-WTP disparity highlighted in other parts of the literature.[4] By construction, the P. R. Blavatskyy, 2014 model does not distinguish between the WTA and WTP cases so it would not be possible to observe this disparity. Again, my model introduces this ability without compromising any of the central features of the P. R. Blavatskyy, 2014 model.

It's telling that P. R. Blavatskyy, 2014 never looks at what the probability equivalent means in terms of the direct value of an uncertain payoff. This is almost certainly because the traditional notion of indifference pricing no longer applies when it comes to the probability equivalent - with the probability equivalent, it is possible for a decision-maker to choose the option with the lower utility value, which is an outcome verified in the empirical literature.[5] As I take the step of using the expected value as a proxy for value, I include a section that investigates the validity of such an idea.[6]

Accepting this proxy has one final benefit investigated in this paper and that is the prediction of the overweighting of small probabilities.[7] This is a direct result from P. R. Blavatskyy, 2014 that only comes from investigating what I term the 'probability equivalent valuation' or PEV.

The rest of the paper is structured thus: section 2 looks at the literature relevant to this paper; section 3 outlines the stochastic utility model, defines and discusses the validity of the PEV, and demonstrates the model's ability to predict the overweighting of small probabilities; section 4 shows how accounting for the endowment effect in my model predicts the WTA-WTP disparity; section 5 demonstrates that my model comes closer to explaining the full preference reversal as opposed to P. R. Blavatskyy, 2014 who only covers the WTA case; section 6 concludes the paper.

3.2 Background and Literature Review

I focus on stochastic decision-making in this paper, which has been widely studied predominantly to try and understand what part of the decision-making process is stochastic and what forms this stochasticity takes. One possibility was that decision-makers made mistakes (Luce, 1959, Camerer and Ho, 1994, Wu and Gonzalez, 1996), another was that the utility values the decision-makers held were themselves stochastic (Thurstone, 1927, Harsanyi, 1973, G. Loomes and Sugden, 1995), and another was that decision-makers deliberately randomised their decisions (Machina, 1985, Swait and Marley, 2013, Fudenberg

[3] The literature on preference reversals has focussed almost exclusively on the WTA case.
[4] Knetsch and Sinden, 1984, Kahneman et al., 1991
[5] Louie et al., 2013
[6] See Section 3.3.1
[7] Burns et al., 2010

and Strzalecki, 2015). Agranov and Ortoleva, 2015 find in favour of the latter by conducting experiments in which they ask participants to make a decision regarding a risky choice and then immediately asking them to make the same decision again. Under most models of decision-making, one would expect someone presented with the same choice twice in a row would make identical choices in both cases. However, since this is not what Agranov and Ortoleva, 2015 observe, they conclude that deliberate randomisation is a real phenomenon. Van de Kuilen, 2009 find that the probability-weighting of subjects who perform repeated decisions with feedback converges to linear.

Other models of stochastic choice include Duffie and Epstein, 1992 who develop a model of stochastic differential utility. "Trembles" are investigated in Harless and Camerer, 1994, the Fechner model with homoscedastic errors is used in Hey and Orme, 1994 and with heteroscedastic errors in Hey, 1995 and in Buschena and Zilberman, 2000.[8] P. R. Blavatskyy, 2007 uses a Fechner model with truncated, heteroscedastic errors while G. Loomes and Sugden, 1995 uses random utility with a probability measure over preference functionals. Baillon et al., 2012 uses modern ambiguity theory (Abdellaoui et al., 2011) and find that probability priors are treated differently depending on context. My paper builds on these by taking a comparatively simple model of stochastic decision-making with a focus on eliminating dominated choices and enhances it in the direction of context-dependence.

Many of the empirical phenomena I look at, including the preference reversal phenomenon, have, in some part, also been explained through non-stochastic means. For example, Cumulative Prospect Theory (CPT) (Luce and Fishburn, 1991, Tversky and Kahneman, 1992) has been an important development in the study of decision-making. It models the overweighting of probabilities and framing effects that help fit utility models to the experimental data. For example, P. R. Blavatskyy, 2013 demonstrates how CPT can be used to justify the common-ratio effect.

A central question of CPT is how real-world probabilities need to be transformed to adequately explain the empirical decision data. Wilcox et al., 2015 finds that optimism is the most prevalent form of rank-dependant weighting leading to the overweighting of small probabilities when it comes to gains and underweighting of small probabilities when it comes to losses (see also Etner and Jeleva, 2013, Ehrlich and G. S. Becker, 1972, and Burns et al., 2010). Ackert et al., 2012 find people can be split into two groups: those that overweight high payoffs and low probabilities, and those that act rationally.

The correct transformation of objective probabilities seems to also rely on stake size. Markowitz, 1952 surmises that risk-seeking behaviour turns to risk-averse behaviour as stake size increases and this is corroborated by Holt, Laury, et al., 2002 and Holt and Laury, 2005 but only when experiments are conducted with real money - hypothetical experiments did not carry this effect. Fehr-Duda et al., 2010 find that probability weighting is less optimistic for larger gains and that decision-makers exhibit Increasing Relative Risk Aversion for gains but not for losses.

P. R. Blavatskyy and Pogrebna, 2007 find that the assumption of loss aversion is violated when stakes are large. P. Blavatskyy and Pogrebna, 2008 find that agents show identical risk aversion when faced with large stakes regardless whether the probability of attaining the payoff is high or low. Linde and Sonnemans, 2012 show that individuals are more risk-averse in a loss situation than in a gain situation in a social context - counter to what prospect theory predicts. My paper confirms the phenomenon of overweighting of small probabilities using the probability equivalent defined in P. R. Blavatskyy, 2014

[8] See Fechner, 1860 for the Fechner model - see also Section 3.3.

albeit in a simplified scenario. We must ask the question, however, whether stochastic or non-stochastic decision-making is a more reasonable explanation for this phenomenon.

G. Loomes and Pogrebna, 2014 find that neither random noise nor CPT can explain deviations from the independence axiom. Cubitt et al., 2015 find that imprecision is a feature of choice. D. Butler et al., 2012 show that the assumption that choices have noise-independence is systematically violated. They also find that the models of P. R. Blavatskyy, 2009 & P. R. Blavatskyy, 2011 and G. M. Becker et al., 1963 & G. Loomes and Sugden, 1995 come closest to adequately explaining empirical phenomena.

There is clear evidence that empirical data supports a stochastic decision-making model and there are many available each with their advantages and disadvantages. Many (Fechner, 1860 and Luce, 1959, for example) can be used to justify puzzles such as the Allais paradox (Allais, 1953, Tversky, 1969) but can sometimes produce behaviour that seems implausible, such as the overly prevalent habit of selecting lotteries that are stochastically dominated.

G. Loomes et al., 2002 conduct experiments to determine which stochastic model from the Fechner model , random preferences, and "trembles" is best and finds that the lack of choice for stochastically dominated options makes the Fechner model a bad fit for the empirical data. Drichoutis and Lusk, 2014 compare the Fechner, Luce, and contextual utility models and find in favour of the latter. However, Holt, Laury, et al., 2002 find that agents don't accurately predict how they'll act in a real situation so hypothetical experiments may not give reliable results.

P. R. Blavatskyy, 2014 presents a model that allows stochastic choice but prevents lotteries that are stochastically dominated from being chosen. These are extremely desirable characteristics merging the best of stochastic models and Expected Utility Theory. This model can also be used to explain such puzzles as the "basic" preference reversal phenomenon (in the WTA case) and common ratio effect. However, its construction does not permit an explanation for the WTA-WTP disparity nor the "full" preference reversal phenomenon (for both the WTA and WTP cases). My paper resolves these problems, however, determining whether the additional parameter is worth its inclusion would require additional empirical tests in the same vein as those listed above.

The preference reversal (PR) phenomenon was shown in Lichtenstein and Slovic, 1971, Lichtenstein and Slovic, 1973, Lindman, 1971, and Slovic, 1975 and confirmed by Grether and Plott, 1979. The set-up for the PR phenomenon is when an agent is presented with two lotteries: one with a high payout but a low chance of winning (the $-bet), and one with a relatively low payout but a relatively high chance of winning (the P-bet). A PR has occurred if, after having been asked to choose between the two lotteries, the agent goes on to value the other lottery more highly.

U. Schmidt and Hey, 2004 shows that the tendency to see standard PR more often than non-standard PR only exists in the WTA case. They happen with similar frequency in the WTP case. A summary of early PR papers can be found in Seidl, 2002.

D. J. Butler and G. C. Loomes, 2007 find that imprecision of preferences can explain PR while P. R. Blavatskyy, 2009 finds that probabilistic preferences can also explain PR. Alternatively, U. Schmidt et al., 2008 present Third-generation Prospect Theory - identifiable by a variable reference point - and finds that both the PR phenomenon and WTP-WTA disparity can be explained without probabilistic or imprecise preferences. U. Schmidt and Hey, 2004 find most PR is the result of pricing errors rather than choice errors. G. Loomes, Pogrebna, et al., 2015 don't find one best method for describing choice and pricing behaviour but do determine that probabilistic preferences are important.

Hollard et al., 2015 confirms the inconsistencies in PR, framing effects, and preference instability but determines that there is no correlation between them. By building on the work done by P. R. Blavatskyy, 2014, I improve its ability to predict the full preference reversal phenomenon.

The WTA-WTP disparity comes about when individuals put a higher price on something if they are selling it (Willingness-to-Accept: WTA) than if they are buying it (Willingness-to-Pay: WTP). Knetsch and Sinden, 1984 identify the WTA-WTP disparity in hypothetical values. Tunçel and Hammitt, 2014 - building on Horowitz and McConnell, 2002 confirms the disparity and explores its size-variability for different goods.

Plott and Zeiler, 2005 suggest that the WTA-WTP disparity can be explained by participants' misconceptions about the experimental procedure during experiments in which they have to set WTP and WTA prices. They use this as evidence that the endowment effect does not provide an explanation for the disparity. On the other hand, Isoni et al., 2011 reruns Plott and Zeiler, 2005's experiments and, although they find the same conclusion in regard to the valuation of mugs (a common item used in valuation experiments), they find that the WTA-WTP disparity persists when subjects are asked to value lotteries.[9] By constructing my model to predict the preference reversal phenomenon, I discover that it also predicts the WTA-WTP disparity as a natural consequence.

3.3 The Stochastic Utility Model

Fechner, 1860 presented a model of stochastic choice whereby a decision-maker chooses lottery L over lottery L' if

$$U(L) - U(L') \geq \xi$$

where ξ is an error term, usually with mean 0, and $U(\cdot)$ defines expected value of the decision-maker's utility function.

P. R. Blavatskyy, 2014 modifies this model so that the same decision-maker chooses lottery L over lottery L' if

$$U(L) - U(L') \geq \epsilon \cdot [U(L \vee L') - U(L \wedge L')]$$

where ϵ is an error term defined on the interval $[-1, 1]$, $L \vee L'$ is the lottery that stochastically dominates L and L' but for which there is no third lottery, $L'' \neq L \vee L'$, that stochastically dominates L and L' but not $L \vee L'$, and $L \wedge L'$ is the lottery that is stochastically dominated by L and L' but for which there is no third lottery, $L''' \neq L \wedge L'$, that is stochastically dominated by L and L' but not by $L \wedge L'$.

In this model, the probability that the decision-maker will select lottery L over lottery L' is equal to

$$P(L, L') = F\left(\frac{U(L) - U(L')}{U(L \vee L') - U(L \wedge L')}\right)$$

where $F : [-1, 1] \to [0, 1]$ is a cumulative distribution function with the characteristic that $F(-v) + F(v) = 1$ for all $v \in [0, 1]$, i.e. symmetry about 0.

[9] Jaspersen, 2015 provides a review of the insurance literature including WTP research. D. Green et al., 1998 find that WTP is highly vulnerable to anchoring.

Subsequently, the probability that a certain amount, x, is preferred to some lottery, L, is

$$P\left(x, L\right) = F\left(\frac{U(x) - U(L)}{U(x \vee L) - U(x \wedge L)}\right)$$

and this is known as a probability equivalent. Suppose that the highest and lowest payouts in lottery L are y and m respectively, then we can generate the "probability equivalent valuation" or PEV of this lottery, denoted $\Pi_{U,L}$, as the expected value of the probability equivalent, which can be written as

$$\Pi_{U,L} = \int_{-\infty}^{\infty} x \, \mathrm{d}P(x, L) = y - \int_{m}^{y} P(x, L) \mathrm{d}x.$$

Take a simple lottery in which one can win a positive amount y with probability $p \in (0, 1)$ or 0 with probability $1 - p$. I denote this lottery using the notation $(y, 0; p)$.

In this notation, the lottery $x \vee L$ is denoted $(y, x; p)$ and the lottery $x \wedge L$ is denoted $(x, 0; p)$.

Suppose the agent is risk-neutral, i.e. $U(X) = E(X)$, then $U(x) = x$, $U(L) = py$, $U(x \vee L) = py + (1 - p)x$, and $U(x \wedge L) = px$. This means that

$$P(x, L) = F\left(\frac{U(x) - U(L)}{U(x \vee L) - U(x \wedge L)}\right) = F\left(\frac{x - py}{(1 - 2p)x + py}\right)$$

You can see that

$$P(y, L) = F\left(\frac{y - py}{(1 - 2p)y + py}\right) = \left(\frac{(1 - p)y}{(1 - p)y}\right) = F(1) = 1$$

indicating that, if $x = y$, the certain amount stochastically dominates the lottery and is sure to be chosen. On the other hand, notice that

$$P(0, L) = F\left(\frac{0 - py}{0 + py}\right) = F(-1) = 0$$

indicating that, if $x = 0$, the certain amount is stochastically dominated by the lottery and will never be chosen.

$F(\cdot)$ can be any symmetric (about 0) cumulative distribution function but I focus on only a couple:

- Rational expectations distribution

$$F(v) = \begin{cases} 0 & , \quad v < 0 \\ 1 & , \quad v \geq 0 \end{cases}$$

- Uniform distribution

$$F(v) = \begin{cases} 0 & , \quad v < -1 \\ \frac{1}{2} + \frac{1}{2}v & , \quad -1 \leq v \leq 1 \\ 1 & , \quad v > 1 \end{cases}$$

- Raised cosine distribution with limit parameter, s,

$$F(v) = \begin{cases} 0 & , \quad v < -s \\ \frac{1}{2}[1 + \frac{v}{s} + \frac{1}{\pi}\sin\left(\frac{\pi v}{s}\right)] & , \quad -s \leq v \leq s \\ 1 & , \quad v > s \end{cases}$$

the last of which was found by P. R. Blavatskyy, 2014 to provide the best fit to the empirical data of the 8 distribution functions he looked at.

In addition, I look at 2 types of decision-maker: risk-neutral and risk-averse. For the risk-averse investor I use an exponential utility function with constant absolute risk aversion equal to α.

- If we use a **risk-neutral** agent and a **rational expectations** distribution, the valuation of a simple lottery, $(y, 0; p)$, is

$$\Pi = py$$

- If we use a **risk-neutral** agent and a **uniform** distribution, the PEV of $(y, 0; p)$ is

$$\Pi = \frac{1}{2}y$$

if $p = \frac{1}{2}$ and

$$\Pi = \frac{1}{2}y\left(1 - \frac{1}{1-2p} + \frac{2p(1-p)}{(1-2p)^2}\ln\left(\frac{1-p}{p}\right)\right)$$

otherwise.

The comparison of valuations using these two distributions is

Price/y | p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Certainty Equivalent	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability Equivalent	0.184	0.283	0.362	0.433	0.5	0.567	0.638	0.717	0.816

The result is a probability weighting function that weights probabilities in a manner consistent with the data i.e. overweighting of small probabilities.

- If we use a **risk-neutral** agent and **raised-cosine** distribution, the comparison of valuations for $(y, 0; p)$ is

Price/y | p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Certainty Equivalent	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
PE: Uniform	0.184	0.283	0.362	0.433	0.5	0.567	0.638	0.717	0.816
PE: Raised-cosine	0.124	0.229	0.323	0.413	0.5	0.587	0.677	0.772	0.876

suggesting that this distribution also provides a consistent probability weighting function with smaller deviations from the rational expectations model.

- If we use a **risk-averse** agent and **rational expectations** distribution, the valuation of $(y, 0; p)$ is

$$\Pi = -\frac{1}{\alpha}\ln(pe^{-\alpha y} + 1 - p)$$

- If we use a **risk-averse** agent and **uniform** distribution, the PEV of $(y, 0; p)$ is

$$\Pi = \frac{1}{\alpha} + \frac{ye^{-\alpha y}}{e^{-\alpha y} - 1}$$

if $p = \frac{1}{2}$ and

$$\Pi = \frac{1}{2}y + \frac{1}{2}\left(\frac{pe^{-\alpha y} + (1-p)}{pe^{-\alpha y} - (1-p)}y + \frac{1}{\alpha}\left(\frac{1}{1-2p} + \frac{pe^{-\alpha y} + (1-p)}{pe^{-\alpha y} - (1-p)}\right)\ln\left(\frac{1-p}{p}\right)\right)$$

otherwise.

The comparison of valuations using these two distributions is

$\alpha = 0.1$	$y = 0.1$			$y = 1$			$y = 10$		
p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
Certainty Equivalent	0.020	0.050	0.080	0.192	0.488	0.792	1.352	3.799	7.046
PE: Uniform	0.028	0.050	0.072	0.276	0.492	0.710	2.204	4.180	6.428
$\alpha = 1$	$y = 0.1$			$y = 1$			$y = 10$		
p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
Certainty Equivalent	0.019	0.049	0.079	0.135	0.380	0.705	0.223	0.693	1.609
PE: Uniform	0.028	0.049	0.071	0.220	0.418	0.643	0.462	1.000	1.847

- If we use a **risk-averse** agent and **raised-cosine** distribution, the comparison of valuations using these two distributions is

$\alpha = 0.1$	$y = 0.1$			$y = 1$			$y = 10$		
p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
Certainty Equivalent	0.020	0.050	0.080	0.192	0.488	0.792	1.352	3.799	7.046
PE: Rasied-cosine	0.023	0.050	0.077	0.221	0.489	0.764	1.616	3.943	6.820
$\alpha = 1$	$y = 0.1$			$y = 1$			$y = 10$		
p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
Certainty Equivalent	0.019	0.049	0.079	0.135	0.380	0.705	0.223	0.693	1.609
PE: Rasied-cosine	0.022	0.049	0.076	0.162	0.394	0.682	0.279	0.774	1.665

3.3.1 Discussion on the PEV

In traditional economics, the value of a lottery is often the certain amount that generates the same utility as the lottery. This usually indicates that the agent is indifferent between holding the certain amount and the lottery and an upward (downward) change in this amount will lead to the agent always choosing the certain amount (the lottery).

Stochastic decision-making has made this distinction tougher with the certain amount that has the same utility merely indicating that the chance of choosing the certain amount is 50%. Changes in the amount just change the probability of it being chosen so the concept of indifference becomes poorly adapted to this situation - indifference happens by degrees rather than as a cut-off point with a distinct value.

This has informed my decision to use the expected value of the probability equivalent as my valuation measure, however, it's difficult, on the face of it, to easily accept this quantity as the value of the lottery due to the imprecise nature of preference that this model introduces. For a start, in traditional economics, one would expect the values of two different lotteries with the same utility to have the same valuation. However, it is easy to see that this is not true for the PEV.

Take a look at the PEVs generated for a risk-neutral agent: if equal utility meant two lotteries had the same PEV, then the value for the lottery with $p = 0.8$ should be double that of the lottery with $p = 0.4$. Although this is the case for the rational expectations model, it is true for neither the uniform ($0.433 \times 2 \neq 0.717$) nor the raised cosine ($0.413 \times 2 \neq 0.772$) distributions.

This is not a problem, however, because if utility were a sufficient proxy for value, there would be no need for the probability equivalent nor for models of stochastic decision-making. Indeed, if we take the literature on the overweighting of small probabilities seriously, it is possible to argue that the PEV is more representative of real investor behaviour.

One can interpret the expectation of the probability equivalent as the amount the agent would expect to have had he chosen the certain amount over the lottery. In this sense, the PEV measures something akin to the opportunity cost associated with the lottery. See Atallah, 2006 and Frederick et al., 2009 for examples of the literature on opportunity costs.

Since the opportunity cost is a real phenomenon in the economics literature, one would expect the PEV to be highly, if not perfectly, positively correlated with the underlying value of the lottery. Despite not being a perfect analogue for the utility of the lottery, it's hard to say that the PEV is in no way informative as to the lottery's value. Indeed, the very nature of stochastic decision-making puts the practice of indifference pricing into existential doubt.

There is indirect evidence in the neurofinance literature that the expectation of the probability equivalent is a suitable way of calculating value. Louie et al., 2013 show that context significantly affects the way humans and monkeys choose and that the same option can easily receive different valuations when the context changes, even if the post-choice utility of that option doesn't change. Specifically, their model involves a normalisation by division of the probability of choosing a particular option in a similar, albeit not identical, manner to the probability equivalent. They argue that "the divisive scaling documented in value coding plays a critical role in decision making and underscore the importance of incorporating normalization processes in the interpretation of decision-making activity and behavior".

Further research into the relationship between the PEV and normalisation described

in Louie et al., 2013 would be fruitful since this method of valuation may help account for some discrepancies in pricing that behavioural finance and neurofinance have, since their inception, uncovered as a result of context-dependence. The confirmation in empirical tests that valuation relies not just on the underlying utility of an option supports the use of the PEV as a proxy for value at the moment of choice.

3.4 WTA-WTP disparity

The P. R. Blavatskyy, 2014 model dictates that the underlying probability density function of the agent's decision-making model be symmetric about 0, thereby ensuring that the WTA and WTP cases are treated equally save for their respective utility values. Among other things, this means that two options with the same utility value must be chosen with equal likelihood.

This, however, means that the model can only produce a single valuation, which does not reflect reality. The difference between the willingness-to-pay and willingness-to-accept valuations is a well-established phenomenon (see Section 3.2) and it would be useful to take the P. R. Blavatskyy, 2014 model, which already holds the attractive feature of preventing the choice of stochastically dominated options, and modify it in a reasonable way so that it, too, can produce multiple valuations that depend on whether the asset is being bought, sold, or neither.

I modify the model with the inclusion of the variable $Q(\omega)$, which is the probability that, given a choice between a lottery with a known utility and a certain amount with identical utility, the agent chooses the one they are currently in possession of, i.e. $Q(\omega) \neq \frac{1}{2}$ represents the endowment effect.

The variable ω denotes the state in which the choice is made and it has three possible values:

$$\begin{array}{ll} \omega_x & : \text{the decision-maker holds the certain amount} \\ \omega_L & : \text{the decision-maker holds the lottery} \\ \omega_0 & : \text{the decision-maker holds neither} \end{array}$$

and the corresponding values of Q are

$$\begin{array}{rcl} Q(\omega_x) & = & q \\ Q(\omega_L) & = & 1-q \\ Q(\omega_0) & = & 1/2 \end{array}$$

where $1 > q \geq 1/2$ is a parameter to be estimated. The stipulation that q be no less than one half is a result of the construction of $P(x, L)$ and the precepts of the endowment effect.

I define the transformation

$$G_{Q(\omega)}(v) = \begin{cases} 2Q(\omega)v & , \quad v \in [0, \frac{1}{2}] \\ Q(\omega) + 2(1 - Q(\omega))(v - \frac{1}{2}) & , \quad v \in (\frac{1}{2}, 1] \end{cases}$$

such that $G_{Q(\omega)}(F(0)) = Q(\omega)$. This transformation defines a valid and continuous cumulative distribution function but introduces a non-differentiable point at $v = \frac{1}{2}$ for all but a few underlying distributions and values of Q. It would serve future research well to

derive a globally continuous and differentiable function, however, for the purposes of this paper at least, this simple modification is both apt and demonstrative.

The PEV for a lottery with minimum and maximum payouts m and y respectively using this new distribution is

$$\Pi_{Q(\omega)} = y - \int_m^y G_{Q(\omega)}(P(x,L))\mathrm{d}x$$

$$= y - 2Q(\omega)\int_m^{CE_L} P(x,L)\mathrm{d}x - 2(1-Q(\omega))\int_{CE_L}^y P(x,L)\mathrm{d}x + (1-2Q(\omega))(y-CE_L)$$

where CE_L is the certain amount such that $U(CE_L) = U(L)$. To prove that this formula generates WTP and WTA valuations that conform to the data, first I show that $\Pi_{0.5}$, the PEV in the absence of the endowment effect, is equal to that of the Blavatskyy (2014) model:

$$\Pi_{0.5} = y - \int_m^{CE_L} P(x,L)\mathrm{d}x - \int_{CE_L}^y P(x,L)\mathrm{d}x + 0$$

$$= y - \int_m^y P(x,L)\mathrm{d}x = \Pi \ \square$$

Next, take the derivative of Π_Q with respect to Q:

$$\frac{\partial \Pi_Q}{\partial Q} = -2\int_m^{CE_L} P(x,L)\mathrm{d}x + 2\int_{CE_L}^y P(x,L)\mathrm{d}x - 2(y-CE_L)$$

$$= -2\left(\int_m^{CE_L} P(x,L)\mathrm{d}x + \int_{CE_L}^y (1-P(x,L))\mathrm{d}x\right) < 0$$

Hence, when Q is larger than $\frac{1}{2}$ i.e. in the case where she has the certain amount and is thinking of buying the lottery (WTP), the PEV will be lower than in the case without endowment. Similarly, when Q is smaller than $\frac{1}{2}$ i.e. in the case where she has the lottery and is thinking of selling it (WTA), the PEV will be higher than in the case without endowment. I conclude that my model conforms with the WTA-WTP disparity evidence.

Example

Returning to our simple lottery, $(y, 0; p)$, setting q to 0.6, and using the uniformly-distributed $F(\cdot)$, we can see how the PEVs change

Π/y | p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Certainty Equivalent	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability Equivalent (ω_0)	0.184	0.283	0.362	0.433	0.5	0.567	0.638	0.717	0.816
(WTP) PE (ω_x)	0.155	0.243	0.316	0.384	0.450	0.518	0.592	0.678	0.787
(WTA) PE (ω_L)	0.213	0.322	0.408	0.482	0.550	0.616	0.684	0.757	0.845

Using, instead, the raised-cosine distribution of $F(\cdot)$ with $s = 1$, we get

Π/y | p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Certainty Equivalent	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability Equivalent (ω_0)	0.124	0.229	0.323	0.413	0.5	0.587	0.677	0.772	0.876
(WTP) PE (ω_x)	0.111	0.208	0.297	0.384	0.470	0.559	0.651	0.751	0.863
(WTA) PE (ω_L)	0.137	0.249	0.349	0.442	0.530	0.616	0.703	0.792	0.889

Tables 3.1 and 3.2 show more examples of PEVs for the lottery $(y, 0; p)$ generated by this model.

3.5 The Preference Reversal Phenomenon

P. R. Blavatskyy, 2014 shows that his model can justify the preference reversal phenomenon (see Seidl, 2002 for a recent review), however, since his model treats the WTA and WTP cases identically, his model contradicts the evidence in U. Schmidt and Hey, 2004 who demonstrate that the preference reversal phenomenon has different characteristics for the two cases.

The phenomenon can be demonstrated with two lotteries of a similar expected value. One lottery, L say, has a payout, y, that can be won with probability $L(y)$. On the other hand, the other lottery, L' say, has a relatively large payout, $z > y$, that can be won with a relatively low probability, $L'(z) < L(y)$. The first is known as the P-bet and the second is known as the $-bet.

A standard preference reversal is such that an individual prefers the P-bet to the $-bet in a direct choice between the two, however, when engaged in a valuation exercise of the two lotteries, assigns a higher value to the $-bet - a contradiction if classical rational behaviour is assumed. A non-standard preference reversal is where the $-bet is chosen in a direct choice but given a lower value in the valuation exercise.

The majority of the literature confirms that standard PR occurs more often than non-standard PR and this is the phenomenon that P. R. Blavatskyy, 2014 addresses. However, U. Schmidt and Hey, 2004 finds that this is only true in the WTA case and that, in the WTP case, they happen at a more or less equal rate.

Using the same method as P. R. Blavatskyy, 2014, I determine the conditions under which my model predicts the entire preference reversal phenomenon. If $P(L', L)$ is the probability that an individual chooses the $-bet over the P-bet and $P(x, L)$ is the probability equivalent of lottery L, standard preference reversals are more likely to happen than non-standard preference reversals if

$$(1-P(L',L))\cdot\int G_{Q(\omega)}(P(x,L))\mathrm{d}G_{Q(\omega)}(P(x,L')) > P(L',L)\cdot\left(1-\int G_{Q(\omega)}(P(x,L))\mathrm{d}G_{Q(\omega)}(P(x,L'))\right)$$

$$\iff \int G_{Q(\omega)}(P(x,L))\mathrm{d}G_{Q(\omega)}(P(x,L')) > P(L',L)$$

Similarly, standard preference reversals are as likely to happen as non-standard preference reversals if

$$\int G_{Q(\omega)}(P(x,L))\mathrm{d}G_{Q(\omega)}(P(x,L')) = P(L',L)$$

P. R. Blavatskyy, 2014 shows that the former is possible when $Q = \frac{1}{2}$. If my model correctly predicts the preference reversal phenomenon, then I would observe that $\int G_{Q(\omega)}(P(x,L))\mathrm{d}G_{Q(\omega)}(P(x,L'))$ is larger in the WTA case ($Q < \frac{1}{2}$) and smaller in the WTP case ($Q > \frac{1}{2}$). This means that a necessary, but not sufficient, condition for my model to predict PR is

$$\left(\frac{\partial \int G_Q(P(x,L))\mathrm{d}G_Q(P(x,L'))}{\partial Q}\right)_{Q=0.5} < 0.$$

This is equivalent to the expression

$$\int_{CE_L}^{z}\left(P(x,L)-\frac{1}{2}\right)\mathrm{d}P(x,L') - \int_{0}^{CE_{L'}} P(x,L)\mathrm{d}P(x,L') > 0 \tag{3.1}$$

for any pair of lotteries L and L' as described. The derivation of equation (3.1) is in Appendix I.[10]

Using numerical methods on the case where both lotteries have the same expected payoff, I find that equation 3.1 holds for all pairs of lotteries in the set

$$\{L,\, L';\;\; 0.99 \geq L(y) \geq 0.02,\;\; (L(y)-0.01) \geq L'(z) \geq 0.01, E(L) = E(L')\}$$

for both risk-neutral and risk-averse agents and using the uniform and rasied cosine ($s = 1$) distributions as the underlying CDF. Figures 3.1-3.4 show the values of the left-hand side of equation 3.1 for this set. Since P. R. Blavatskyy, 2014 proves the existence of preference reversals within his model, I can conclude that my addition improves the fit of his model to the existing data under reasonable conditions.

I have shown that accounting for the endowment effect in the model presented by P. R. Blavatskyy, 2014 can replicate the WTA-WTP disparity seen in the empirical data and bring the model closer to replicating the full preference reversal phenomenon (not just the WTA case) without compromising the most important characteristic of the original model, that is, a stochastic utility model that can prevent stochastically dominated options from being chosen. I achieve this with the addition of a single parameter that represents the probability that an individual will, from two options, choose the option they already possess given both options impart to them the same utility.

This simple addition explains much, however, it does not permit modifications anywhere other than at the point in the CDF where the two options have equal utility and also generates an underlying CDF that is not globally differentiable. This paper easily demonstrates that it's possible to improve the P. R. Blavatskyy, 2014 model, however, this caveat may mean that it's possible to finesse the modification and improve the fit of the model to the data further.

[10]Finding a sufficient condition to correctly predict the preference reversal phenomenon is extremely difficult due to the large number of degrees of freedom in the number of possible specifications. This test serves to show that my model at least satisfies an important necessary condition.

3.6 Conclusion

The P. R. Blavatskyy, 2014 model solves one problem with stochastic decision-making models - that is the tendency to predict unreasonable behaviour regarding dominated choices - but doesn't account for other phenomena such as the WTA-WTP disparity or preference reversals.

Without any modification, I conjecture that the expectation of the probability equivalent in P. R. Blavatskyy, 2014 can be used to generate valuations (PEVs) for lotteries and other risky assets. It subsequently predicts that individuals will overweight small probabilities without resorting to using a probability weighting function.

I conjecture that this could be the result of normalisation - a neural mechanism that affects decision-making...

With an additional single parameter representing the endowment effect added to the P. R. Blavatskyy, 2014 model, I can maintain all the contributions that his model achieved but also explain other phenomena that his model does not.

First of these is the WTA-WTP disparity that is widely reported in the literature and has a presence in the markets in the form of the bid-ask spread. By changing the model such that an individual is more likely to want to hold on to something they already have than trade for something else that has an identical utility, I can generate two valuations that influence the WTA and WTP cases which the original model cannot. These valuations are consistent with existing data on the WTA-WTP disparity.

Finally, although the original model can explain preference reversals in the WTA case, its inability to differentiate between the WTA and WTP cases means it cannot explain preference reversals in the WTP case. I state an inequality in equation (3.1) that must be satisfied if my model is to correctly predict the empirical data on the differences between the two cases. Indeed, the change I make to the model satisfies this necessary condition and suggests a better fit to the empirical data than the P. R. Blavatskyy, 2014 model.

My contribution demonstrates that an existing model can be improved to better fit the data, how it can be improved, and why it should be. My proposed modification adds a single parameter, however, it focuses on the point in the CDF where one would find the certainty equivalent. This is an arbitrary decision and, although it greatly simplifies the calculations in this paper and permits the inclusion of only one additional parameter, this is most likely the place where future improvements could be made.

There is also the matter of taking into account the context in which the decision takes place. I deliberately look at the simplest situation possible in order to determine whether my model is appropriate but, given the nature of the field of research, there are many other factors that should be considered including but not only the level of existing wealth, the case in which the agent is short-selling, and the agent's existing portfolio profile.

My contribution to the literature is to modify an existing decision-making model to include the endowment effect without undoing the contributions the model made in the first place. This provides a much more potent decision-making model that has a broader range of applications.

Figure 3.1: Values of the LHS of equation 3.1 (agent is risk-neutral)

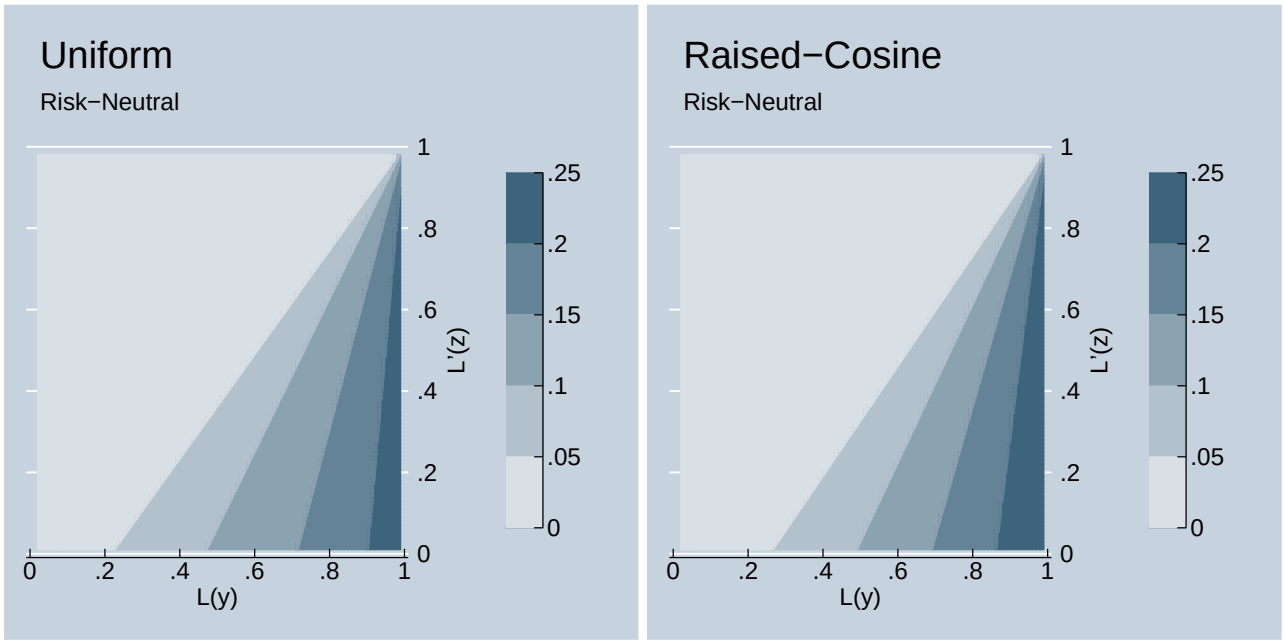


Figure 3.2: Values of the LHS of equation 3.1 (agent is risk-averse and low payoff (y) is 0.1)

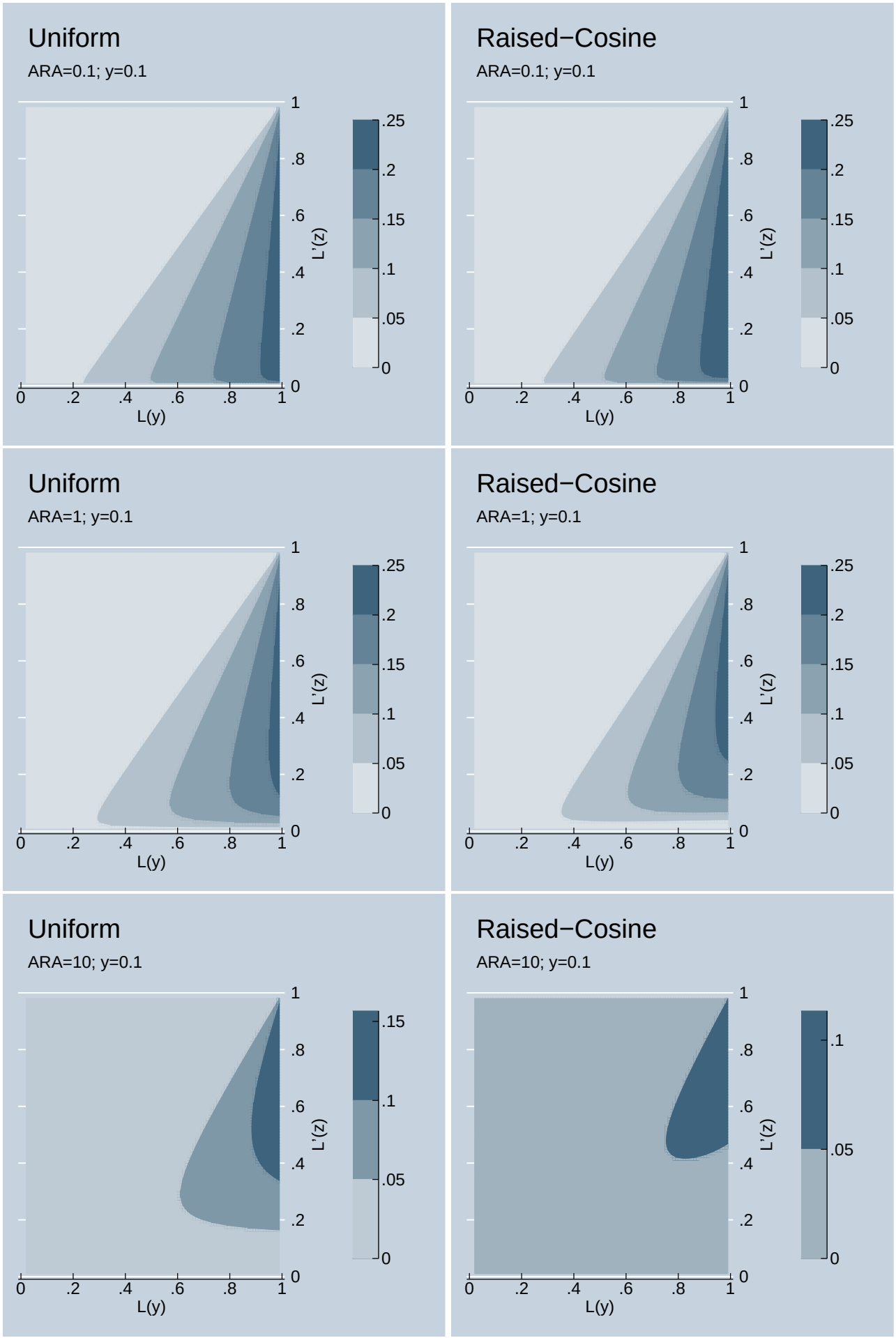


Figure 3.3: Values of the LHS of equation 3.1 (agent is risk-averse and low payoff (y) is 1)

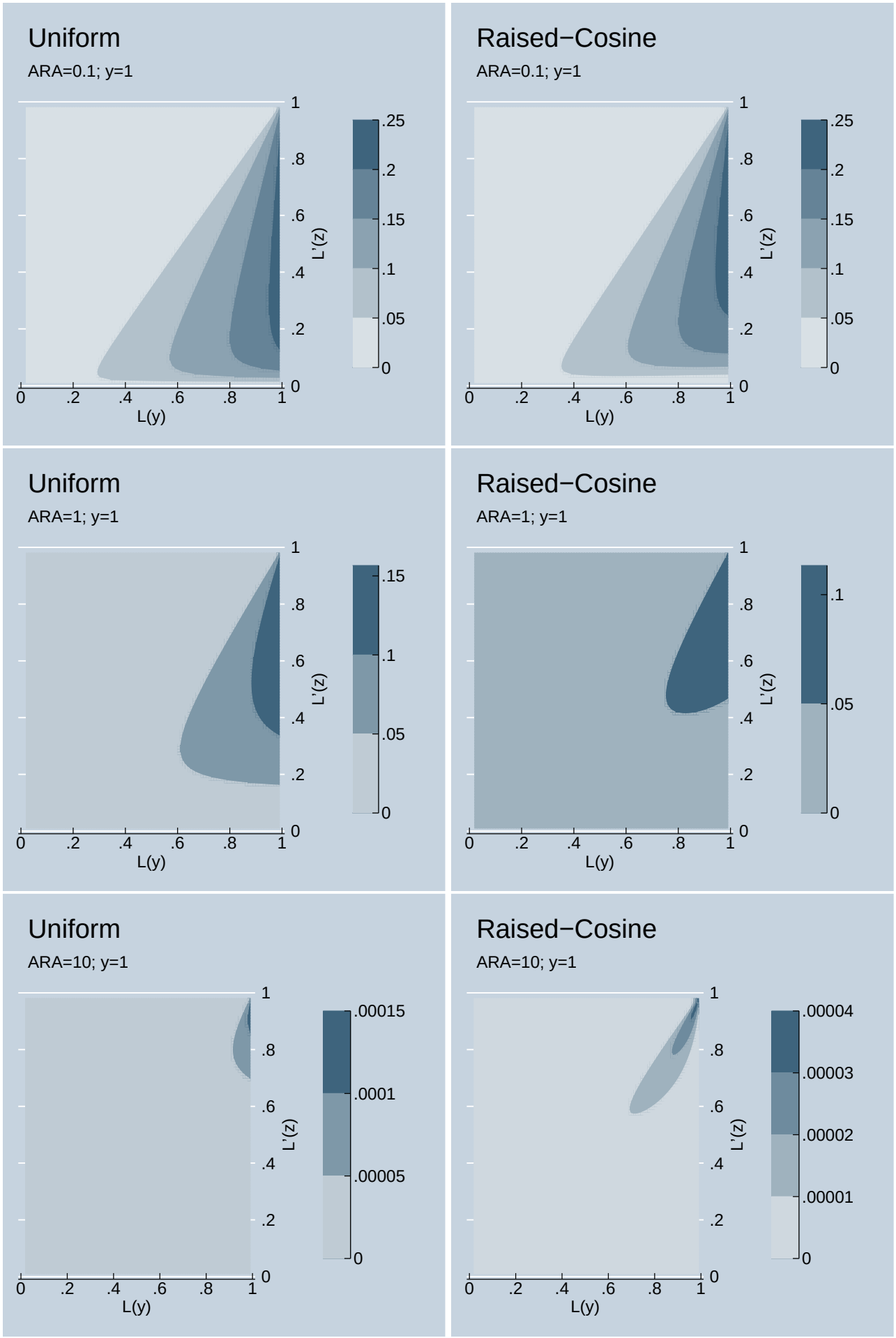


Figure 3.4: Values of the LHS of equation 3.1 (agent is risk-averse and low payoff (y) is 10)

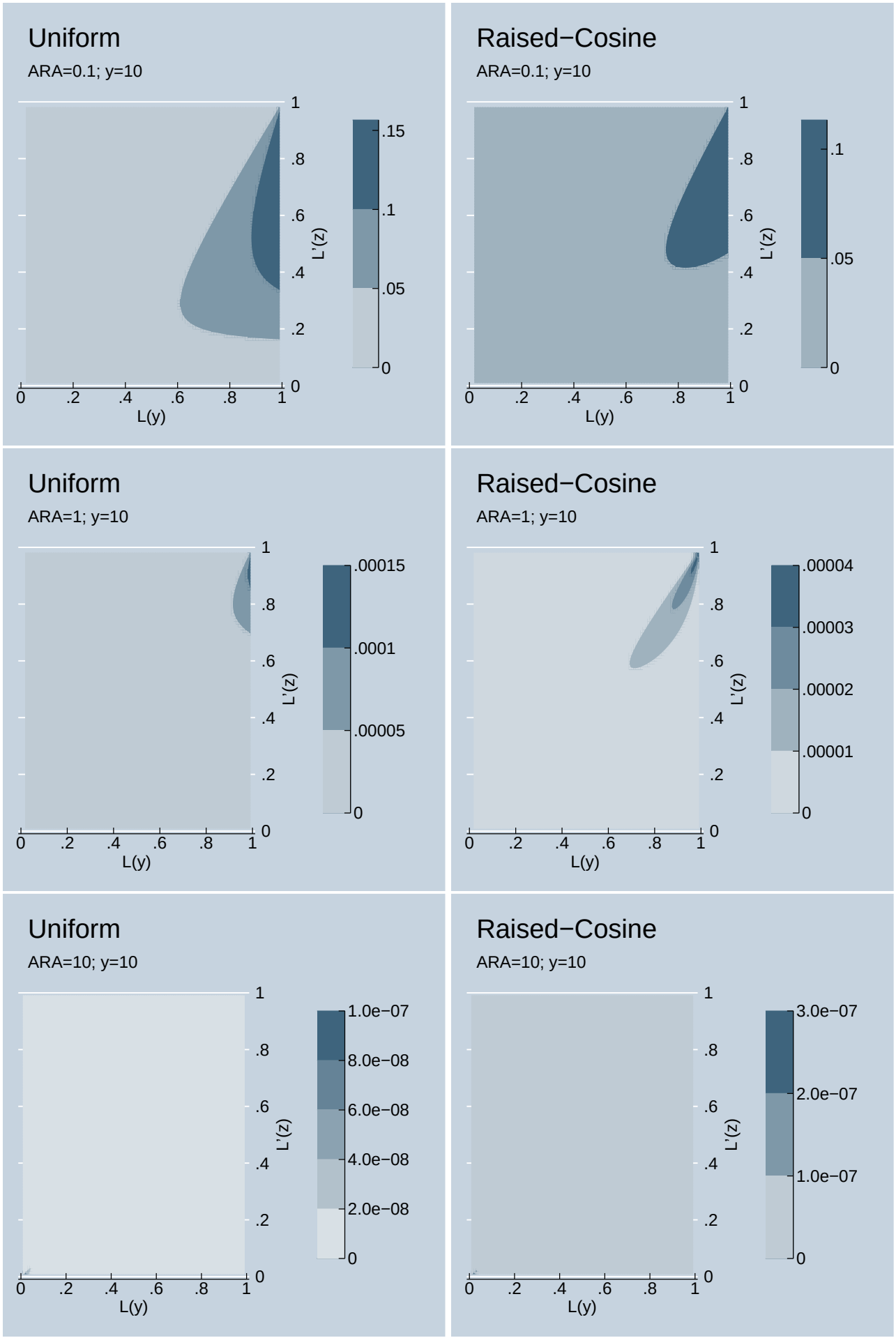


Table 3.1: Probability Equivalent using the Uniform Distribution

Risk-neutral PEVs (as a proportion of y)

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
CE_L	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1-q (WTA)									
0.1	0.299	0.441	0.545	0.629	0.700	0.763	0.821	0.876	0.931
0.2	0.270	0.402	0.499	0.580	0.650	0.714	0.775	0.836	0.902
0.3	0.241	0.362	0.454	0.531	0.600	0.665	0.729	0.797	0.873
0.4	0.213	0.322	0.408	0.482	0.550	0.616	0.684	0.757	0.845
0.5	0.184	0.283	0.362	0.433	0.500	0.567	0.638	0.717	0.816
0.6	0.155	0.243	0.316	0.384	0.450	0.518	0.592	0.678	0.787
0.7	0.127	0.203	0.271	0.335	0.400	0.469	0.546	0.638	0.759
0.8	0.098	0.164	0.225	0.286	0.350	0.420	0.501	0.598	0.730
0.9	0.069	0.124	0.179	0.237	0.300	0.371	0.455	0.559	0.701
q (WTP)									

Risk-averse PEVs (α represents constant absolute risk aversion)

$\alpha = 0.1$	$y = 0.1$			$y = 1$			$y = 10$		
p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
CE_L	0.020	0.050	0.080	0.192	0.488	0.792	1.352	3.799	7.046
1-q (WTA)									
0.1	0.044	0.070	0.088	0.432	0.692	0.872	3.523	6.101	8.231
0.2	0.040	0.065	0.084	0.393	0.642	0.831	3.193	5.621	7.780
0.3	0.036	0.060	0.080	0.354	0.592	0.791	2.863	5.141	7.330
0.4	0.032	0.055	0.076	0.315	0.542	0.751	2.534	4.660	6.879
0.5	0.028	0.050	0.072	0.276	0.492	0.710	2.204	4.180	6.428
0.6	0.024	0.045	0.068	0.237	0.442	0.670	1.874	3.700	5.978
0.7	0.020	0.040	0.064	0.198	0.392	0.630	1.545	3.220	5.527
0.8	0.016	0.035	0.060	0.159	0.342	0.589	1.215	2.740	5.077
0.9	0.012	0.030	0.056	0.120	0.292	0.549	0.886	2.260	4.626
q (WTP)									

$\alpha = 1$	$y = 0.1$			$y = 1$			$y = 10$		
p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
CE_L	0.019	0.049	0.079	0.135	0.380	0.705	0.223	0.693	1.609
1-q (WTA)									
0.1	0.043	0.069	0.087	0.352	0.610	0.823	0.759	1.554	2.657
0.2	0.039	0.064	0.083	0.319	0.562	0.778	0.685	1.415	2.455
0.3	0.035	0.059	0.079	0.286	0.514	0.733	0.611	1.277	2.252
0.4	0.032	0.054	0.075	0.253	0.466	0.688	0.536	1.138	2.049
0.5	0.028	0.049	0.071	0.220	0.418	0.643	0.462	1.000	1.847
0.6	0.024	0.044	0.067	0.187	0.370	0.598	0.388	0.861	1.644
0.7	0.020	0.039	0.063	0.155	0.322	0.553	0.313	0.723	1.442
0.8	0.016	0.034	0.059	0.122	0.274	0.508	0.239	0.584	1.239
0.9	0.012	0.029	0.055	0.089	0.226	0.463	0.165	0.445	1.037
q (WTP)									

Table 3.2: Probability Equivalent using the Rasied Cosine Distribution

Risk-neutral PEVs (as a proportion of y)

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
CE_L	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1-q (WTA)									
0.1	0.176	0.312	0.427	0.528	0.619	0.703	0.781	0.855	0.927
0.2	0.163	0.291	0.401	0.499	0.589	0.674	0.755	0.834	0.914
0.3	0.150	0.270	0.375	0.470	0.560	0.645	0.729	0.813	0.902
0.4	0.137	0.249	0.349	0.442	0.530	0.616	0.703	0.792	0.889
0.5	0.124	0.229	0.323	0.413	0.500	0.587	0.677	0.772	0.876
0.6	0.111	0.208	0.297	0.384	0.470	0.559	0.651	0.751	0.863
0.7	0.099	0.187	0.271	0.355	0.441	0.530	0.625	0.730	0.850
0.8	0.086	0.166	0.245	0.326	0.411	0.501	0.599	0.709	0.837
0.9	0.073	0.145	0.219	0.297	0.381	0.472	0.573	0.688	0.824
q (WTP)									

Risk-averse PEVs (α represents constant absolute risk aversion)

$\alpha = 0.1$	$y = 0.1$			$y = 1$			$y = 10$		
p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
CE_L	0.020	0.050	0.080	0.192	0.488	0.792	1.352	3.799	7.046
1-q (WTA)									
0.1	0.031	0.062	0.085	0.302	0.608	0.849	2.252	5.062	7.836
0.2	0.029	0.059	0.083	0.282	0.578	0.828	2.093	4.782	7.582
0.3	0.027	0.056	0.081	0.262	0.549	0.806	1.934	4.503	7.328
0.4	0.025	0.053	0.079	0.241	0.519	0.785	1.775	4.223	7.074
0.5	0.023	0.050	0.077	0.221	0.489	0.764	1.616	3.943	6.820
0.6	0.021	0.047	0.075	0.200	0.459	0.742	1.457	3.664	6.566
0.7	0.019	0.044	0.073	0.180	0.430	0.721	1.298	3.384	6.312
0.8	0.017	0.041	0.071	0.160	0.400	0.699	1.139	3.105	6.058
0.9	0.014	0.038	0.069	0.139	0.370	0.678	0.980	2.825	5.804
q (WTP)									

$\alpha = 1$	$y = 0.1$			$y = 1$			$y = 10$		
p	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
CE_L	0.019	0.049	0.079	0.135	0.380	0.705	0.223	0.693	1.609
1-q (WTA)									
0.1	0.030	0.061	0.085	0.225	0.506	0.784	0.396	1.038	2.076
0.2	0.028	0.058	0.083	0.209	0.478	0.758	0.367	0.972	1.973
0.3	0.026	0.055	0.081	0.193	0.450	0.733	0.337	0.906	1.871
0.4	0.024	0.052	0.079	0.178	0.422	0.707	0.308	0.840	1.768
0.5	0.022	0.049	0.076	0.162	0.394	0.682	0.279	0.774	1.665
0.6	0.020	0.046	0.074	0.146	0.366	0.657	0.250	0.708	1.563
0.7	0.018	0.043	0.072	0.130	0.338	0.631	0.221	0.643	1.460
0.8	0.016	0.040	0.070	0.114	0.311	0.606	0.192	0.577	1.358
0.9	0.014	0.037	0.068	0.098	0.283	0.580	0.163	0.511	1.255
q (WTP)									

Conclusion

This collection of essays takes a broad look at some of the different aspects that make up the field of behavioural finance: the neurological, the behaviour of individuals, and the impact on markets. Each one providing some light on how people make financial decisions, why they make those decisions, and what the results of those decisions are.

The first essay demonstrates that irrational behaviour, and more specifically the language used to describe it, has a negative impact on stock markets. This is unsurprising when one considers how investors often wish to avoid uncertainty. With the previous focus on negativity and positivity in media and their effect on stock markets, our research into the effect of language suggestive of irrational behaviour is a useful addition to the study on how media can impact our financial institutions.

We find that not only does it lower subsequent stock returns, it also increases stock market volatility. We also find that small stocks take longer to react to this news than large stocks suggesting that this is complex information that takes longer to be integrated into less liquid stocks.

The second essay confirms that individuals use simplifying heuristics when confronted with more options than they can reasonably compare in a short space of time. What it also does, however, is test to see if there is an innate sense of harmony in certain ratios that presents itself in economic experiments in which participants have to split a given amount of money. I conclude that, although it is not a major predictor of economic behaviour, it is not wholly unpredictive either. This research confirms an important aspect of behavioural finance and has provided a number of important avenues for further study of the heuristic phenomenon.

The final essay takes a theoretical approach by looking at the mathematical models we've built to explain not only some of the many paradoxes seen in empirical economics, but also the process by which the brain makes economic decisions. Building on existing models of stochastic decision-making, I demonstrate that by taking into account the endowment effect, it's possible to not only predict the overweighting of small probabilities, the WTA-WTP disparity, and the preference reversal phenomenon, it's also possible to safeguard the important characteristics of the underlying model.

Together, these essays provide a glimpse into the breadth and reach of behavioural finance while offering a number of important and illuminating results.

Appendices

Appendix A

Appendix for Does Market Irrationality in the Media Affect Stock Returns?

A.1 The Irrationality Lexicon

Words marked with an asterisk (*) were not used to select the news articles in the first part of constructing the market irrationality sentiment measure but were later included when scoring the articles. Many are variations of other words that appear in the lexicon - added later for completeness, however, BUBBLE, BURST, and CRASH were deliberately omitted to err on the conservative side when detecting stock market irrationality in news articles. Those marked with a dagger (†) also appeared in the Loughran and McDonald *Fin-Neg* lexicon used in this paper.

ABERRANT
ABSURD
ABSURDITY
ACCURSED
ALARMING
ANARCHIC
ANARCHY
ANXIOUS
BAFFLE
BAFFLED
BAFFLING
BARBAROUS
BELLIGERENT
BERSERK
BIZARRE
BONKERS
BRAINLESS
BUBBLE*
BUBBLES*
BURST*
CALAMITY†
CAPRICIOUS
CHAOS
CHAOTIC
CHILDISH
COMMOTION
CONFOUND*
CONFOUNDED
CONFOUNDING
CONFUSE*†
CONFUSED†
CONFUSION†
CONTRADICTORY
CRASH*
CRASHED*
CRASHES*
CRASHING*
CRAZE
CRAZED
CRAZINESS
CRAZY
DAFT
DELIRIOUS
DELIRIUM
DELUSION
DELUSIONAL
DEMENTED
DEMENTIA
DEPRAVED
DERANGED
DESPAIR
DESPAIRED*
DESPAIRING
DISORDER
DISORDERED
DISORGANISE
DISORGANISED*
DISORGANIZE*
DISORGANIZED*
DISTRUST
DISTRUSTFUL
DISTRUSTING*
DIZZY

ECCENTRIC	INCONCEIVABLE	PARANOIA
ECCENTRICITY	INCONSISTENCIES*†	PARANOID
ENVIOUS	INCONSISTENCY†	PERPLEX
ERRATIC†	INCONSISTENT†	PERPLEXED
FANATIC	INSANE	PERPLEXING
FANATICAL	INSANITY	PERVERSE
FOOLISH	INSTABILITY†	PREPOSTEROUS
FOOLISHNESS	IRMATIONAL	PSYCHO
FRANTIC	IRMATIONALITY	PSYCHOTIC
FRANTICALLY	IRMESPONSIBLE	REASONLESS
FRAUGHT	JITTERY	STUPID
HAVOC	LUDICROUS	STUPIDITY
HYPOCRISY	LUNACY	SUPERSTITION
HYPOCRITE	LUNATIC	SUPERSTITIOUS
HYPOCRITICAL	MAD	UNHINGED
HYSTERIA	MADMAN	UNREASONABLE†
HYSTERIC	MADNESS	UNREASONABLY*†
HYSTERICAL	MANIA*	UNRELIABILITY
IDIOCY*	MANIC*	UNRELIABLE*†
IDIOT	MOODY	UNRELIABLY*
IDIOTIC	NEEDLESS	UNSETTLE*
IGNORANCE	NEUROTIC	UNSETTLED*
IGNORANT	NONSENSE	UNSETTLING
ILLOGICAL	NONSENSICAL	UNSOUND
IMPATIENT	OBSTINATE	UNSTABLE
INCOHERENCY*	PANIC	UNUSUAL†
INCOHERENT	PANICKING*	UNWISE

A.2 The Experts

Prior to our study, a lexicon consisting of words directly relating to the concept of irrationality was not readily available, despite diligent searching. In response to this, we worked to construct our own lexicon, starting by looking through the entire Harvard IV Psychosocial Dictionary for relevant words and compiling them into a single list. We then gathered a team of experts from the fields of neuroscience and psychology, gave them the word list, and the following set of instructions.

"We are looking at the use of language in connection with the way financial journalists describe stock markets and the people who trade on the stock market

You MUST complete this task independently from each other and with as little influence from others as possible. We recommend you go through the list on your own at least twice if you have the time.

1. Study the list of words IRMATIONALITY LEXICON.txt.

- If you think a word should appear in the IRMATIONAL category, indicate this with a '1' following the word.

- If you think a word should appear in the IRMATIONAL category but only when preceded by the word TOO, indicate this with a '1T' following the word.
- If you think a word should appear in the IRMATIONAL category but only when preceded by the word NOT, indicate this with a '1N' following the word.

- If you think a word should appear in the EXCESSIVE category, indicate this with a '2' following the word.
- If you think a word should appear in the EXCESSIVE category but only when preceded by the word TOO, indicate this with a '2T' following the word.
- If you think a word should appear in the EXCESSIVE category but only when preceded by the word NOT, indicate this with a '2N' following the word.

- If you think a word should appear in the INSUFFICIENT category, indicate this with a '3' following the word.
- If you think a word should appear in the INSUFFICIENT category but only when preceded by the word TOO, indicate this with a '3T' following the word.
- If you think a word should appear in the INSUFFICIENT category but only when preceded by the word NOT, indicate this with a '3N' following the word.

If you think a word should be rejected entirely, indicate this with a '0' following the word.

If a word fits into more than one of these categories, please list all of the appropriate categories with the most pertinent first.

If you think a word should appear in a related category other than the above, use successive numbers ('4', '5', '6', etc.) and a key explaining what category these numbers pertain to.

If the word fits into a custom category but only when preceded by the word TOO, append the number with the letter 'T'.

If the word fits into a custom category but only when preceded by the word NOT, append the number with the letter 'N'.

Notes.

Please rate words based on their most common usage. If a word could be considered IRMATIONAL in one context but is more likely to occur in a context where this is not the case, please do not include it in the IRMATIONAL category.

Categories.

IRMATIONAL. Words that strongly imply that the subject is acting irrationally.

EXCESSIVE. Words that strongly imply that the actions the subject is taking are inappropriate in the sense that they go too far (overaction or overreaction).

INSUFFICIENT. Words that strongly imply that the actions the subject is taking are inappropriate in the sense that they don't go far enough (insubstantial action or reaction).

2. When we've received the reports from all three judges, we will put all words into the categories that at least 2 out of 3 judges have put them in.

New words that were included by a single judge will be put into a separate list alongside the number corresponding to its suggested category.

New words that were included by more than one judge and put into a category that both judges agree on, will be placed into that category without further investigation.

New categories will be supplied as separate lists and will contain any words that a judge felt belonged in that category that were not put into an existing category by both of the other judges.

New categories will also be assigned their own number e.g. '4'.

Words that were not assigned in the first phase or were introduced in the first phase, either by a single judge or by multiple judges with conflicting categorisations, will be provided in a new list.

This list should then be studied and processed in the same way the first one was, except now with the new categories in mind."

After the round 1 lexicons had been submitted, we made some additional rules on the definition of agreement in order to prevent too many words from being unassigned. If 2 experts agreed on a category but one added the preceding word "Too" and the other one didn't, we counted that as an agreement and included the word in that category WITH the preceding word "Too".

If one expert included a word with "Too" in the Excessive (Insufficient) category and another expert included the same word with "Not" in the Insufficient (Excessive) category, and the third expert did not contradict this (by suggesting the word fit into no category for example), we counted this as an agreement and included the word in both categories with the suggested preceding words.

Note: We did not use the EXCESSIVE or INSUFFICIENT lexicons in the final study, nor did we use any of the words with the preceding "too" or "not", thus those words are not included in Appendix I to prevent confusion.

The CVs of the three experts who validated the market irrationality lexicon are available upon request.

A.3 The Articles

The articles were chosen by searching the Factiva database from the 1st January 1997 to the 31st December 2012 for any articles that contained any of the words from the Irrationality lexicon within a five word radius of any of the words "Market", "Markets", "Dow", "NASDAQ", or "NYSE" and did not include the phrases "Moody's", "Dow Jones reported", or "Dow Jones said". "Moody's" is excluded because "moody" is contained in the Irrationality lexicon. Articles are limited to those written in English and relating to the North American continent and were sourced from the entire range of Dow Jones newswires.

The articles are categorised based on several rules.

- Any article whose headline contains either "Highlights" or "Summary" is categorised as a summary. In general, these articles are compliations of news stories that cover a broad range of news items. More often than not, the specific news item we are

interested in is published separately and appears in multiple summaries over a 24-hour period receiving perhaps 3 or 4 extra matches than we would expect from a standard news article. We keep these articles for robustness checks.

- Similarly, any article that has strictly more than 10% of its lines beginning with a numeral is categorised as a summary. This is because these articles normally consist of a time-stamped rundown of events over an extended period of time.
- Any article whose headline contains "RealTick" is excluded. These articles only ever come in the form of tables and we have no interest in those.
- Any article containing fewer than 50 words is excluded because these are usually just headlines with no body.
- Any line within an article that contains 3 or more consecutive spaces followed by any text is excluded. This has the effect of removing most of the tables embedded in an article without losing any of the major information; it also removes some subheadings, some cases of the authors' names in the texts, and links to other articles that have no relation to the article in question due to their tendency to be indented. Any article that has 50% or more of its lines categorised in this way is excluded entirely.
- Any line within an article that contains the string "http" is excluded as this very rarely refers to anything other than an advert for the site that published the article.

A.4 Company Selection

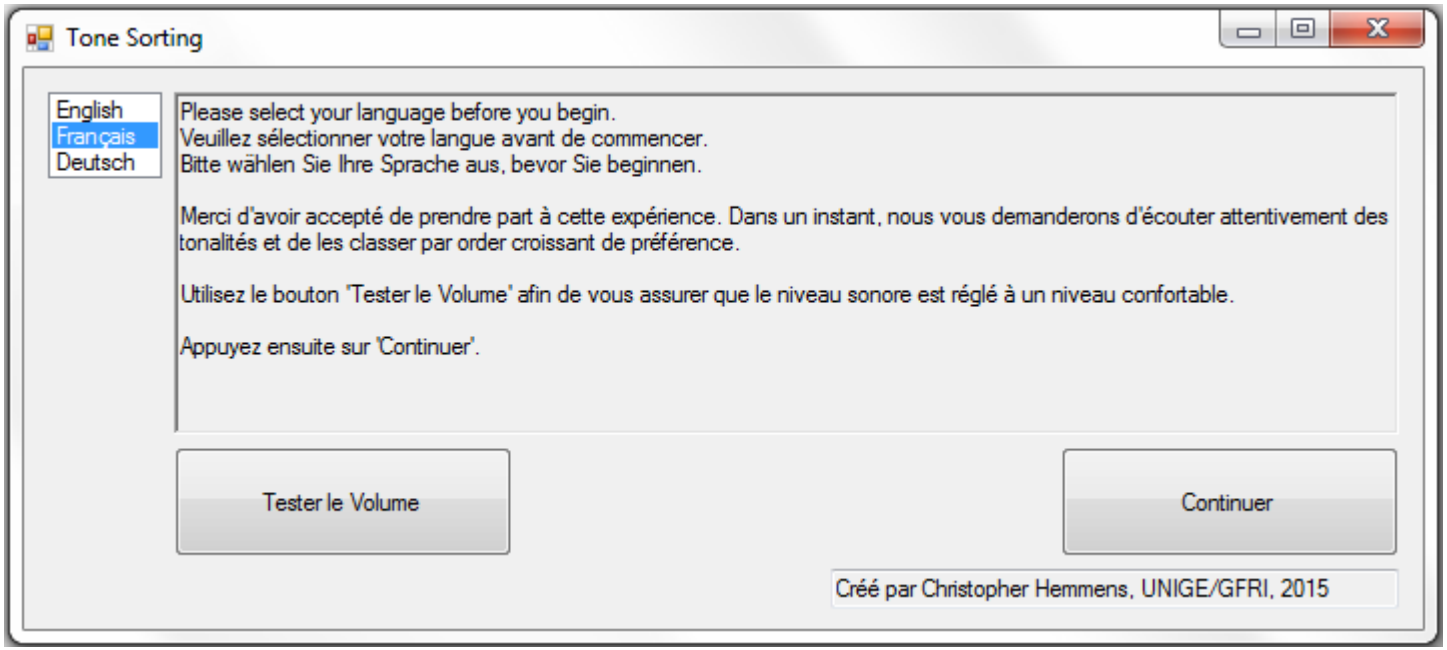
Start with 720 unique CUSIPs.

- 3 companies have overlapping inclusions in the index based on PERMCO. This leaves us with 717 unique companies.
- We submit these CUSIPS to CRSP and retrieve their returns data. 687 unique companies remain.
- We keep only observations whose trading status is active, whose share code is 10 or 11, and whose exchange code identifies the NYSE, AMEX, or NASDAQ. 656 unique companies remain.
- We remove any observations that occurred prior to 365 calendar days before a company's entry into the index and any observations following its exit, any observations with no return data, and any observations belonging to a company with fewer than 250 observations following these changes. This leaves us with a final total of 637 unique companies.

Appendix B

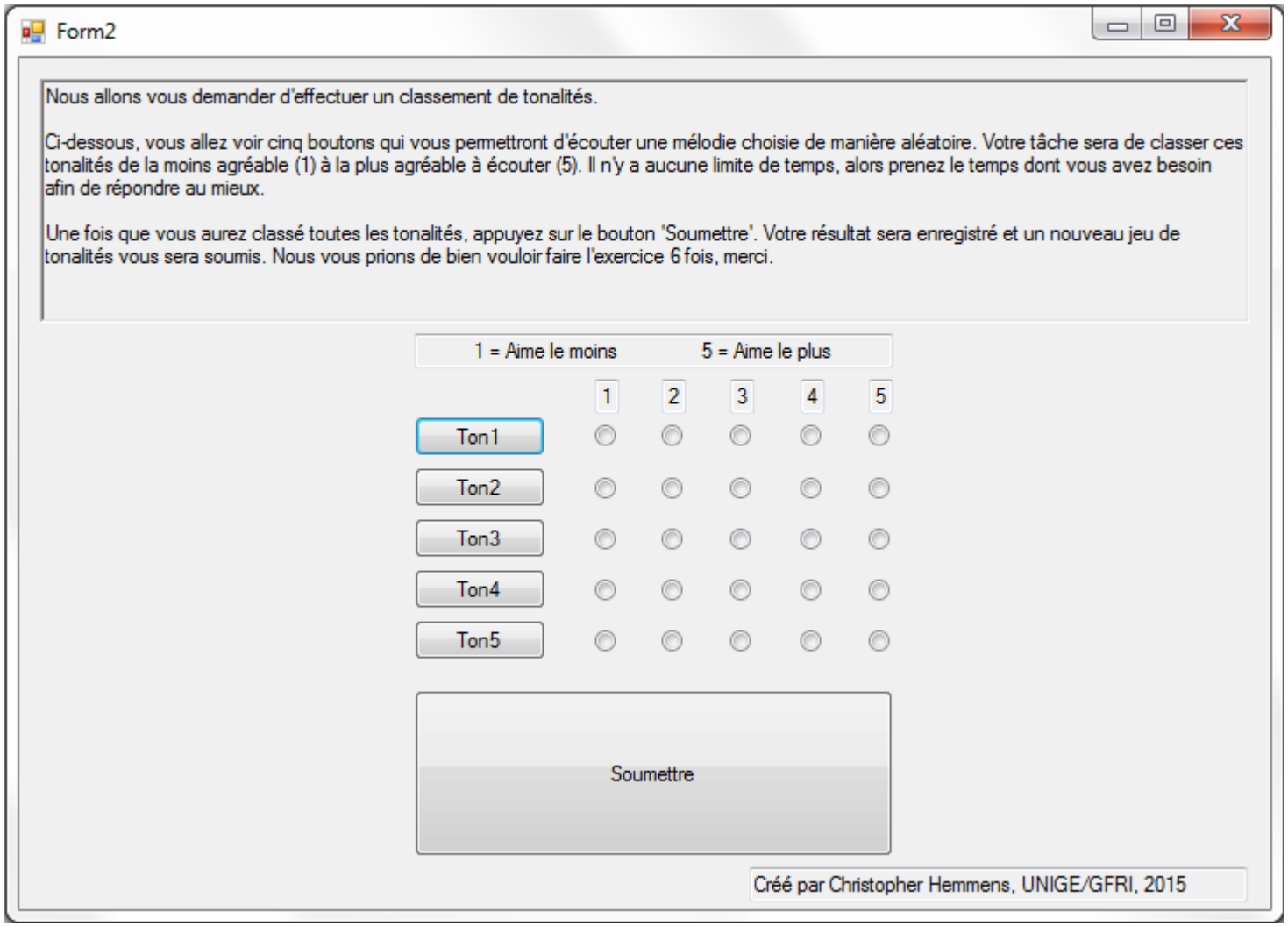
Appendix for Introduction of the aesthetic heuristic in analysing money-sharing experiments

Figure B.1: Steps in the Music Experiment 1



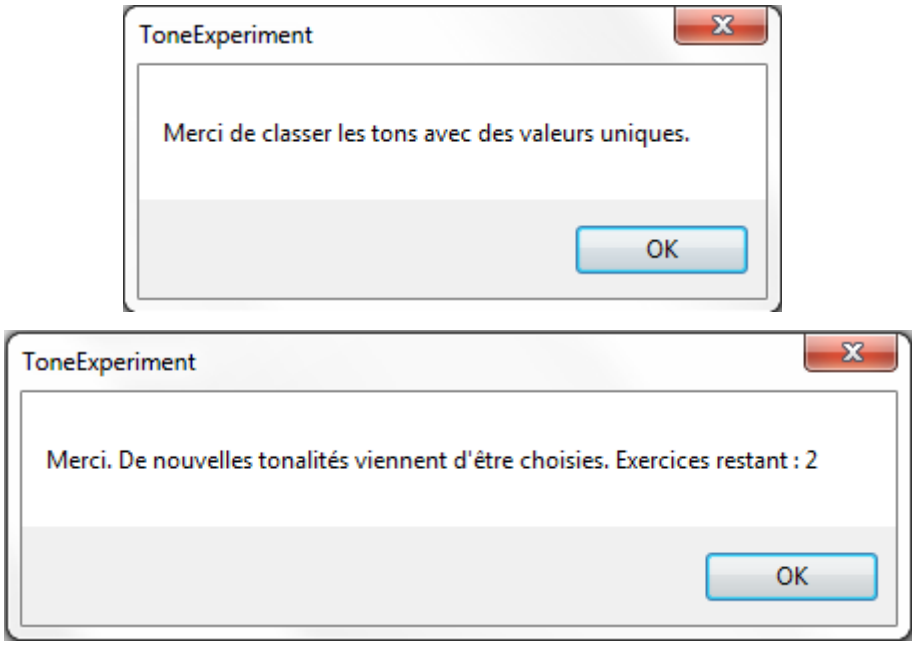
The first window presented to the participant gives the main instructions for the experiment: asking them to choose the best language for them and to test the volume on their computer. The text changed dynamically whenever a different language was chosen and the default language upon opening the program was French. When the 'Test Volume' button was pressed, a single piano note played to allow the participant to modify the sound to a volume they found comfortable; they could press this button as many times as they liked. If they tried to continue without testing the volume, they received a warning telling them they could not continue without doing so. I later ignored the results of these participants because I inferred they were not taking the experiment seriously.

Figure B.2: Steps in the Music Experiment 2



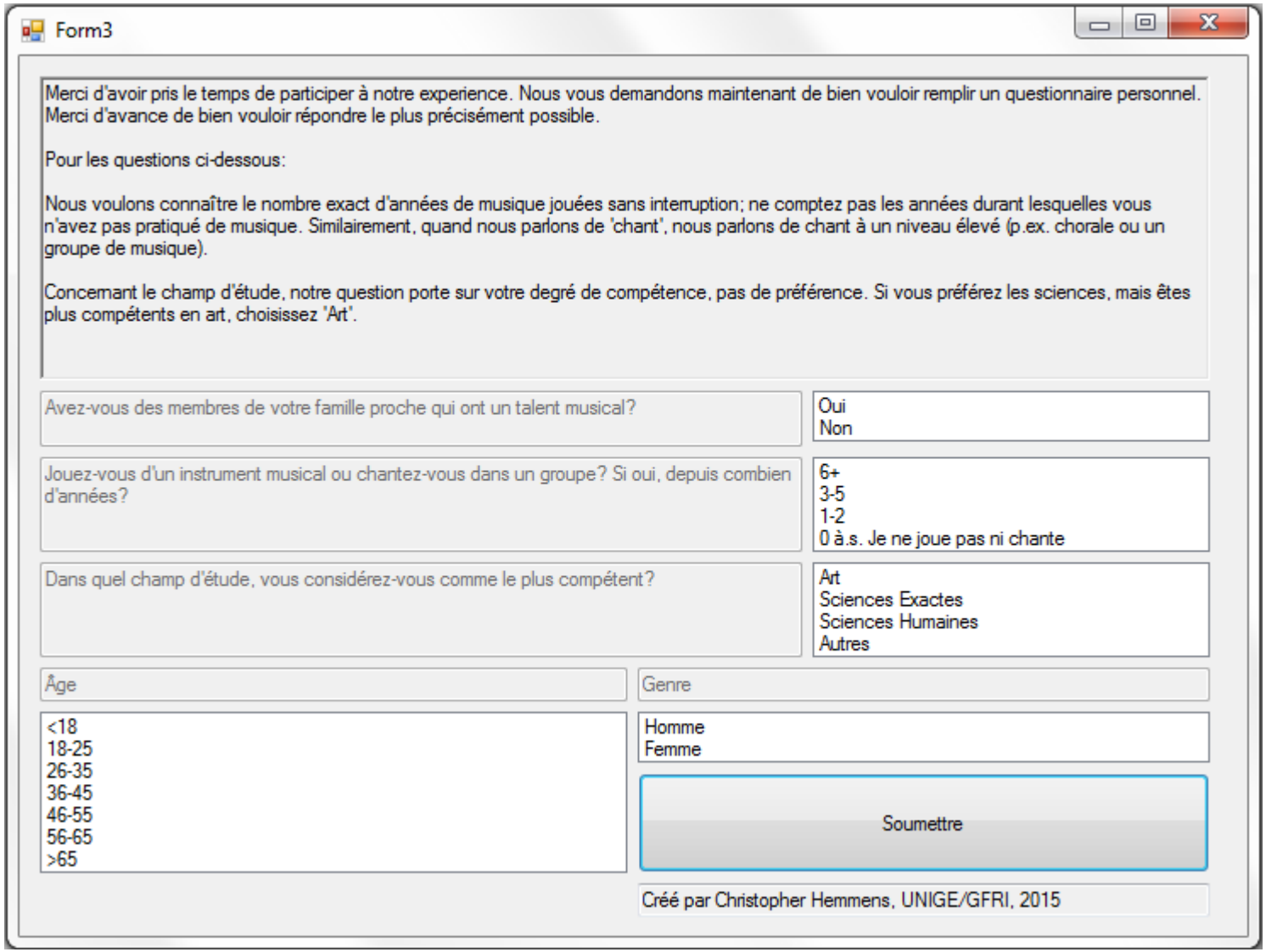
On continuing, participants were presented with the second window. 5 tones were randomly selected from a list of 33 and assigned randomly to the buttons 'Tone1', 'Tone2', etc. Participants could click these buttons to hear each tone as many times as they liked. They were only able to continue when they had selected one and only one option in column 1, column 2, etc. and clicked 'Submit'.

Figure B.3: Steps in the Music Experiment 3



If a participant tried to continue without selecting one and only one option from each column, they saw the top message shown here and were not allowed to continue until they had done so. If they did it correctly, they were shown the bottom message shown here with the relevant number of trials remaining and a new selection of 5 randomly selected tones were assigned to the buttons in the second window shown on the previous page. If they were completing the last trial, they would see neither of these messages but rather the third window shown on the following page.

Figure B.4: Steps in the Music Experiment 4



The final window asked the participant a number of demographic questions shown here. They were not allowed to continue until all questions had been answered and, once their answers had been submitted, the program told them to indicate this to the experimenter.

Appendix C

Appendix for Stronger utility and the endowment effect

If the model presented in this paper correctly predicts the preference reversal phenomenon as described, then it must satisfy the condition

$$\left(\frac{\partial \int G_Q(P(x,L))\mathrm{d}G_Q(P(x,L'))}{\partial Q}\right)_{Q=0.5} < 0.$$

In the tables below, step I splits $\int G_Q(P(x,L))\mathrm{d}G_Q(P(x,L'))$ into its component parts; step II takes the partial derivative with respect to Q; step III equates Q to 0.5. The sum of the components in step III represent the LHS of the inequality above.

Case: $\underline{CE_L \leq CE_{L'}}$

	$0 - CE_L$	$CE_L - CE_{L'}$	$CE_{L'} - z$
I	$4Q^2 \int P(L)\mathrm{d}P(L')$	$4Q(1-Q) \int P(L)\mathrm{d}P(L') + (4Q^2 - 2Q) \int \mathrm{d}P(L')$	$4(1-Q)^2 \int P(L)\mathrm{d}P(L') + (-4Q^2 + 6Q - 2) \int \mathrm{d}P(L')$
II	$8Q \int P(L)\mathrm{d}P(L')$	$4(1-2Q) \int P(L)\mathrm{d}P(L') + (8Q - 2) \int \mathrm{d}P(L')$	$-8(1-Q) \int P(L)\mathrm{d}P(L') + (-8Q + 6) \int \mathrm{d}P(L')$
III	$4 \int P(L)\mathrm{d}P(L')$	$2 \int \mathrm{d}P(L')$	$-4 \int P(L)\mathrm{d}P(L') + 2 \int \mathrm{d}P(L')$

$$(\div \text{-}4\ ) \implies \int_{CE_{L'}}^{z} P(x,L)\mathrm{d}P(x,L') - \frac{1}{2}\int_{CE_{L'}}^{z} \mathrm{d}P(x,L') - \frac{1}{2}\int_{CE_L}^{CE_{L'}} \mathrm{d}P(x,L') - \int_{0}^{CE_L} P(x,L)\mathrm{d}P(x,L') > 0$$

$$\implies \int_{CE_L}^{z} P(x,L)\mathrm{d}P(x,L') - \frac{1}{2}\int_{CE_L}^{z} \mathrm{d}P(x,L') - \int_{0}^{CE_{L'}} P(x,L)\mathrm{d}P(x,L') > 0$$

$$\implies \int_{CE_L}^{z} \left(P(x,L) - \frac{1}{2}\right)\mathrm{d}P(x,L') - \int_{0}^{CE_{L'}} P(x,L)\mathrm{d}P(x,L') > 0 \ \square$$

Case: $\underline{CE_{L'} \leq CE_L}$

	$0 - CE_{L'}$	$CE_{L'} - CE_L$	$CE_L - z$
I	$4Q^2 \int P(L)\mathrm{d}P(L')$	$4Q(1-Q) \int P(L)\mathrm{d}P(L')$	$4(1-Q)^2 \int P(L)\mathrm{d}P(L') + (-4Q^2 + 6Q - 2) \int \mathrm{d}P(L')$
II	$8Q \int P(L)\mathrm{d}P(L')$	$4(1-2Q) \int P(L)\mathrm{d}P(L')$	$-8(1-Q) \int P(L)\mathrm{d}P(L') + (-8Q + 6) \int \mathrm{d}P(L')$
III	$4 \int P(L)\mathrm{d}P(L')$	0	$-4 \int P(L)\mathrm{d}P(L') + 2 \int \mathrm{d}P(L')$

$$(\div \text{-}4\ ) \implies \int_{CE_L}^{z} P(x,L)\mathrm{d}P(x,L') - \frac{1}{2}\int_{CE_L}^{z} \mathrm{d}P(x,L') - \int_{0}^{CE_{L'}} P(x,L)\mathrm{d}P(x,L') > 0$$

$$\implies \int_{CE_L}^{z} \left(P(x,L) - \frac{1}{2}\right)\mathrm{d}P(x,L') - \int_{0}^{CE_{L'}} P(x,L)\mathrm{d}P(x,L') > 0 \ \square$$

List of Tables

List of Figures

Bibliography

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, and Peter P Wakker (2011). "The rich domain of uncertainty: Source functions and their experimental implementation". In: *The American Economic Review* 101.2, pp. 695–723.

Ackert, Lucy F, Brian D Kluger, and Li Qi (2012). "Irrationality and beliefs in a laboratory asset market: Is it me or is it you?" In: *Journal of Economic Behavior & Organization* 84.1, pp. 278–291.

Agranov, Marina and Pietro Ortoleva (2015). "Stochastic Choice and Preferences for Randomization". In: *Available at SSRN 2644784*.

Allais, Maurice (1953). "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine". In: *Econometrica: Journal of the Econometric Society*, pp. 503–546.

Anderson, Lisa R, Yana V Rodgers, and Roger R Rodriguez (2000). "Cultural differences in attitudes toward bargaining". In: *Economics Letters* 69.1, pp. 45–54.

Andreoni, James, Marco Castillo, and Ragan Petrie (2003). "What do bargainers' preferences look like? Experiments with a convex ultimatum game". In: *American Economic Review*, pp. 672–685.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang (2006). "The Cross-Section of Volatility and Expected Returns". In: *The Journal of Finance* 61, pp. 259–299.

Antweiler, Werner and Murray Z. Frank (2004). "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards". In: *The Journal of Finance* 59, pp. 1259–1294.

Atallah, Gamal (2006). "Opportunity costs, competition, and Firm Selection". In: *International Economic Journal* 20.4, pp. 409–430.

Baillon, Aurélien, Laure Cabantous, and Peter P Wakker (2012). "Aggregating imprecise or conflicting beliefs: An experimental investigation using modern ambiguity theories". In: *Journal of risk and uncertainty* 44.2, pp. 115–147.

Barber, Brad, Terrance Odean, and Ning Zhu (2006). "Do noise traders move markets?" In: *EFA 2006 Zurich meetings paper*.

Barker, Andrew (2010). "Mathematical Beauty Made Audible: Musical Aesthetics in Ptolemy's Harmonics". In: *Classical Philology* 105.4, pp. 403–420.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak (1963). "Stochastic models of choice behavior". In: *Behavioral science* 8.1, pp. 41–55.

Bejan, Adrian (2009). "The golden ratio predicted: Vision, cognition and locomotion as a single design in nature". In: *International Journal of Design & Nature and Ecodynamics* 4.2, pp. 97–104.

Bethwaite, Judy and Paul Tompkinson (1996). "The ultimatum game and non-selfish utility functions". In: *Journal of Economic Psychology* 17.2, pp. 259–271.

Biais, Bruno, Peter Bossaerts, and Chester Spatt (2010). "Equilibrium asset pricing and portfolio choice under asymmetric information". In: *Review of Financial Studies* 23.4, pp. 1503–1543.

Blavatskyy, Pavlo R (2007). "Stochastic expected utility theory". In: *Journal of Risk and Uncertainty* 34.3, pp. 259–286.

— (2009). "Preference reversals and probabilistic decisions". In: *Journal of Risk and Uncertainty* 39.3, pp. 237–250.

— (2011). "A model of probabilistic choice satisfying first-order stochastic dominance". In: *Management Science* 57.3, pp. 542–548.

— (2013). "A probability weighting function for cumulative prospect theory and meangini approach to optimal portfolio investment". In: *Available at SSRN 2380484*.

— (2014). "Stronger utility". In: *Theory and decision* 76.2, pp. 265–286.

Blavatskyy, Pavlo R and Ganna Pogrebna (2007). "Loss aversion? Not with half-a-million on the table!" In:

Blavatskyy, Pavlo and Ganna Pogrebna (2008). "Risk aversion when gains are likely and unlikely: Evidence from a natural experiment with large stakes". In: *Theory and Decision* 64.2-3, pp. 395–420.

Bornstein, Gary and Ilan Yaniv (1998). "Individual and group behavior in the ultimatum game: Are groups more "rational" players?" In: *Experimental Economics* 1.1, pp. 101–108.

Brennan, Thomas J and Andrew W Lo (2011). "The origin of behavior". In: *The Quarterly Journal of Finance* 1, pp. 55–108.

— (2012). "An evolutionary model of bounded rationality and intelligence". In: *PloS one* 7.11, e50310.

Brocas, Isabelle and Juan D Carrillo (2014). "Dual-process theories of decision-making: A selective survey". In: *Journal of economic psychology* 41, pp. 45–54.

Brown, Gregory W (1999). "Volatility, sentiment, and noise traders". In: *Financial Analysts Journal*, pp. 82–90.

Burns, Zach, Andrew Chiu, and George Wu (2010). "Overweighting of small probabilities". In: *Wiley Encyclopedia of Operations Research and Management Science*.

Buschena, David and David Zilberman (2000). "Generalized expected utility, heteroscedastic error, and path dependence in risky choice". In: *Journal of Risk and Uncertainty* 20.1, pp. 67–88.

Butler, David J and Graham C Loomes (2007). "Imprecision as an account of the preference reversal phenomenon". In: *The American Economic Review*, pp. 277–297.

Butler, David, Andrea Isoni, and Graham Loomes (2012). "Testing the 'standard' model of stochastic choice under risk". In: *Journal of Risk and Uncertainty* 45.3, pp. 191–213.

Camerer, Colin F and Teck-Hua Ho (1994). "Violations of the betweenness axiom and nonlinearity in probability". In: *Journal of risk and uncertainty* 8.2, pp. 167–196.

Cameron, Lisa A (1999). "Raising the stakes in the ultimatum game: Experimental evidence from Indonesia". In: *Economic Inquiry* 37.1, pp. 47–59.

Carhart, Mark M (1997). "On persistence in mutual fund performance". In: *The Journal of finance* 52, pp. 57–82.

Cubitt, Robin P, Daniel Navarro-Martinez, and Chris Starmer (2015). "On preference imprecision". In: *Journal of Risk and Uncertainty* 50.1, pp. 1–34.

Da, Zhi, Joseph Engelberg, and Pengjie Gao (2011). "In search of attention". In: *The Journal of Finance* 66, pp. 1461–1499.

— (2015). "The Sum of All FEARS Investor Sentiment and Asset Prices". In: *Review of Financial Studies* 28, pp. 1–32.

Daniel, Kent and Sheridan Titman (2000). *Market efficiency in an irrational world*. Tech. rep.

Das, Sanjiv R. and Mike Y. Chen (2007). "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web". In: *Management Science* 53, pp. 1375–1388.

De Long, J Bradford, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann (1990). "Noise trader risk in financial markets". In: *Journal of political Economy*, pp. 703–738.

Dehaene, Stanislas (2003). "The neural basis of the Weber–Fechner law: a logarithmic mental number line". In: *Trends in cognitive sciences* 7.4, pp. 145–147.

Diamond, Peter and Botond Köszegi (2003). "Quasi-hyperbolic discounting and retirement". In: *Journal of Public Economics* 87, pp. 1839–1872.

Dow, James and Gary Gorton (2006). *Noise traders*. Tech. rep.

Drichoutis, Andreas C and Jayson L Lusk (2014). "Judging statistical models of individual decision making under risk using in-and out-of-sample criteria". In: *PloS one* 9.7, e102269.

Duffie, Darrell and Larry G Epstein (1992). "Stochastic differential utility". In: *Econometrica: Journal of the Econometric Society*, pp. 353–394.

Eckel, Catherine C and Philip J Grossman (1996). "Altruism in anonymous dictator games". In: *Games and economic behavior* 16, p. 181.

— (2001). "Chivalry and solidarity in ultimatum games". In: *Economic Inquiry* 39.2, pp. 171–188.

Ehrlich, Isaac and Gary S Becker (1972). "Market insurance, self-insurance, and self-protection". In: *Journal of political Economy* 80.4, pp. 623–648.

Ellingsen, Tore and Magnus Johannesson (2001). "Sunk Costs, Fairness, and Disagreement". In: *Stockholm School of Economics Working Paper*.

Etner, Johanna and Meglena Jeleva (2013). "Risk perception, prevention and diagnostic tests". In: *Health economics* 22.2, pp. 144–156.

Falk, Armin and Urs Fischbacher (2006). "A theory of reciprocity". In: *Games and economic behavior* 54.2, pp. 293–315.

Fama, Eugene F and Kenneth R French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of financial economics* 33, pp. 3–56.

Fama, Eugene F and James D MacBeth (1973). "Risk, return, and equilibrium: Empirical tests". In: *The journal of political economy*, pp. 607–636.

Fechner, Gustav Theodor (1860). "Elements of psychophysics". In:

Fehr, Ernst and Simon Gächter (1999). "Cooperation and punishment in public goods experiments". In: *Institute for Empirical Research in Economics working paper* 10.

Fehr, Ernst and Klaus M Schmidt (2000). "Fairness, incentives, and contractual choices". In: *European Economic Review* 44.4, pp. 1057–1068.

Fehr-Duda, Helga, Adrian Bruhin, Thomas Epper, and Renate Schubert (2010). "Rationality on the rise: Why relative risk aversion increases with stake size". In: *Journal of Risk and Uncertainty* 40.2, pp. 147–180.

Fershtman, Chaim and Uri Gneezy (2001). "Strategic delegation: An experiment". In: *RAND Journal of Economics*, pp. 352–368.

Forsythe, Robert, Joel L Horowitz, N Eugene Savin, and Martin Sefton (1994). "Fairness in simple bargaining experiments". In: *Games and Economic behavior* 6.3, pp. 347–369.

Frederick, Shane, Nathan Novemsky, Jing Wang, Ravi Dhar, and Stephen Nowlis (2009). "Opportunity cost neglect". In: *Journal of Consumer Research* 36.4, pp. 553–561.

Fudenberg, Drew and Tomasz Strzalecki (2015). "Dynamic logit with choice aversion". In: *Econometrica* 83.2, pp. 651–691.

García, Diego (2013). "Sentiment during Recessions". In: *The Journal of Finance* 68, pp. 1267–1300.

Gibson, Rajna and Songtao Wang (2016). "Earnings Belief Risk and the Cross- Section of Stock Returns". In: *Working Paper*.

Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr (2003). "Explaining altruistic behavior in humans". In: *Evolution and Human Behavior* 24.3, pp. 153–172.

Green, Donald, Karen E Jacowitz, Daniel Kahneman, and Daniel McFadden (1998). "Referendum contingent valuation, anchoring, and willingness to pay for public goods". In: *Resource and Energy Economics* 20.2, pp. 85–116.

Greiner, Ben et al. (2003). *The online recruitment system ORSEE: a guide for the organization of experiments in economics*. Max-Planck-Inst. for Research into Economic Systems, Strategic Interaction Group.

Grether, David M and Charles R Plott (1979). "Economic theory of choice and the preference reversal phenomenon". In: *The American Economic Review* 69.4, pp. 623–638.

Güth, Werner, Steffen Huck, and Wieland Müller (2001). "The relevance of equal splits in ultimatum games". In: *Games and Economic Behavior* 37.1, pp. 161–169.

Güth, Werner, Nadege Marchand, and Jean-Louis Rulliere (1997). *On the reliability of reciprocal fairness: An experimental study*. Wirtschaftswissenschaftliche Fakultät Berlin.

Hagen, Edward H and Peter Hammerstein (2006). "Game theory and human evolution: A critique of some recent interpretations of experimental games". In: *Theoretical population biology* 69.3, pp. 339–348.

Handgraaf, Michel JJ, Eric van Dijk, Henk AM Wilke, and Riël C Vermunt (2004). "Evaluability of outcomes in ultimatum bargaining". In: *Organizational Behavior and Human Decision Processes* 95.1, pp. 97–106.

Harless, David W and Colin F Camerer (1994). "The predictive utility of generalized expected utility theories". In: *Econometrica: Journal of the Econometric Society*, pp. 1251–1289.

Harsanyi, John C (1973). "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points". In: *International Journal of Game Theory* 2.1, pp. 1–23.

Hey, John D (1995). "Experimental investigations of errors in decision making under risk". In: *European Economic Review* 39.3, pp. 633–640.

Hey, John D and Chris Orme (1994). "Investigating generalizations of expected utility theory using experimental data". In: *Econometrica: Journal of the Econometric Society*, pp. 1291–1326.

Hirshleifer, David (2001). "Investor psychology and asset pricing". In: *The Journal of Finance* 56, pp. 1533–1597.

— (2015). "Behavioral finance". In: *Annual Review of Financial Economics* 7, pp. 133–159.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon Smith (1994). "Preferences, property rights, and anonymity in bargaining games". In: *Games and Economic Behavior* 7.3, pp. 346–380.

Hollard, Guillaume, Hela Maafi, and Jean-Christophe Vergnaud (2015). "Consistent inconsistencies? Evidence from decision under risk". In: *Theory and Decision*, pp. 1–26.

Holt, Charles A and Susan K Laury (2005). "Risk aversion and incentive effects: New data without order effects". In: *The American economic review* 95.3, pp. 902–904.

Holt, Charles A, Susan K Laury, et al. (2002). "Risk aversion and incentive effects". In: *American economic review* 92.5, pp. 1644–1655.

Horowitz, John K and Kenneth E McConnell (2002). "A review of WTA/WTP studies". In: *Journal of Environmental Economics and Management* 44.3, pp. 426–447.

Huynh, Thanh D and Daniel R Smith (2013). "News Sentiment and Momentum". In: *FIRN Research Paper*.

Isoni, Andrea, Graham Loomes, and Robert Sugden (2011). "The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations: Comment". In: *The American Economic Review* 101.2, pp. 991–1011.

Jaspersen, Johannes G (2015). "Hypothetical Surveys and Experimental Studies of Insurance Demand: A Review". In: *Journal of Risk and Insurance*.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler (1991). "Anomalies: The endowment effect, loss aversion, and status quo bias". In: *The journal of economic perspectives* 5.1, pp. 193–206.

Knetsch, Jack L and John A Sinden (1984). "Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value". In: *The Quarterly Journal of Economics*, pp. 507–521.

Laibson, David (1997). "Golden eggs and hyperbolic discounting". In: *The Quarterly Journal of Economics* 112, pp. 443–478.

Li, Feng (2006). "Do stock market investors understand the risk sentiment of corporate annual reports". In: *Available at SSRN*, pp. 1–53.

Lichtenstein, Sarah and Paul Slovic (1971). "Reversals of preference between bids and choices in gambling decisions." In: *Journal of experimental psychology* 89.1, p. 46.

— (1973). "Response-induced reversals of preference in gambling: An extended replication in Las Vegas." In: *Journal of Experimental Psychology* 101.1, p. 16.

Linde, Jona and Joep Sonnemans (2012). "Social comparison and risky choices". In: *Journal of Risk and Uncertainty* 44.1, pp. 45–72.

Lindman, Harold R (1971). "Inconsistent preferences among gambles." In: *Journal of Experimental Psychology* 89.2, p. 390.

Loomes, Graham, Peter G Moffatt, and Robert Sugden (2002). "A microeconometric test of alternative stochastic theories of risky choice". In: *Journal of risk and Uncertainty* 24.2, pp. 103–130.

Loomes, Graham and Ganna Pogrebna (2014). "Testing for independence while allowing for probabilistic choice". In: *Journal of Risk and Uncertainty* 49.3, pp. 189–211.

Loomes, Graham and Robert Sugden (1995). "Incorporating a stochastic element into decision theories". In: *European Economic Review* 39.3, pp. 641–648.

Loomes, Graham, Ganna Pogrebna, et al. (2015). "Do preference reversals disappear when we allow for probabilistic choice?" In:

Loughran, Tim and Bill McDonald (2011). "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". In: *The Journal of Finance* 66, pp. 35–65.

Louie, Kenway, Mel W Khaw, and Paul W Glimcher (2013). "Normalization is a general neural mechanism for context-dependent decision making". In: *Proceedings of the National Academy of Sciences* 110.15, pp. 6139–6144.

Luce, R Duncan (1959). "On the possible psychophysical laws." In: *Psychological review* 66.2, p. 81.

Luce, R Duncan and Peter C Fishburn (1991). "Rank-and sign-dependent linear utility models for finite first-order gambles". In: *Journal of risk and Uncertainty* 4.1, pp. 29–59.

Luce, R Duncan and David M Green (1974). "The response ratio hypothesis for magnitude estimation". In: *Journal of Mathematical Psychology* 11.1, pp. 1–14.

Machina, Mark J (1985). "Stochastic choice functions generated from deterministic preferences over lotteries". In: *The economic journal* 95.379, pp. 575–594.

Manela, Asaf and Alan Moreira (2013). "News implied volatility and disaster concerns". In: *Available at SSRN 2382197*.

Markowitz, Harry (1952). "The utility of wealth". In: *The Journal of Political Economy*, pp. 151–158.

Mehra, Rajnish and Edward C Prescott (1985). "The equity premium: A puzzle". In: *Journal of monetary Economics* 15.2, pp. 145–161.

Nash, John (1951). "Non-Cooperative Games". In: *Annals of Mathematics* 54.2, pp. 286–295.

Nelissen, Rob MA, Dorien SI Van Someren, and Marcel Zeelenberg (2009). "Take it or leave it for something better? Responses to fair offers in ultimatum bargaining". In: *Journal of Experimental Social Psychology* 45.6, pp. 1227–1231.

Oosterbeek, Hessel, Randolph Sloof, and Gijs Van De Kuilen (2004). "Cultural differences in ultimatum game experiments: Evidence from a meta-analysis". In: *Experimental Economics* 7.2, pp. 171–188.

Plott, Charles R and Kathryn Zeiler (2005). "The Willingness to Pay–Willingness to Accept Gap, the". In: *The American Economic Review* 95.3, pp. 530–545.

Roch, Sylvia G, John AS Lane, Charles D Samuelson, Scott T Allison, and Jennifer L Dent (2000). "Cognitive load and the equality heuristic: A two-stage model of resource overconsumption in small groups". In: *Organizational behavior and human decision processes* 83.2, pp. 185–212.

Ruffle, Bradley J (1998). "More is better, but fair is fair: Tipping in dictator and ultimatum games". In: *Games and Economic Behavior* 23.2, pp. 247–265.

Schmidt, Ulrich and John D Hey (2004). "Are preference reversals errors? An experimental investigation". In: *Journal of Risk and Uncertainty* 29.3, pp. 207–218.

Schmidt, Ulrich, Chris Starmer, and Robert Sugden (2008). "Third-generation prospect theory". In: *Journal of Risk and Uncertainty* 36.3, pp. 203–223.

Schotter, Andrew, Avi Weiss, and Inigo Zapater (1996). "Fairness and survival in ultimatum and dictatorship games". In: *Journal of Economic Behavior & Organization* 31.1, pp. 37–56.

Seidl, Christian (2002). "Preference reversal". In: *Journal of Economic Surveys* 16.5, pp. 621–655.

Shapiro, Samuel Sanford and Martin B Wilk (1965). "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52.3/4, pp. 591–611.

Shiller, R.J. (2005). *Irrational Exuberance*.

Slembeck, Tilman (1999). "Reputations and fairness in bargaining-experimental evidence from a repeated ultimatum game with fixed opponents". In:

Slovic, Paul (1975). "Choice between equally valued alternatives." In: *Journal of Experimental Psychology: Human Perception and Performance* 1.3, p. 280.

Swait, Joffre and AAJ Marley (2013). "Probabilistic choice (models) as a result of balancing multiple goals". In: *Journal of Mathematical Psychology* 57.1, pp. 1–14.

Tausczik, Yla R. and James W. Pennebaker (2010). "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods". In: *Journal of Language and Social Psychology* 29, pp. 24–54.

Tetlock, Paul C. (2007). "Giving Content to Investor Sentiment: The Role of Media in the Stock Market". In: *The Journal of Finance* 62, pp. 1139–1168.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy (2008). "More Than Words: Quantifying Language to Measure Firms' Fundamentals". In: *The Journal of Finance* 63, pp. 1437–1467.

Thurstone, Louis L (1927). "A law of comparative judgment." In: *Psychological review* 34.4, p. 273.

Tirunillai, Seshadri and Gerard J. Tellis (2012). "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance". In: *Marketing Science* 31, pp. 198–215.

Tunçel, Tuba and James K Hammitt (2014). "A new meta-analysis on the WTP/WTA disparity". In: *Journal of Environmental Economics and Management* 68.1, pp. 175–187.

Tversky, Amos (1969). "Intransitivity of preferences." In: *Psychological review* 76.1, p. 31.

Tversky, Amos and Daniel Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty". In: *Journal of Risk and uncertainty* 5.4, pp. 297–323.

Ubeda, Paloma (2014). "The consistency of fairness rules: An experimental study". In: *Journal of Economic Psychology* 41, pp. 88–100.

Van de Kuilen, Gijs (2009). "Subjective probability weighting and the discovered preference hypothesis". In: *Theory and Decision* 67.1, pp. 1–22.

Von Neumann, John and Oskar Morgenstern (1944). *Theory of games and economic behaviour*. Princeton University Press.

Ward, Lawrence M (1973). "Repeated magnitude estimations with a variable standard: Sequential effects and other properties". In: *Perception & Psychophysics* 13.2, pp. 193–200.

Weg, Eythan and Vernon Smith (1993). "On the failure to induce meager offers in ultimatum game". In: *Journal of economic psychology* 14.1, pp. 17–32.

Wilcox, Nathaniel T et al. (2015). *Unusual Estimates of Probability Weighting Functions*. Tech. rep.

Wu, George and Richard Gonzalez (1996). "Curvature of the probability weighting function". In: *Management science* 42.12, pp. 1676–1690.

Yuan, Yu (2015). "Market-wide attention, trading, and stock returns". In: *Journal of Financial Economics* 116, pp. 548–564.